

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

## Rt or RDt, that is the question!

### **This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1789676> since 2021-06-06T10:51:18Z

*Published version:*

DOI:10.19191/EP20.5-6.S2.102

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

# ***R<sub>t</sub> or RD<sub>t</sub>, that is the question!***

**Maria Teresa Giraud<sup>1</sup>, Manuele Falcone<sup>2</sup>, Ennio Cadum<sup>3</sup>, Silvia Deandrea<sup>4</sup>, Salvo Scondotto<sup>5</sup>, Andrea Mattaliano<sup>6</sup>, Carlo Di Pietrantonj<sup>7</sup>, Lucia Bisceglia<sup>8</sup>, Piergiorgio Duca<sup>9</sup>, Cesare Cislighi<sup>10</sup>**

1 Dipartimento di matematica G. Peano, Università di Torino

2 Agenzia regionale di sanità Toscana, Firenze

3 Dipartimento igiene e prevenzione sanitaria; Unità operativa complessa salute e ambiente e progetti innovativi, Agenzia di tutela della salute, Pavia

4 Unità operativa complessa salute e ambiente e progetti innovativi; Unità operativa semplice epidemiologia ambientale, Agenzia di tutela della salute, Pavia

5 Dipartimento attività sanitarie e osservatorio epidemiologico, Regione Sicilia, Palermo

6 KPMG Italy, Palermo

7 Servizio di riferimento regionale di epidemiologia per la sorveglianza, la prevenzione e il controllo delle malattie infettive, Azienda sanitaria locale, Alessandria

8 Area epidemiologia e *care intelligence*, Agenzia regionale strategica per la salute e il sociale Puglia, Bari

9 Università degli Studi di Milano

10 Economista epidemiologo

## **BOX**

### **COSA SI SAPEVA GIA'**

- Per misurare la capacità di contagio di un agente infettivo in assenza di misure di contenimento si utilizza l'indice di contagiosità di base chiamato  $R_0$ , erre con zero.
- Lo sviluppo dell'epidemia può essere monitorato con lo stesso indicatore calcolato al passare del tempo anche in funzione delle misure di contenimento, quindi non considerando più tutta la popolazione come suscettibile; in questo caso si usa  $R_t$ , erre con t, come nome dell'indice.
- Le difficoltà di utilizzo dell'indice derivano sia dalla complessità di calcolo e dalla necessità di competenze statistiche e di informazioni sui cosiddetti tempi di generazione del contagio che dalla disponibilità dell'informazione sulla data di inizio sintomi, di difficile e talvolta impossibile rilevazione.

### **COSA SI AGGIUNGE**

- Gli unici dati di incidenza dell'epidemia da Covid-19 diffusi pubblicamente sono state le frequenze giornalieri di nuovi casi di soggetti risultati positivi al test molecolare (tamponi).
- Per analizzare l'"accelerazione" o la "decelerazione" della diffusione dei contagi si è messo a punto un indice detto di *replicazione diagnostica* ( $RD_t$ ), in grado anch'esso di segnalare le variazioni nello sviluppo dell'epidemia.
- Le analisi qui presentate indicano come i due indici,  $R_t$  e  $RD_t$ , forniscano indicazioni molto simili e segnalino che l'indice  $RD_t$  risulta più facilmente utilizzabile e comprensibile anche da parte dei decisori di sanità pubblica che non dispongano di specifiche competenze statistiche.

## **INTRODUZIONE**

Nelle giornate drammatiche dell'esplosione epidemica tutti aspettavano con ansia di sapere le variazioni dell'intensità della dinamica dell'epidemia e tra gli indici più seguiti c'erano sicuramente l' $R_t^{(1)}$  e, in alcuni casi, l' $RD_t^{(2)}$ . Il primo è considerato più ufficiale in quanto diffuso dall'Istituto Superiore di Sanità<sup>(3)</sup> e il secondo più usufruibile data la minor difficoltà di calcolo e la completa disponibilità dei dati tramite cui ricavarlo. I due indici hanno un diverso rationale e una diversa modalità di calcolo, ma soprattutto richiedono tipologie diverse di dati per essere elaborati.

Nel seguito verranno introdotti i due indici, i cui metodi di calcolo sono specificati negli allegati all.1S e all.2S, e i dati su cui si basano. Verranno poi presentati il confronto tra i due indici e tra valori dell' $RD_t$  diversamente computati relativi a un'area del Nord Italia, riportando negli allegati all.3S e all.4S i risultati analoghi per altre tre zone. Si presenteranno quindi una discussione inerente la variabilità casuale dell'indice  $RD_t$ , dettagliata nell'allegato all.5S, l'interpretazione degli indici e la loro praticabilità e infine alcune conclusioni.

### **Il rationale**

Il rationale dell'indice  $R_t$  si richiama a quello dell'indice  $R_0^{(4)}$  che viene usato dagli infettivologi come misura di contagiosità di un determinato agente infettivo. L'indice  $R_0$  indica quanti soggetti mediamente vengono infettati dal paziente "zero" in mancanza di qualsiasi misura di contenimento del contagio.

L'indice  $R_t$  fornisce una misura simile durante lo svolgimento dell'epidemia ed è quindi determinato non solo dall' $R_0$  dipendente dall'agente infettivo, ma anche dalle misure preventive attuate per il contenimento dei possibili contagi. Dalle definizioni derivano anche il loro nome:  $R_0$  "indice di riproduzione di base",  $R_t$  "numero di riproduzione" o "indice di trasmissibilità" o "indice di contagio al tempo  $t$ ".

Il rationale dell'indice  $RD_t$  è invece differente e deriva dalle tecniche della *time series analysis* che si usano per analizzare le serie storiche, cioè per l'andamento di una misura o di una frequenza di un evento che si sviluppa in funzione del tempo. Per fare semplici esempi possiamo ricordare l'andamento della temperatura giornaliera o la crescita di una somma di denaro data in investimento. L' $RD_t$  non prende in considerazione il momento del contagio bensì la data della diagnosi di positività e per questo motivo viene chiamato "indice di replicazione diagnostica", in quanto vuole descrivere l'intensità dello sviluppo della frequenza di casi riconosciuti come positivi nella popolazione.

Per entrambi gli indici è chiaro che se il loro valore è  $>1$  significa che gli eventi analizzati, contagi o positività, sono in crescita, se è  $=1$  significa che sono stazionari, se è  $<1$  significa invece che sono in decrescita.

## **METODI**

### **I metodi di calcolo**

Esistono vari algoritmi di calcolo per ricavare i valori dell'indice  $R_t^{(1),(5),(6),(7)}$ , basati su interpretazioni diverse del suo significato. Seppure apparentemente molto diversi, i due metodi qui utilizzati per il calcolo degli indici  $R_t$  e  $RD_t$  sono più simili di quanto si pensi in quanto fondamentalmente sono il rapporto tra una stima degli eventi al tempo  $t$  e degli eventi al tempo  $t-s$  dove  $s$  sono i giorni che si stima siano necessari affinché un caso possa riprodurre un altro. L'intervallo  $s$  viene denominato in vario modo: tempo di generazione, serial interval, lag.

La differenza è che, mentre per  $R_t$  il rapporto è tra due medie non pesate di due periodi separati da lag differenti, nel caso dell' $RD_t$  il rapporto è con i casi dei giorni precedenti a un determinato  $s$  pesati con una distribuzione teorica dei tempi di generazione. Si può utilizzare per esempio una distribuzione Gamma<sup>(8)</sup>, che è una distribuzione di probabilità continua che comprende, come casi particolari, anche le distribuzioni esponenziale e chi quadro. Per specificare la distribuzione da utilizzare se ne devono definire il valore atteso e la varianza.

Nel caso dell' $RD_t$ , invece, il rapporto è tra una media mobile centrata non pesata di  $n$  elementi al tempo  $t$  e una media mobile centrata non pesata di  $n$  elementi al tempo  $t-s$ , dove  $s$  può assumere diversi possibili valori solitamente pari a 4, 5, 6, 7 o 8 giorni. La scelta del lag è legata alle necessità interpretative del fenomeno che richiedono di considerare i tempi di replicazione del contagio e potrebbe quindi essere diversa, ai fini di risultare informativa, per altri tipi di epidemia.

E' comunque interessante notare che, se i dati su cui vengono calcolati sono gli stessi, i due indici  $R_t$  ed  $RD_t$  assumono valori molto simili e le discrepanze che si osservano non sembrano comunque poter dare adito a diversità interpretative del fenomeno. Una maggiore divergenza nei valori si può però riscontrare quando le frequenze diventano molto esigue; in questi casi tuttavia entrambi gli indici perdono probabilmente molto del loro significato ed anzi possono dare indicazioni ingannevoli.

Per la descrizione più puntuale dei metodi di calcolo dell' $R_t$  e dell' $RD_t$  si vedano gli allegati (vedi file all.1S per  $R_t$  e file all.2S per  $RD_t$ )

## **DATI**

### ***I dati su cui vengono calcolati***

La reale diversità che si può rilevare nei valori dei due indici  $R_t$  e  $RD_t$  può dipendere dai dati che vengono utilizzati per il loro calcolo. Le date che rivestono rilevanza sono le date in cui avviene il contagio, in cui sorgono i primi sintomi, quelle in cui il soggetto contatta la struttura sanitaria, quelle in cui viene fatta una diagnosi clinica, quelle in cui viene fatto un test molecolare (tampone) che determina inequivocabilmente la positività al virus e infine quelle in cui i risultati del test vengono inseriti nel sistema informativo. L' $R_t$  utilizza la data dell'inizio sintomi, l' $RD_t$  la data di registrazione della positività.

Sarebbe sicuramente ottimale poter conoscere con esattezza la data in cui è avvenuto il contagio, ma ciò non è di solito possibile dato che per lo più non si riesce a determinare il contesto esatto in cui si è verificato. Si cerca quindi di stimarlo utilizzando la data dell'inizio sintomi e prevedendo appunto che i tempi tra il contagio e l'esordio dei sintomi seguano una distribuzione Gamma i cui parametri vengano stimati utilizzando dei casi in cui entrambi gli eventi siano determinabili con sufficiente precisione.

*La supposta costanza dei tempi di generazione:* Va innanzitutto osservato come sia debole l'ipotesi che la distribuzione dei *serial interval*, cioè dei tempi di generazione tra un contagio ed il successivo, rimanga costante durante tutto il periodo dell'epidemia e sia la stessa in diversi ambiti o in diverse aree. Questo tempo dipende da molti fattori, ma, oltre che dal possibile delay tra momento dell'infezione e capacità di contagiare, dipende soprattutto dalla frequenza dei rapporti interpersonali in quanto dove questi sono poco frequenti il *serial interval* si dilata, mentre in situazioni di contagio avvenute in comunità, come nelle RSA o negli ospedali, esso dipende probabilmente solo dalle capacità del virus di svilupparsi nell'individuo contagiato. Ciò significa che la distanza temporale tra contagio ed inizio sintomi non dovrebbe essere costante, né in media né in varianza, per situazioni differenti. Pensando di scomporre l'intervallo seriale in durata

di infettività e tempo di contatto interpersonale, si potrebbe eventualmente considerare costante la prima ed assegnare una distribuzione di probabilità al secondo, ma quest'ultima potrebbe comunque sempre variare a seconda delle circostanze.

*La rilevazione e la registrazione della data di inizio sintomi:* La data di inizio sintomi viene necessariamente ricavata solamente dall'anamnesi dei pazienti al momento dell'esito positivo del tampone. Ci sono casi in cui la rilevazione è impossibile, vuoi perché il paziente non è in grado di esprimersi o di ricordare, vuoi perché il paziente è immediatamente deceduto, vuoi perché il paziente è asintomatico; ci sono poi tutte le imprecisioni originate dal *recall bias* ed è comunque difficile attribuire i sintomi indicati proprio al coronavirus. Le basi dati di test relative a quattro aree italiane (una provincia e una Regione del Nord Italia e due Regioni del Sud Italia) utilizzate per l'analisi sono state estratte dai rispettivi sistemi informativi il 20 maggio 2020 e si riferiscono al periodo compreso tra il primo caso e la settimana precedente per quanto riguarda l'analisi sulla completezza della data inizio sintomi e tra il primo caso e il 30 aprile per l'analisi su  $R_t$  e  $RD_t$ . Facciamo quindi presente come si possa ritenere che le caratteristiche di completezza delle basi dati riflettano quanto disponibile a livello di sistemi informativi locali o regionali per la gestione dell'operatività durante il corso del fenomeno epidemico e non il livello ottimale di qualità del dato che le stesse realtà possono raggiungere in fase di analisi e approfondimento post-epidemico.

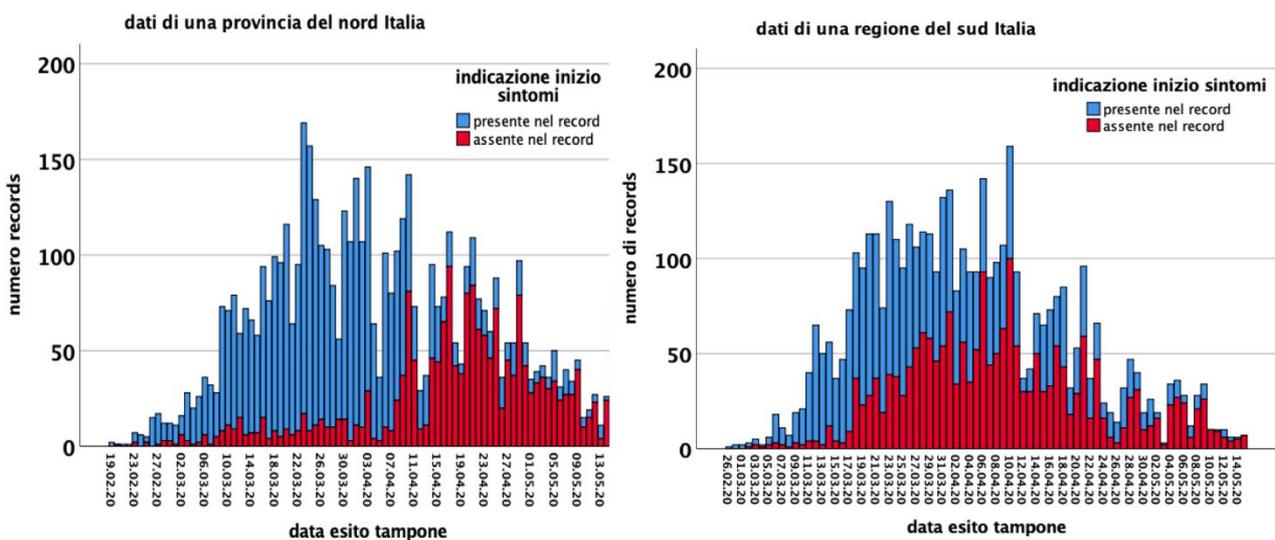


Figura 1: Proporzioni di rilevazioni di casi per data esito tampone rispetto alla presenza del dato su inizio sintomi

Figure 1: Proportions of detected case incidence with respect to availability of symptoms onset date

Oltre ai problemi di rilevazione vi sono quelli dei tempi di registrazione. Come indicano i due grafici nella Figura 1, ma soprattutto quello relativo alla provincia del Nord, la registrazione nella cartella clinica o la trascrizione nel sistema informativo sono state molto più lente nelle fasi avanzate dell'epidemia, quando il numero dei casi giornalieri è cresciuto di 5-10 volte e oltre la capacità del personale disponibile di registrarlo su supporto informatico in tempo reale (ma con recupero nel periodo successivo).

*La completezza dell'informazione inizio sintomi:* I problemi precedenti comportano quindi delle grosse carenze di completezza dell'informazione concernente la data di inizio sintomi, completezza che varia poi anche nel tempo e tra località diverse come si vede dai grafici nella Figura 2 relativi a due zone del Nord Italia e due del Sud.

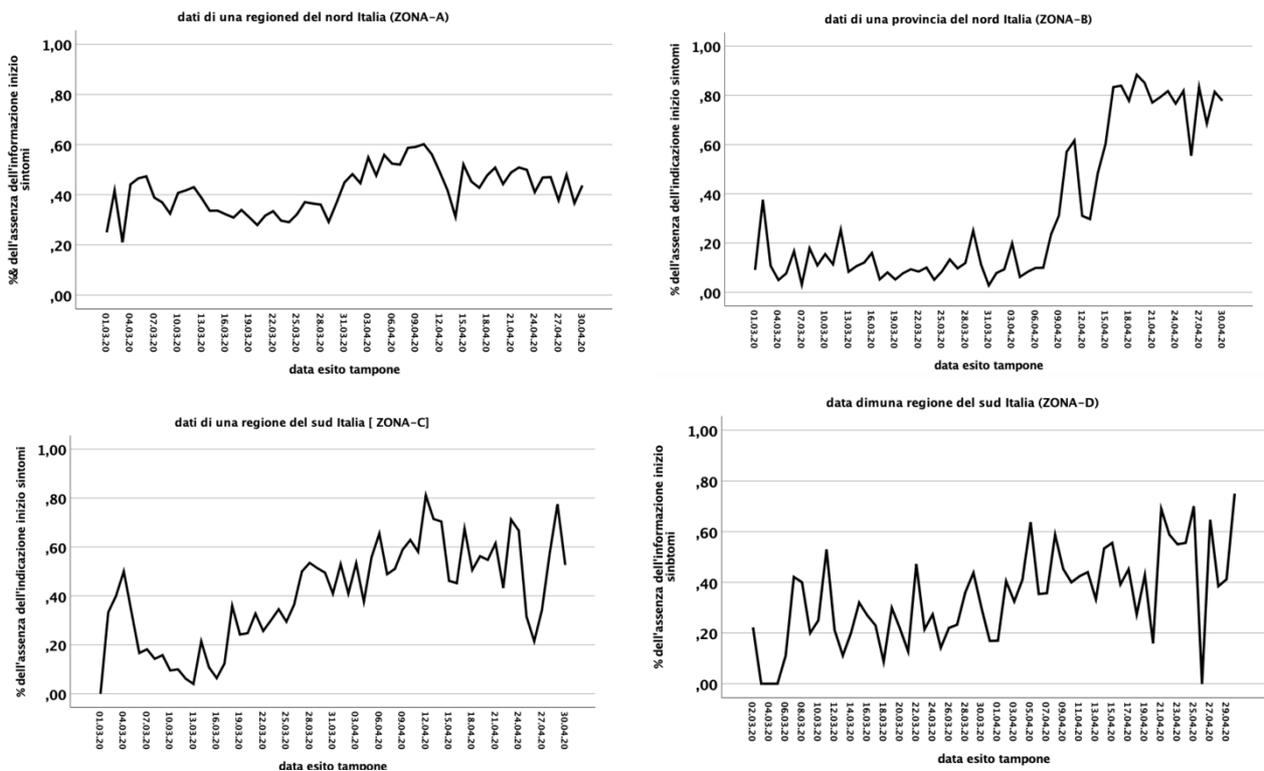


Figura 2: Percentuali di mancanza del dato su inizio sintomi rispetto alle incidenze giornaliere di casi rilevati tramite tampone nelle quattro zone considerate (A Regione del Nord, B provincia del Nord, C e D regioni del Sud)

Figure 2: Percentage of lack in date of symptoms onset with respect to daily incidence of cases detected by swab in the four areas considered (A Northern Region, B Northern Province, C and D two Southern Regions)

*Inizio sintomi contemporaneo all'esito tampone o anche molto successivo:* Una caratteristica della data di inizio sintomi che fa molto dubitare sulla sua utilizzabilità come proxy posticipata della data del contagio è il fatto che ci siano molti casi per cui viene dichiarata come contestuale alla data del tampone o anche molto successiva.

Dal grafico riportato in Figura 3 risulta che nelle quattro aree analizzate i dati di inizio sintomi realmente precedenti (anche solo di due giorni) alla data del tampone sono solo tra il 60% e l'80% dei dati validi, ma se consideriamo tutti i record dei casi positivi quelli che hanno la data del tampone di almeno due giorni successivi all'inizio sintomi sono tra il 35% ed il 52%.

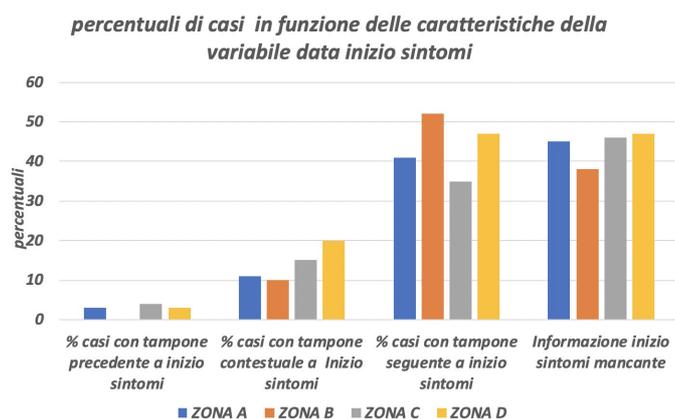


Figura 3: Percentuali di casi sul totale con diversa collocazione temporale del tampone rispetto all’inizio sintomi nelle quattro zone considerate (come in Figura 2)

Figure 3: Percentage of cases with different temporal location of the swab with respect to symptoms onset in the four areas considered (as in Figure 2)

Non si può esser certi se questi dati debbano essere letti come indicazione o meno della qualità dell’informazione relativa all’inizio sintomi, ma certamente non possono non creare delle perplessità sull’utilizzo di questa informazione che non sembra sufficientemente solida.

### **Inizio sintomi e tampone**

Se l’informazione sulla data inizio sintomi non appare sufficientemente disponibile ed attendibile, l’informazione sulla data del tampone appare invece decisamente migliore anche se la distanza tra questa data e l’inizio del contagio non è certo costante. La data del tampone dipende in larga misura dalle regole diagnostiche stabilite e che si sono modificate nel corso dell’epidemia. Nei primi giorni, data anche la disponibilità di un numero molto ridotto di tamponi che era possibile effettuare, questi venivano fatti solo a pazienti con sintomatologia grave che si presentavano al Pronto Soccorso.

Successivamente, i tamponi sono stati eseguiti in misura sempre maggiore anche su richiesta dei MMG o anche solo per verificare casi sospetti come quelli di coloro che avevano contattato dei contagiati. Per questo motivo le analisi dell’andamento dell’epidemia effettuate utilizzando la data dell’inizio sintomi o la data del tampone non coincidono.

	mese di marzo			mese di aprile			mese di maggio		
	media	mediana	dev. stand.	media	mediana	dev. stand.	media	mediana	dev. stand.
<b>zona A</b>	<b>5,24</b>	<b>4</b>	<b>7,46</b>	<b>7,67</b>	<b>5</b>	<b>11,44</b>	<b>16,58</b>	<b>7</b>	<b>20,90</b>
<b>zona B</b>	<b>4,06</b>	<b>4</b>	<b>7,15</b>	<b>7,13</b>	<b>6</b>	<b>5,93</b>	<b>7,67</b>	<b>5</b>	<b>9,81</b>
<b>zona C</b>	<b>4,64</b>	<b>4</b>	<b>5,73</b>	<b>5,77</b>	<b>3</b>	<b>9,81</b>	<b>7,82</b>	<b>2</b>	<b>13,25</b>
<b>zona D</b>	<b>4,04</b>	<b>3</b>	<b>5,87</b>	<b>5,58</b>	<b>2</b>	<b>9,61</b>	<b>17,59</b>	<b>5</b>	<b>22,78</b>

Tabella 1: Statistiche relative alla distanza tra data inizio sintomi e data tampone nelle quattro zone considerate (come in Figura 2)

Table 1: Statistical indexes for the distance between symptom onset date and swab date in the four areas considered (as in Figure 2)

Come si può vedere nella Tabella 1, le distanze medie tra data inizio sintomi e data tampone hanno avuto nelle quattro zone in marzo (data del tampone) una media dai 4,04 ai 5,24 giorni ed una mediana da 3 a 4, mentre in aprile una media molto più alta, da 5,28 a 7,67 giorni, con una mediana tra 2 e 6. Nel mese di maggio i valori delle medie si sono molto diversificati e le mediane si sono attestate tra 2 e 7 giorni. Questa variabilità della distanza tra data inizio sintomi e data tampone è il principale motivo per cui l'utilizzo delle due date non produce esattamente gli stessi risultati calcolando sia  $R_t$  che  $RD_t$ .

## **RISULTATI**

### ***Il confronto dei metodi attraverso i risultati ottenuti***

Il confronto tra le diverse applicazioni dei metodi, eseguito utilizzando a tal fine i medesimi dati per data inizio sintomi, e il confronto tra l'uso delle diverse tipologie di dati di monitoraggio eseguito per evidenziare le differenze tra l'informazione ricavabile dai due differenti set per inizio sintomi e per esito tampone, sono stati effettuati limitatamente alle quattro aree per cui è stato possibile disporre di dati individuali completi relativi ai mesi di marzo e di aprile.

Il confronto presentato nel seguito riguarda esclusivamente l'area popolosa corrispondente ad una Regione del Nord Italia; gli altri confronti sono riportati in allegato (vedi file all.3S). Per il calcolo dell' $R_t$  è stato utilizzato il software  $R^{(9),(10)}$ , mentre per il calcolo dell' $RD_t$  si è fatto ricorso al foglio di calcolo di Excel. I valori di  $R_t$  sono stati ottenuti utilizzando per gli intervalli seriali, grazie all'analisi di dati di cluster dell'epidemia in Lombardia, una distribuzione Gamma con valore medio pari a 6.6 giorni e deviazione standard pari a 4.88 giorni<sup>(8)</sup>.

### ***Confronto tra $R_t$ ed $RD_t$ calcolati sui dati di incidenza per data inizio sintomi***

I grafici in Figura 4 mostrano l'andamento giornaliero dell' $R_t$  e dell' $RD_t$  calcolato a vari lag, da 4 ad 8 giorni, relativo rispettivamente ai mesi di marzo e aprile e, come particolare, al solo mese di aprile, utilizzando le frequenze disponibili dei casi per data inizio sintomi.

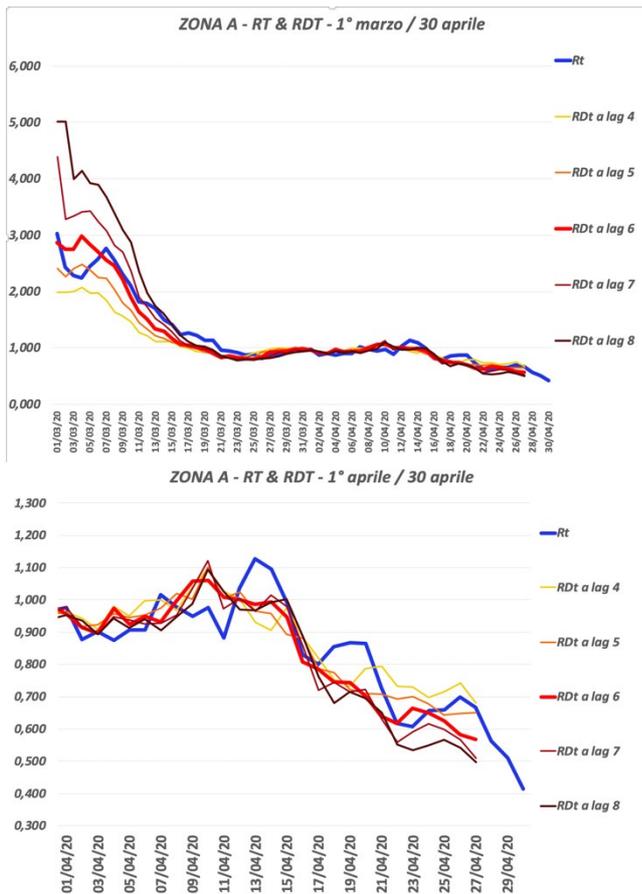


Figura 4: Andamento giornaliero dell' $R_t$  e dell' $RD_t$  per la zona A calcolato in base alla data inizio sintomi a vari lag per i mesi di marzo e aprile (in alto) e per il solo mese di aprile (in basso)  
 Figure 4: Daily trend of  $R_t$  and  $RD_t$  for area A calculated according to the date of symptoms onset at various lags in March and April (top) and in April only (bottom)

Si nota innanzitutto che, laddove l'andamento dell'epidemia cambia repentinamente, i valori dell' $RD_t$  si differenziano maggiormente tra i vari lag in quanto, ad esempio, la modifica dell'andamento a quattro giorni risulta meno marcata di quella dell'andamento a otto giorni. Si può rilevare la stessa divaricazione al termine del mese di aprile quando nuovamente le frequenze cambiano velocemente, anche se per valori molto inferiori.

L'andamento generale dell' $RD_t$ , per lo meno ad un lag intermedio (6 ovvero 7), sembra effettivamente molto simile a quello dell' $R_t$  e la correlazione risulta infatti molto alta, come mostrato nella Figura 5. La correlazione massima tra  $R_t$  e  $RD_t$  viene raggiunta a lag 7 con un valore di  $r$  che supera il 97% ( $R^2 = 0,944$ ), per cui ciascun indice spiega la variabilità dell'altro per quasi il 95%. In corrispondenza di tale lag lo scarto assoluto medio tra i due indici assume il valore minimo di 0.14.

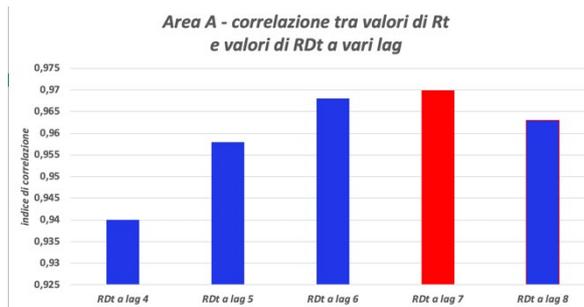


Figura 5: Correlazione tra i valori degli indici  $R_t$  e  $RD_t$  ai diversi lag riportati nella Figura 4 in alto  
 Figure 5: Correlation between the values of  $R_t$  and  $RD_t$  indices shown in Figure 4 above

Osservando l'andamento nel solo mese di marzo, con una scala chiaramente maggiore sulle ordinate, emergono alcune differenze tra i due indici che possono essere spiegate innanzitutto dal fatto che il calcolo dell' $R_t$  qui effettuato considera solo quattro giornate come riferimento della situazione precedente, mentre l' $RD_t$  è calcolato come rapporto di medie mobili di incidenze a sette giorni. Un'altra ragione può derivare dal fatto che, mentre l' $RD_t$  assegna a ogni giorno lo stesso peso nel calcolo della media mobile, l' $R_t$  utilizza la distribuzione Gamma per i tempi di generazione del contagio per cui può talvolta dare più rilevanza ad una giornata in cui vi siano stati, anche solo casualmente, più o meno casi. A questo proposito va rilevato che nell'incidenza registrata si osserva un andamento settimanale ripetitivo con dei minimi nelle giornate da domenica a martedì e dei massimi tra giovedì e venerdì, conseguenti ai ritmi di lavorazione dei laboratori che eseguivano analisi dei test molecolari.

*Confronto tra diversi  $RD_t$  calcolati su dati degli stessi soggetti positivi per data inizio sintomi o per data esito tampone*

Per illustrare i differenti risultati che si ottengono calcolando l' $RD_t$  ai vari lag per gli stessi casi positivi, ma utilizzando le frequenze per data inizio sintomi oppure per data esito tamponi, riportiamo di seguito nella Figura 6 solo i grafici relativi alla Regione del Nord Italia rimandando all'allegato i grafici analoghi relativi alle altre aree (vedi file all.4S).

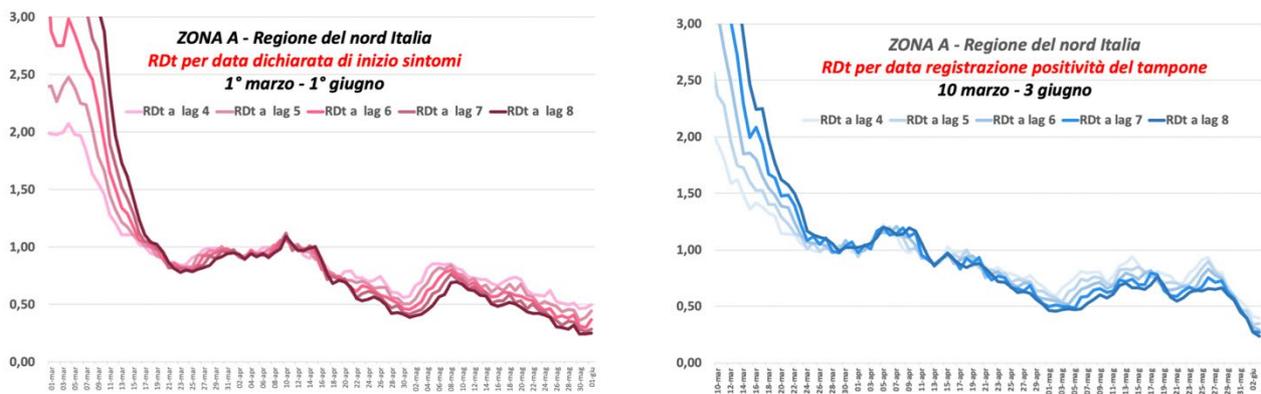


Figura 6: Valori dell' $RD_t$  ai vari lag per la zona A considerando le incidenze sintomatiche (a sinistra) e di diagnosi (a destra)  
 Figure 6:  $RD_t$  values at various lags from 10 March to 10 June for area A considering symptomatic (left) and diagnostic incidence (right)

L'andamento dei due grafici è simile, ma non uguale perché innanzitutto il numero dei soggetti non è lo stesso in quanto molti casi non riportano, come si è già detto, la data di inizio sintomi, e inoltre perché la distanza tra l'esordio dei sintomi e l'esito dei tamponi è variabile sia tra i soggetti che in media tra le diverse fasi dell'epidemia.

Come si può vedere dal pannello a sinistra della Figura 7, la media giornaliera della distanza tra inizio sintomi e esito tampone nella Regione qui descritta varia tra un valore di circa 2 giorni a inizio marzo per diventare di 6 al 1° aprile e poi aumentare fino quasi a raddoppiare.

E' quindi evidente come questa distanza tra i due eventi monitorati produca degli andamenti differenti negli indici di replicazione diagnostica con discrepanze tra di essi a loro volta variabili nel tempo.

Nel pannello a destra della Figura 7 è riportato l'andamento dell'indice  $RD_t$  calcolato in base alle frequenze per data esito tampone (linea blu) e alle frequenze per data inizio sintomi (linea punteggiata rossa). Vengono inoltre riportate alcune copie di quest'ultima linea traslate in verso positivo e negativo di vari giorni (linee sottili nere) e viene in particolare evidenziata in rosso quella spostata in avanti di 6 giorni.

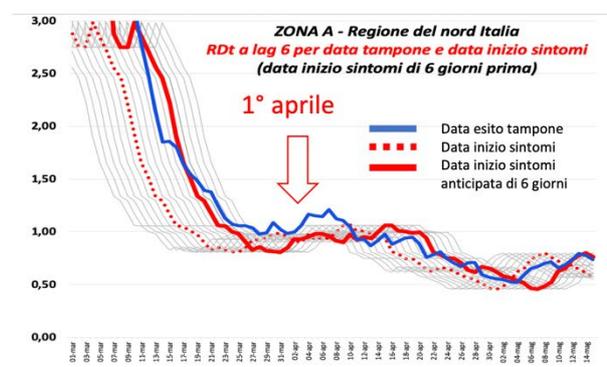
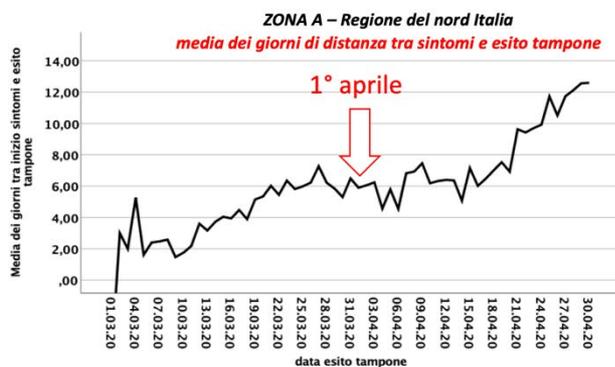


Figura 7: Media giornaliera della distanza tra inizio sintomi e esito tampone per la zona A (a sinistra) e andamento dell'indice  $RD_t$  in base alla data esito tampone (linea blu) e alla data inizio sintomi (linee rosse e nere)

Figure 7: Daily average of the distance between symptoms onset and swab outcome for area A (left) and trend of the  $RD_t$  index according to swab outcome (blue line) and symptoms onset (red and black lines).

Si può notare che in corrispondenza del 1° aprile, giorno in cui la distanza tra gli eventi era di 6 giorni, le linee sono quasi coincidenti mentre non è così né prima né dopo.

Da queste considerazioni emerge che gli indici di replicazione valutati in base alla data di esordio sintomi o di esito tampone rappresentano due situazioni differenti tra di loro che sono abbastanza simili, ma non pienamente corrispondenti. Se interessa maggiormente capire l'evoluzione del contagio forse la curva dell' $RD_t$  per inizio sintomi è più informativa, mentre se l'interesse risiede nella presa in carico dei casi positivi da parte del sistema sanitario allora risulta più efficace monitorare l'andamento dell' $RD_t$  per data esito tampone. Si osservi comunque che i dati di inizio sintomi non sono completi e la loro validità è legata al ricordo talvolta difficoltoso del paziente; in alcuni casi inoltre la data non può essere determinata perché lo stesso non è in grado di parlare o di ricordare o perché è asintomatico. Quest'ultima situazione diventa via via più diffusa al diminuire dei casi di contagio e quindi a maggior ragione l'unico indice che riesce meglio a monitorare l'andamento dell'epidemia è quello che utilizza la data dell'esito del tampone.

## DISCUSSIONE

### La variabilità casuale dell'indice $RD_t$

Come tutti gli indicatori anche i numeri di riproduzione hanno una variabilità che in parte può essere attribuita al caso. Riportiamo in allegato (vedi file all.5S) la descrizione dei metodi tramite cui è possibile ricavare un intervallo di confidenza per l'indice  $RD_t$ ; qui preme solo rilevare come sia opportuno leggere questi indici rendendosi conto della possibile incertezza ad essi associata. Nella Figura 8 sono rappresentati gli andamenti dell' $RD_t$  in Lombardia e in Puglia per i mesi di maggio e giugno, nei quali le frequenze lombarde sono state poco elevate e quelle pugliesi addirittura minime.

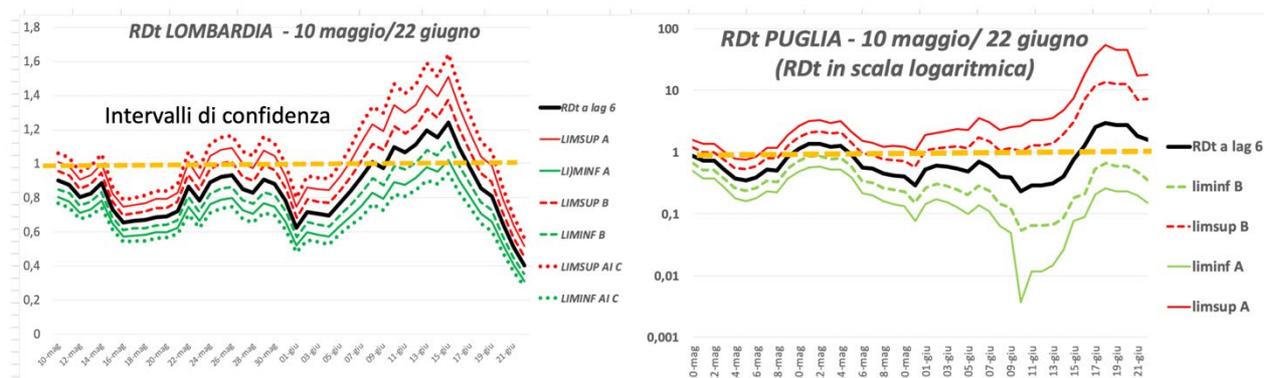


Figura 8: Andamento dell' $RD_t$  in Lombardia e in Puglia per i mesi di maggio e giugno con gli intervalli di confidenza e di massima variabilità

Figure 8: Trends of the  $RD_t$  index in Lombardy and Apulia for May and June with confidence and maximum variability intervals

Gli intervalli di confidenza descritti per la Lombardia sono stati calcolati tramite varie formule che derivano da diverse ipotesi distributive dell'indice stesso. Riteniamo che non sia così essenziale decidere quale sia l'ipotesi più corretta dato che è necessario limitarsi ad un impiego descrittivo dell'indice stesso e quindi non lo si può utilizzare per fini rigidamente decisionali. Nel caso dell'andamento dell'indice in Puglia, invece, l'ampiezza dell'intervallo di confidenza è così rilevante da sottintendere che l'indice non può essere utilizzato neppure a fini descrittivi.

Per queste ragioni crediamo sia opportuno sottolineare che l'analisi dell'andamento di questi indici deve accompagnarsi sempre all'analisi dell'andamento delle frequenze giornaliere<sup>(11)</sup>. Possiamo simbolicamente dire che l'andamento delle frequenze giornaliere evidenzia la velocità di sviluppo dell'epidemia e  $RD_t$  invece ne mostra l'accelerazione, positiva o negativa a seconda che sia maggiore o minore dell'unità.

I due grafici riportati nella Figura 9 costituiscono un esempio di come si possano rappresentare congiuntamente le due dimensioni giorno per giorno in quanto riportano in ordinata l' $RD_t$  e in ascissa l'incidenza giornaliera. I dati sono gli stessi di quelli prima utilizzati relativamente alle Regioni Lombardia e Puglia per i giorni da metà maggio a metà giugno.

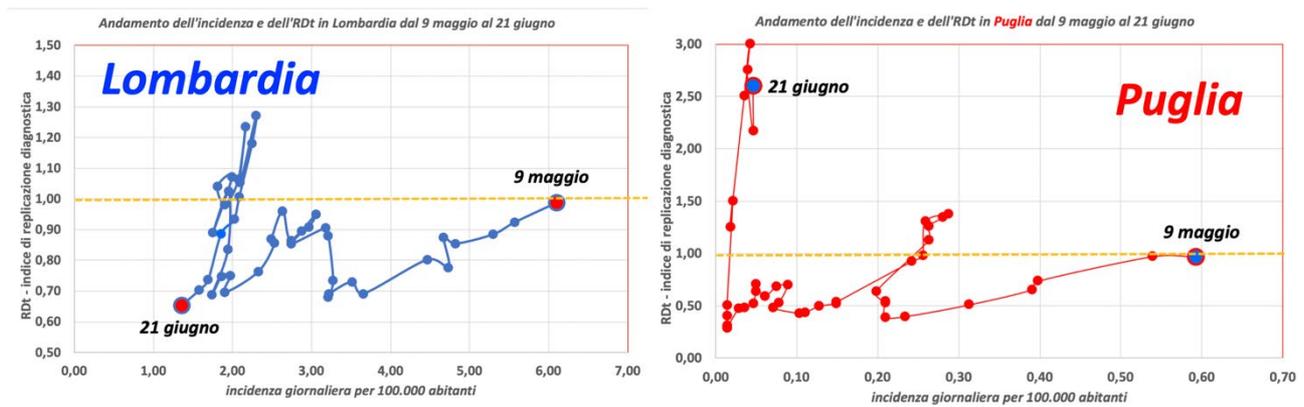


Figura 9: Valori dell' $RD_t$  in funzione dell'incidenza giornaliera per Lombardia e Puglia nel periodo dal 9 maggio al 21 giugno

Figure 9:  $RD_t$  values as a function of daily incidence for Lombardy and Apulia from 9 May to 21 June

Ogni punto corrisponde a un giorno e si nota come dal 9 maggio l'incidenza diminuisca non sempre con continuità sino ad arrivare a dei valori in cui la diminuzione si arresta (in Puglia perché l'incidenza praticamente si annulla). In queste situazioni l'indice  $RD_t$  assume valori anche elevati, che però hanno scarso significato epidemiologico date le basse frequenze su cui viene calcolato.

#### *Interpretazione degli indici come rischi relativi.*

Possiamo ritenere che, al di là dei diversi significati originariamente attribuiti agli indici di replicazione, nella loro sostanza possano essere interpretati come dei rischi relativi dato che le popolazioni nei due periodi considerati distanti un numero di giorni pari ai tempi di generazione del contagio sono praticamente le stesse. Gli indici sono quindi assimilabili a un rapporto tra incidenze al tempo  $t$  e al tempo  $t$  meno alcuni giorni.

L'indice  $R_t$  può essere letto nella sua formulazione concettuale come il rischio relativo di contagiarsi di una popolazione rispetto al rischio di alcuni giorni precedenti, e l'indice  $RD_t$  come il rischio di avere un esito positivo ad un tampone al tempo  $t$  rispetto allo stesso rischio ad alcuni lag precedenti.

La differenza tra i due indici infatti non consiste tanto nel metodo di calcolo, che in situazioni normali porta a risultati praticamente coincidenti, quanto nell'utilizzo di informazioni differenti, appunto l'inizio sintomi e l'esito tamponi.

Se l'intervallo tra inizio sintomi e esito tampone rimane costante tra aree e/o nel tempo i due indici forniscono praticamente la stessa informazione. Le differenze possono derivare dall'ampiezza della finestra di giorni su cui vengono calcolati o dalla possibile presenza di dati giornalieri anomali, in senso sia positivo che negativo, che l'indice  $RD_t$  tende a smorzare attraverso l'utilizzo di medie mobili settimanali mentre l'indice  $R_t$  potrebbe enfatizzare se assegna loro come peso un valore elevato della distribuzione Gamma.

#### *Praticabilità di calcolo e di comprensione*

Un'altra differenza tra i due indici è la diversa praticabilità di calcolo; per l' $R_t$  serve comunque uno strumento informatico dedicato alle analisi statistiche, come ad esempio il software R, e con delle routine già predisposte per il suo calcolo. E' difficile ipotizzare che l'indice  $R_t$  possa essere calcolato da chi non abbia nozioni sufficienti di Statistica e non sia in grado di utilizzare gli strumenti adatti.

L'indice  $RD_t$  invece è basato su calcoli e nozioni elementari e può essere ricavato facilmente tramite un foglio di calcolo come ad esempio quello disponibile in Excel.

A tale facilità corrisponde inoltre anche l'immediatezza della comprensione, caratteristica importante che può permettere ai decisori di fare affidamento su di un indice di cui hanno compreso sino in fondo le modalità di calcolo.

## **CONCLUSIONI**

Possiamo concludere quindi affermando che i due indici presentati sono differenti, ma entrambi validi e, pur mostrando aspetti differenti, forniscono fondamentalmente informazioni del tutto simili al decisore di Sanità pubblica.

E' comunque importante chiamarli con il loro nome,  $R_t$  indice di riproduzione di base,  $RD_t$  indice di replicazione diagnostica, e considerare che le date di monitoraggio a cui si riferiscono sono date differenti anche se relative agli stessi individui. Non si riferiscono alle date di contagio, ma l' $R_t$  ad un evento successivo di circa sei giorni nel caso del coronavirus e l' $RD_t$  ad un momento più variabile, quello della data dell'esito del test molecolare, che dipende molto non tanto dai diversi comportamenti del virus quanto dalle modalità diagnostiche dei servizi in fasi differenti dell'epidemia.

Più che le diversità dei metodi di calcolo risultano importanti le caratteristiche dei dati sia per la loro completezza sia per la loro accuratezza. In questo senso si consideri che la data dell'esito del tampone solitamente corrisponde alla realtà, mentre la data di inizio sintomi risente sia della minore importanza data alla sua registrazione da parte dell'operatore, sia dei problemi di *recall bias* del malato che può interpretare molto diversamente sintomi precedenti anche di diversi giorni come attinenti o meno alla sua patologia.

In conclusione non riteniamo necessario assegnare un giudizio di preferenza a uno dei due indici, ma solamente invitare l'utilizzatore a capirne le diverse potenzialità in modo da poter scegliere quello che ritiene più opportuno rispetto ai fini per i quali lo si voglia impiegare.

Conflitti di interesse dichiarati: nessuno

## **BIBLIOGRAFIA**

- (1) Cori A., Ferguson N.M., Fraser C. and Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers during epidemics. *American Journal of Epidemiology* 2013, 178 (9):1505–1512
- (2) Cislighi C.,  $R_0$ ,  $R_t$ ,  $RD_t$  che sono e che significano? <https://www.epiprev.it/r0-rt-rdt-che-sono-e-che-significano>
- (3) [https://www.iss.it/comunicati-stampa/-/asset\\_publisher/fjTKmjgSgdK/content/id/5398713](https://www.iss.it/comunicati-stampa/-/asset_publisher/fjTKmjgSgdK/content/id/5398713)

- (4) Smith D.L., Battle K. E., Hay S. I., Barker C.M., Scott T. W., McKenzie F. E. Ross, Macdonald and a Theory for the Dynamics and control of Mosquito-Transmitted Pathogens. PLOS Pathogens 2012, 8 (4): e100258
- (5) Fraser C. Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. PLoS One 2007, 2(8): e758
- (6) Wallinga J. and Teunis P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. American Journal of Epidemiology 2004, 160 (6): 509–516
- (7) Cintrón-Arias A., Castillo-Chávez C., Bettencourt L.M.A., Lloyd A.L. and Banks A.T. The Estimation of the Effective Reproduction Number from Disease Outbreak Data. Math Biosci Eng 2009, 6(2):261-82
- (8) Cereda D., Tirani M., Rovida F. et al. The early phase of the COVID-19 outbreak in Lombardy, Italy, arXiv:2003.09320v1 [q-bio.PE]
- (9) R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- (10) <https://cran.r-project.org/web/packages/EpiEstim/index.html>
- (11) Adam D. A guide to *R* – the pandemic misunderstood metric. Nature 2020, 583: 346-348