

L'indice Rt è inaffidabile

Sfortunatamente la discesa è irrilevante, anzi è sintomo preoccupante delle crepe del monitoraggio

Giorgio Parisi Fisico, presidente dell'Accademia Nazionale dei Lincei

In tutta Italia si stanno commentando i dati recentemente diffusi sul fattore di riproduzione Rt. Si legge che il fattore Rt del Paese è sceso, con grande soddisfazione generale, da 1,72 a 1,43. Sfortunatamente però il calcolo di Rt in queste ultime settimane è diventato sempre più inaffidabile e la discesa dell'Rt stimato è completamente irrilevante, anzi questa discesa è un sintomo preoccupante delle crepe che si stanno aprendo sul sistema di monitoraggio: non è un segnale che da solo autorizzi ottimismo.

Questa non affidabilità del valore Rt è chiaramente scritta nei documenti ufficiali. Se la settimana scorsa avevamo un Rt che con il 95% di probabilità era compreso tra 1,45 e 1,83, questa settimana il valore è compreso tra 1,08 e 1,81. Esattamente come per gli exit poll, non c'è certezza, e i dati sono comunicati sotto forma di "forchetta", come hanno ben imparato i giornalisti che presentano i risultati la sera delle elezioni. Sfortunatamente nella comunicazione giornalistica dei dati epidemiologici le forchette spariscono e rimangono solo nei comunicati ufficiali.

In una settimana l'ampiezza della forchetta si è raddoppiata, e questo aumento è estremamente preoccupante. La sintesi dell'ultimo comunicato è che l'incremento del numero dei contagiati giornalieri potrebbe o essere rimasto quasi costante o essere quasi raddoppiato o aver fatto qualcosa di intermedio: conclusioni drammaticamente incerte, sulle quali sono perfettamente d'accordo.

Se non sono affidabili i dati, cioè i numeri che vengono usati per calcolare Rt, il suo valore stimato non è affidabile. Come siamo arrivati a questa situazione? Ad aprile la procedura scelta per il calcolo di Rt era perfettamente adeguata alla situazione sanitaria dell'epoca ed era basata sull'assunzione che i dati provenienti dal monitoraggio nazionale fossero affidabili. Purtroppo dal 20 ottobre non è più così. Per motivi che non è facile identificare con precisione, ma probabilmente connessi all'eccessivo numero di casi, il sistema di monitoraggio non riesce a dare un'informazione precisa sugli sviluppi dell'epidemia.

Il sistema di calcolo di Rt si basa sul numero di persone che sviluppano sintomi, ma se - per inefficienze di varia natura - questo numero non è corretto, il valore stimato di Rt diventa anch'esso un numero non corretto. Questo fatto è ben noto a tutte le persone esperte di statistica che hanno provato a calcolare da sé l'evoluzione di Rt a partire dai dati ufficiali dei vari Paesi. Il raddoppio dell'ampiezza della forchetta sull'Rt del paese nell'arco di una sola settimana ben riflette questa situazione.

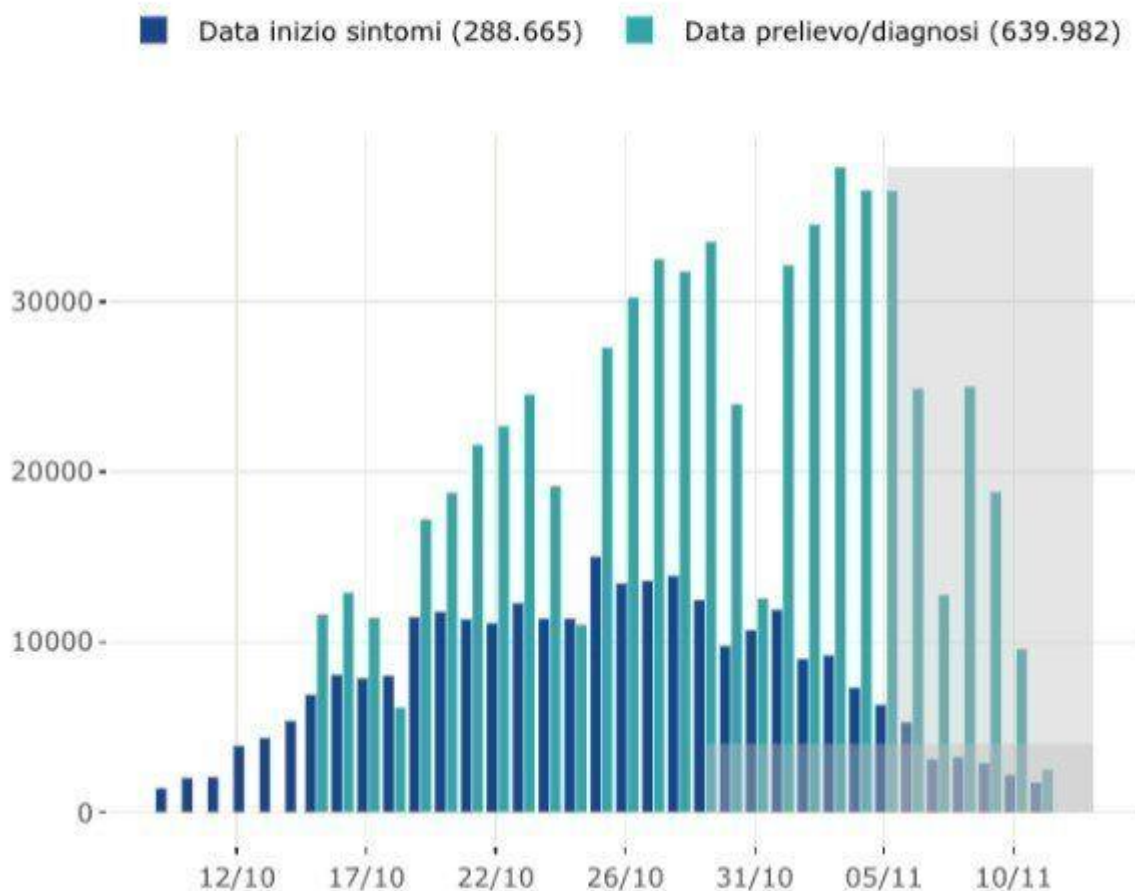
Gli indicatori della crescita epidemica sono numerosissimi: il numero di persone che sviluppano i sintomi, il numero di persone che risulta positivo, la percentuale di persone che risultano positive, il loro stato di salute al momento del tampone, la distribuzione di età, le chiamate ai pronto soccorso, il numero delle persone che entrano ed escono dagli ospedali e dalle terapie intensive, il numero dei decessi, ecc... Ciascuno di questi indicatori ci dà informazioni sulla crescita epidemica da un punto di vista differente. Nel migliore dei mondi, dovrebbero essere coerenti tra loro. Se questa coerenza cessa di esistere, basarsi su uno solo indicatore, quello che sei mesi fa era il più affidabile, risulta del tutto arbitrario.

È quindi assolutamente necessario usare in maniera combinata i vari indicatori per riuscire ad avere delle stime di Rt affidabili. Sono certo che gli addetti ai lavori se ne rendono perfettamente conto. Ma non è facile farlo: serve uno sforzo scientifico coordinato per capire quale sia il metodo più affidabile. Ma il consenso scientifico può essere raggiunto solo attraverso discussioni pubbliche e questo può avvenire solo se tutti questi dati diventano pubblici. Si parla tanto di intelligenza artificiale, di reti neurali, di big-data, ma i dati che sono disponibili sull'epidemia sono dei micro-data. Come ha richiamato chiaramente la commissione COVID-19 dell'Accademia dei Lincei, in assenza di trasparenza, ogni conclusione diviene contestabile sul piano scientifico e, quindi, anche sul piano politico. Che senso ha decidere l'apertura o la chiusura delle Regioni basandosi su un numero non affidabile, con un'incertezza enorme?

Ma dove si vede che il 20 ottobre i dati per il numero di sintomatici cessano di essere affidabili? Basta guardare con occhio di esperto i grafici disponibili su www.epicentro.it per gli ultimi 30 giorni.

Curva epidemica dei casi di COVID-19 diagnosticati in Italia negli ultimi 30 giorni

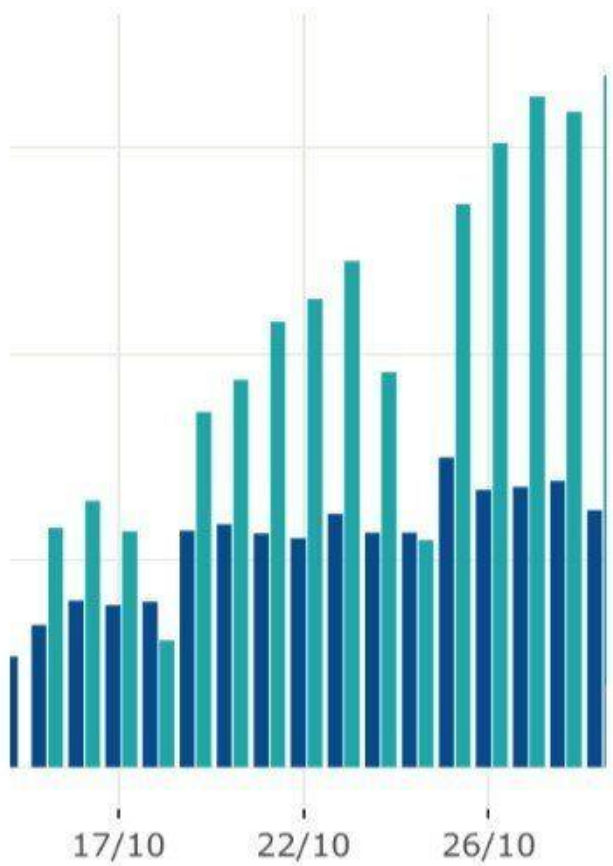
Nota: il numero dei casi riportato negli ultimi giorni (riquadri grigi) deve essere considerato provvisorio sia per possibili ritardi di segnalazione che di diagnosi.



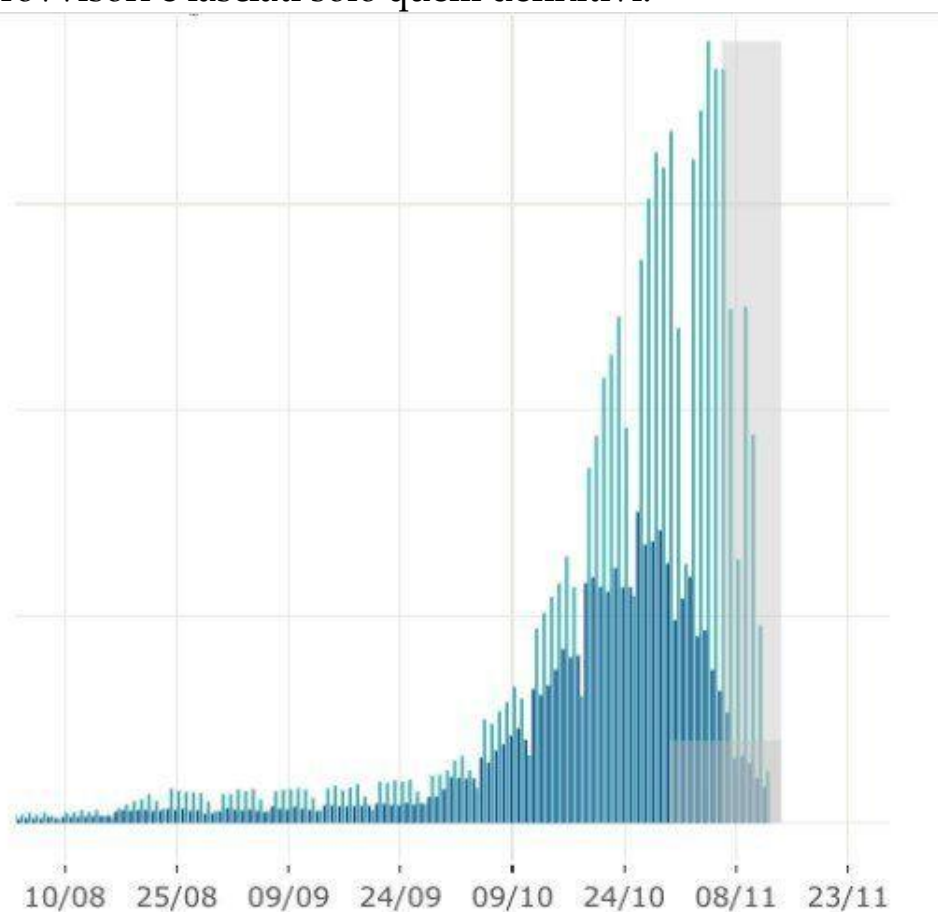
Curva epidemica in Italia negli ultimi 30 giorni

Qui sopra c'è il grafico estratto dalla dashboard del 13 novembre.

Il verde rappresenta il numero di casi positivi registrati in ciascun giorno mentre il blu rappresenta il numero di persone positive che hanno accusato i primi sintomi in quella data. I dati per il numero di casi positivi (verde) prima del 13 novembre non sono indicati. Come è detto chiaramente sul sito, i dati sui positivi successivi al 5 Novembre (verdi) e sui sintomatici (blu) dopo il 30 novembre sono provvisori. Quindi bisogna evitare di guardare i dati dopo queste due date.



Questo è un ingrandimento della figura precedente dove sono stati levati i dati provvisori e lasciati solo quelli definitivi.



Qui ci sono gli stessi dati delle figure precedenti per un periodo più ampio.

Come si vede ad occhio, (i dati del grafico non sono scaricabili!) il rapporto tra casi sintomatici e casi positivi è circa costante al variare del tempo: i sintomatici salgono nello stesso modo dei positivi. Tuttavia, nella settimana del 19 ottobre i sintomatici cessano di crescere nonostante il grande aumento dei positivi e anche nella settimana successiva questa differenza (in proporzione) diventa sempre più grande. Il rallentamento - fin quasi ad arrestarsi - della crescita dei sintomatici a partire dal 20 ottobre non ci dice niente sulla crescita dell'epidemia e non ha senso usarlo per dedurre un valore più basso di R_t negli ultimi giorni: ci dice solo che si è deteriorato il flusso di informazioni a questo riguardo, che partendo dalla periferia arrivano al centro.

Per capire meglio dove il flusso si sia bloccato, bisognerebbe fare un'analisi Regione per Regione, ma questo compito è estremamente difficile in quanto i dati regionali non sono scaricabili. Guardando solamente i dati nazionali, risulta impossibile avere una valutazione affidabile di R_t .

Vorrei chiudere con una nota di molto cauto ottimismo: è assolutamente vero che, indipendentemente da R_t , ci sono vari segnali che potrebbero indicare un rallentamento della crescita epidemica, anche se è difficile arrivare a una conclusione quantitativa.

- Il 14 novembre, per la prima volta, non solo la media settimanale dei nuovi positivi è diminuita, ma anche il numero di nuovi positivi è inferiore al numero registrato la settimana scorsa.
- Il numero di persone in ospedale e nelle sale di rianimazione sta aumentando con un ritmo rallentato - dato che sarebbe più chiaro se fossero disponibili di dati separati delle entrate e delle uscite.

Questi segnali sono ambigui in quanto è difficile estrarre l'effetto di saturazione delle risorse da un vero rallentamento. Fortunatamente ci sono anche segnali meno ambigui che però potrebbero essere una semplice fluttuazione statistica:

- Le chiamate al 118 della zona milanese per patologie respiratorie sono rimaste costanti sulle 500 al giorno da una decina di giorni, ma da due giorni sono scese sui 400. Se non si tratta di una fluttuazione statistica, è una tendenza positiva molto interessante.
- Il numero di decessi che precedentemente raddoppiava ogni settimana, sta aumentando più lentamente, circa una volta e mezzo ogni settimana, e questa tendenza si sta rafforzando.

Senza un sfera di cristallo e senza l'aiuto del polpo Paul, è difficile sapere se nei prossimi giorni la discesa della media settimanale dei casi diventerà sempre più decisa e se il numero dei morti continuerà a seguire la curva dei casi. Bisogna anche dire che questi segnali forse positivi sono in parte controbilanciati da un segnale decisamente negativo: l'aumento dell'età media dei contagiati, pessimo segnale che indica che le infezioni di molti giovani sfuggono al sistema di monitoraggio.

Bisogna mettere assieme tutti queste informazioni per avere un quadro coerente della situazione. Dobbiamo inventarci una nuova metodologia per stimare accuratamente R_t , metodologia che serve non per fare previsioni del futuro, ma per capire lo sviluppo dell'epidemia e determinare l'effetto delle varie misure contenimento già prese. Questa informazione è fondamentale per poter uscire il più velocemente dal picco epidemico con i minori danni possibili. Questo può essere fatto solo coinvolgendo tutta la comunità scientifica partendo dalla pubblicazione tempestiva di tutti i dati rilevanti in formato digitale.

Articolo pubblicato il 15 novembre 2020 su
<https://www.huffingtonpost.it/author/giorgio-parisi/>