

Stime della trasmissibilità di SARS-CoV-2 in Italia

Giorgio Guzzetta e Stefano Merler - Fondazione Bruno Kessler, Trento

7 dicembre 2020

La stima della trasmissibilità di SARS-CoV-2 viene effettuata con un metodo statistico consolidato [1], il più usato nel mondo per valutare la trasmissibilità del virus. In particolare, la Fondazione Bruno Kessler (FBK) utilizza questo metodo fin da gennaio 2020 per stimare la trasmissibilità in Cina e, da febbraio 2020, per stimare la trasmissibilità in Italia. Il motivo alla base di questa scelta è che il metodo [1] si basa su una semplice statistica sui dati delle nuove infezioni giornaliere. Non richiede in particolare nessuna ipotesi sui meccanismi di trasmissione del virus, ancora incerti su alcuni aspetti, come invece richiedono le stime basate su tecniche di modellistica matematica. Non richiede inoltre di assumere crescita esponenziale dell'epidemia, come richiesto da altre tecniche basate sull'analisi del tasso di crescita dei casi. Infatti, durante epidemie controllate (come quella di COVID-19) l'ipotesi di crescita esponenziale è falsa.

L'8 giugno 2020, nel corso di una teleconferenza con tutte le Regioni, è stato spiegato questo metodo, approfondendone gli aspetti positivi e i punti deboli. Le slide presentate alle Regioni sono adesso disponibili come appendice a questo documento. Si è anche reso esplicito a tutte le Regioni che FBK avrebbe utilizzato questa tecnica di stima per calcolare la trasmissibilità di SARS-CoV-2 in tutte le Regioni italiane, a supporto del monitoraggio regionale.

I dati utilizzati per le stime

Per le stime si utilizzano le curve epidemiche dei casi sintomatici organizzati per data inizio sintomi (Figura 1). La definizione di caso sintomatico nel sistema di sorveglianza integrata include tutti i casi a cui sia associata una data inizio sintomi, a esclusione di quelli per cui all'individuo non sia mai stato associato uno stato clinico diverso da "asintomatico" o "guarito" durante il follow-up. Questa circostanza si verifica principalmente per le Regioni che raccolgono i dati di sorveglianza tramite sistemi informativi propri preesistenti alla piattaforma messa a disposizione dell'Istituto Superiore di Sanità (ISS). Alcuni di tali sistemi richiedono infatti l'inserimento obbligatorio di una data inizio sintomi anche per i casi asintomatici. La definizione restrittiva adottata serve ad avere una conferma che i casi considerati nella curva epidemica siano effettivamente casi sintomatici. Occasionalmente, la definizione può escludere casi effettivamente sintomatici in caso di errori di assegnazione dello stato clinico, andando ad aggiungersi ai casi non notificati dai sistemi sanitari, ma l'osservazione dei dati individuali suggerisce che questa sia un'eventualità rara. Il metodo per la stima di $R(t)$, comunque, è robusto rispetto a tassi costanti di sottotifica e non c'è ragione di ipotizzare che la frazione di casi sintomatici eventualmente persi per via della definizione adottata possa cambiare nel tempo.

Tale scelta per la definizione di caso sintomatico è stata condivisa con tutte le Regioni, in occasione della teleconferenza dell'8 giugno 2020.

L'acquisizione dei dati epidemiologici è affetta da una serie di ritardi, alcuni dei quali non comprimibili: in particolare, il tempo tra l'evento infettivo e lo sviluppo dei sintomi (tempo di incubazione), quello tra i sintomi e l'esecuzione del tampone, quello tra l'esecuzione del tampone e la conferma di positività, e quello tra la conferma di positività e l'inserimento nel

sistema di sorveglianza integrata ISS. Il ritardo complessivo tra infezioni e loro rilevamento nel sistema di sorveglianza è valutato e aggiornato settimanalmente analizzando la stabilità del numero di casi (sintomatici o ospedalizzati) riportato a ciascuna data. Su queste valutazioni si basa la scelta della data più recente alla quale si possono considerare sufficientemente stabili le varie stime di $R(t)$. Si noti che i possibili rallentamenti nell'effettuazione e analisi dei tamponi, conseguenti all'aumentata incidenza di infezione, impattano allo stesso modo tanto i conteggi aggregati di nuovi positivi riportati quotidianamente dal Dipartimento della Protezione Civile quanto i dati contenuti nel sistema di sorveglianza integrata.

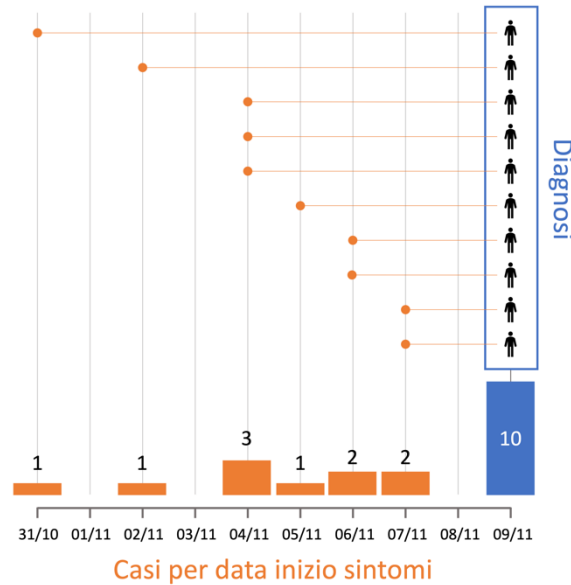


Figura 1. Costruzione della serie temporale dei casi sintomatici per data di inizio sintomi

Stima della trasmissibilità giornaliera da casi sintomatici, $G(t)$

Il metodo [1] fornisce una stima della distribuzione a posteriori della trasmissibilità giornaliera $G(t)$ basata su un algoritmo Markov Chain Monte Carlo (MCMC) applicato alla seguente funzione di verosimiglianza:

$$\mathcal{L} = \prod_{t=1}^T P\left(C(t) - I(t); G(t) \sum_s l(s)C(t-s)\right)$$

dove, oltre a $G(t)$:

- $P(k; \lambda)$ è la densità di una distribuzione Poisson, ovvero la probabilità di osservare k eventi se questi avvengono a una frequenza media λ ; $\mathcal{L}$ permette di confrontare a ogni tempo t il numero di casi stimati con l'equazione di rinnovamento (renewal equation [2]), $G(t) \sum_s l(s)C(t-s)$, con i casi trasmessi localmente, $C(t) - I(t)$.
- $C(t)$ è il numero di casi sintomatici con data di inizio sintomi al giorno t , con $t=0, \dots, T$;

Poiché la trasmissibilità giornaliera $G(t)$ può essere influenzata da eventi particolari (che possono risultare in valori di trasmissibilità anormalmente alti o bassi), si considera più rappresentativa della situazione al tempo t il valore medio dell'ultima settimana (7 giorni, corrispondenti a circa un tempo di generazione). L'algoritmo sopra descritto può essere utilizzato per stimare la distribuzione a posteriori delle variabili casuali

$$R(t) = 1/7 \sum_{s=0}^6 G(t-s).$$

Si veda ad esempio [4,5] per stime di $R(t)$ nelle diverse Regioni e Province italiane. Come per la trasmissibilità giornaliera $G(t)$, è possibile stimare la distribuzione a posteriori di $R(t)$, che chiamiamo stima puntuale per distinguerla da un valore medio descritto più avanti, da cui è in particolare possibile calcolare media e intervalli di credibilità al 95% (Figura 3). I valori medi e gli intervalli di credibilità di $R(t)$ sono quelli utilizzati dalla cabina di regia per il monitoraggio regionale.

Nota tecnica

La distribuzione di $R(t)$ viene campionata utilizzando tutti i valori della catena di Markov, cioè i valori di $G(t)$ campionati durante il processo di stima con MCMC. Questi valori servono direttamente per stimare la distribuzione a posteriori di $G(t)$, come visto sopra. Inoltre, ad ogni passo di MCMC, si può definire:

$$R_i(t) = 1/7 \sum_{s=0}^6 G_i(t-s)$$

per $i=1, \dots, N$

Questi N valori possono essere utilizzati per stimare la distribuzione a posteriori $R(t)$. La media, in particolare, è definita come:

$$\bar{R}(t) = 1/N \sum_{i=1}^N R_i(t)$$

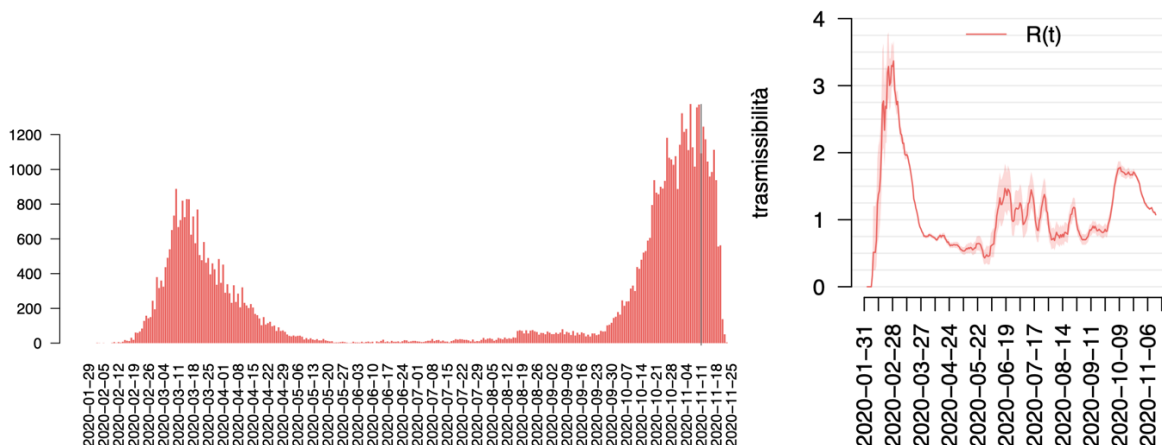


Figura 3. Sinistra: serie temporale del numero di casi $C(t)$ per data di inizio sintomi in Emilia-Romagna. Destra: distribuzione della trasmissibilità puntuale $R(t)$ (linea continua: media; area ombreggiata: 95%CI). I picchi di trasmissibilità legati alla prima onda (febbraio-marzo 2020), i picchi estivi associati a piccoli focolai, come quelli osservati nel settore della distribuzione, il picco di Ferragosto, e la seconda onda (ottobre-novembre 2020), sono molto meglio identificabili rispetto alle stime giornaliere fornite da $G(t)$.

Stima della trasmissibilità media da casi sintomatici, $\bar{M}(t)$

È inoltre utile calcolare la variabilità dei valori medi $\bar{R}(t)$ per capire ad esempio se le stime sono stabili nel tempo, oppure sono caratterizzate da qualche *trend* di crescita o decrescita, oppure fluttuano in un intervallo molto ampio (ad esempio perché il numero di casi $C(t)$ è basso e c'è alta incertezza statistica). A tale scopo si considera il valore:

$$\bar{M}(t) = \frac{1}{14} \sum_{s=0}^{13} \bar{R}(t-s)$$

dove $\bar{R}(t)$ è il valore medio della distribuzione $R(t)$, al tempo t . Questa stima è riportata nei bollettini ISS. Oltre a $\bar{M}(t)$, i bollettini ISS riportano il minimo e il massimo sui 14 giorni del 95% percentile delle distribuzioni $R(t)$, ovvero

$$M_{min}(t) = \min\{R_{0.025}(t-13), R_{0.025}(t-12), \dots, R_{0.025}(t)\}$$

e

$$M_{max}(t) = \max\{R_{0.975}(t-13), R_{0.975}(t-12), \dots, R_{0.975}(t)\},$$

dove $R_q(t)$ indica il q -quantile della distribuzione (Figura 4).

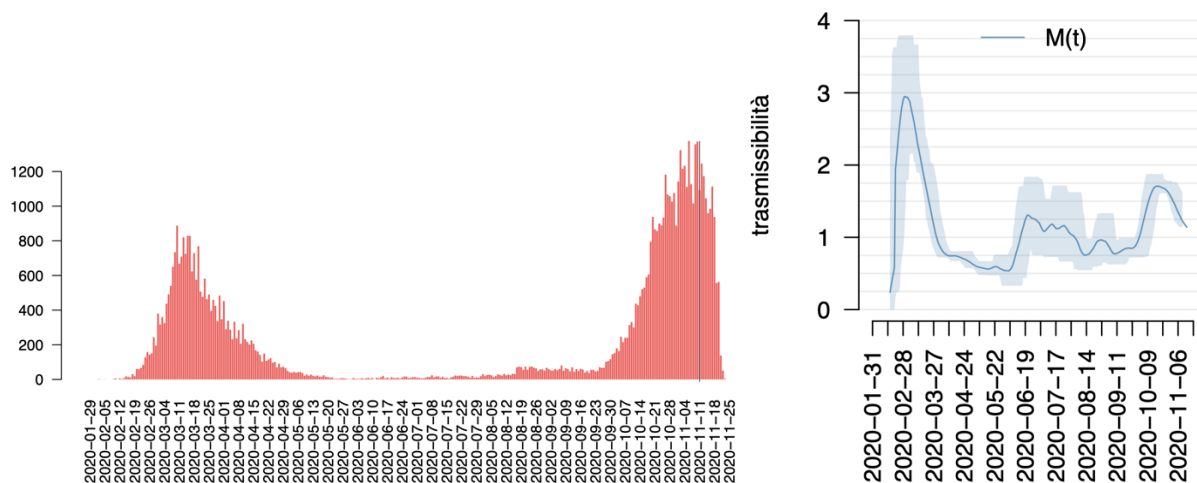


Figura 4. Sinistra: serie temporale del numero di casi $C(t)$ per data di inizio sintomi in Emilia-Romagna. Destra: distribuzione della trasmissibilità media $\bar{M}(t)$ (linea continua) e range (area ombreggiata). I picchi di trasmissibilità legati alla prima onda (febbraio-marzo 2020), i picchi estivi associati a piccoli focolai, come quelli osservati nel settore della distribuzione, il picco di Ferragosto, e la seconda onda (ottobre-novembre 2020), sono molto meglio identificabili rispetto alle stime fornite da $G(t)$. I range riportati danno conto sia

dell'alta variabilità delle stime sui focolai estivi, che dei periodi in cui la trasmissibilità è variata significativamente (es. fase di crescita e decrescita sia nella prima che nella seconda onda).

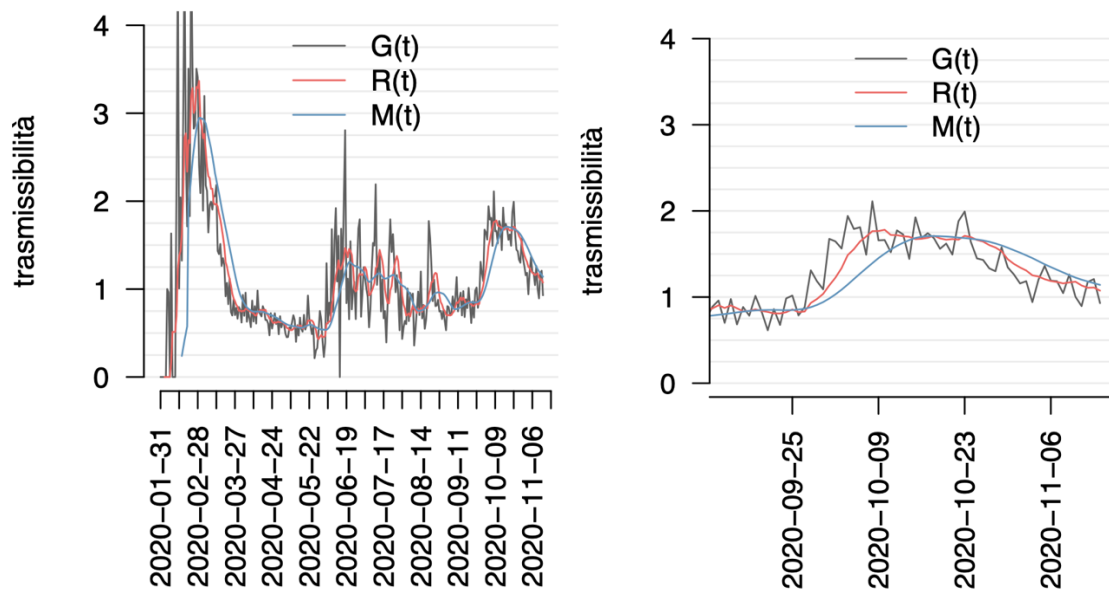


Figura 5. Confronto tra i valori medi $\bar{G}(t)$, $\bar{R}(t)$ e $\bar{M}(t)$ in Emilia-Romagna da inizio epidemia (sinistra) e nel periodo più recente (destra).

Stima della trasmissibilità puntuale da casi ospedalizzati

La stima puntuale da casi ospedalizzati, che chiamiamo $R^{\text{hosp}}(t)$, viene prodotta con gli stessi metodi descritti sopra solo che $C(t)$ rappresenta in questo caso il numero di nuovi casi ospedalizzati al tempo t , cioè si considera il numero di casi ospedalizzati per data di ospedalizzazione, traslati indietro di 3 giorni per tenere conto del ritardo medio tra data inizio sintomi e data di ospedalizzazione (Figura 6).

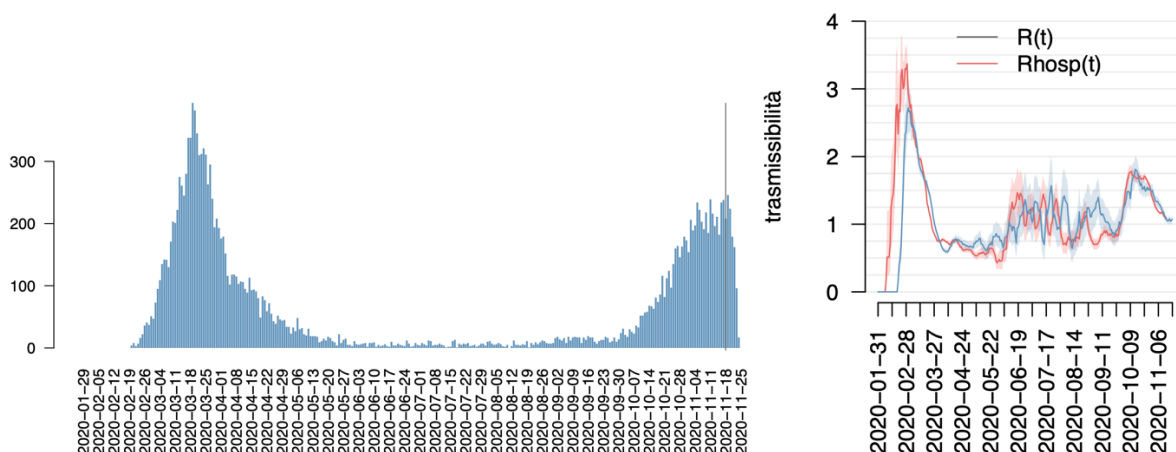


Figura 6. Sinistra: serie temporale del numero di casi ospedalizzati $C(t)$ per data di ospedalizzazione in Emilia-Romagna. Destra: confronto tra la distribuzione della trasmissibilità puntuale $R(t)$ (rosso) e la distribuzione della trasmissibilità puntuale da ospedalizzazioni $R^{\text{hosp}}(t)$ (blu). Le linee continue rappresentano i valori medi, le aree

ombreggiate rappresentano i 95%CI. È possibile osservare che nei periodi caratterizzati da alta incidenza (es. prima e seconda onda) le stime delle distribuzioni $R(t)$ e $R^{hosp}(t)$ sono molto coerenti. Questa coerenza diminuisce sostanzialmente durante il periodo estivo caratterizzato da bassa incidenza.

Stime della trasmissibilità puntuale con tecniche di data imputation

Come detto sopra, i dati sulle infezioni sintomatiche e sulle ospedalizzazioni vengono comunicati con un certo ritardo (che non può essere compresso più di tanto). Inoltre, questo ritardo può variare nel tempo, in quanto dipende dalla possibile saturazione dei sistemi di prelievo e analisi dei tamponi quando l'incidenza dell'infezione cresce. A titolo di esempio, può essere inserito oggi nel database della sorveglianza ISS un nuovo caso sintomatico che però ha avuto la data di inizio sintomi 20 giorni fa.

Su base statistica, confrontando i database recenti della sorveglianza ISS su COVID-19 si determina la data, che chiamiamo D , a cui il database risulti sufficientemente completo, cioè non modificabile più di tanto da future introduzioni di casi di cui non si era a conoscenza. Sufficientemente completo significa completo almeno al 90%. Ciò significa che è possibile che nel futuro vengano inseriti ancora dei dati riguardanti casi che hanno avuto insorgenza dei sintomi prima di D , ma non più del 10%. Tutte le stime della trasmissibilità sono fornite alla cabina di regia per quella data D . Si noti che la data D può essere diversa tra stime della trasmissibilità da sintomatici o ospedalizzati. Chiaramente, le stime della trasmissibilità possono essere rese più incerte (di solito vengono fornite delle sottostime) da questi ritardi di inserimento del database della sorveglianza ISS. Se però l'incompletezza dei dati è uniforme nel tempo, la sottostima è quasi trascurabile (Figura 7).

Si è comunque utilizzata una nota tecnica di *data imputation* per fornire stime di $R(t)$ più ravvicinate nel tempo. La tecnica si basa su una stima del grado di completezza del database della sorveglianza ISS, ottenuta confrontando tutti i database giorno per giorno. Sia T il dato a cui il database è considerato completo (nella definizione data sopra). Su base statistica si calcola la percentuale di completezza per i tempi $T+1, \dots, T+K$, che chiamiamo c_i , con $i=1, \dots, K$. Questo significa che ci si aspetta che il numero di casi ai tempi $T+1, \dots, T+K$ aumenterà, per effetto di future aggiunte di casi, di un fattore $1/c_i$. Si definiscono quindi i nuovi casi ai tempi $T+1, \dots, T+K$ come $C(T+1)/c_1, \dots, C(T+K)/c_K$ per tener conto del futuro aumento dei casi.

A questo punto possono venir calcolate delle stime di $R(t)$, che chiamo stime puntuali aumentate, $R^{aug}(t)$, fino al tempo $T+K$. Ovviamente, bisogna sempre considerare che il livello di affidabilità di queste stime è molto minore di quelle fornite al tempo T .

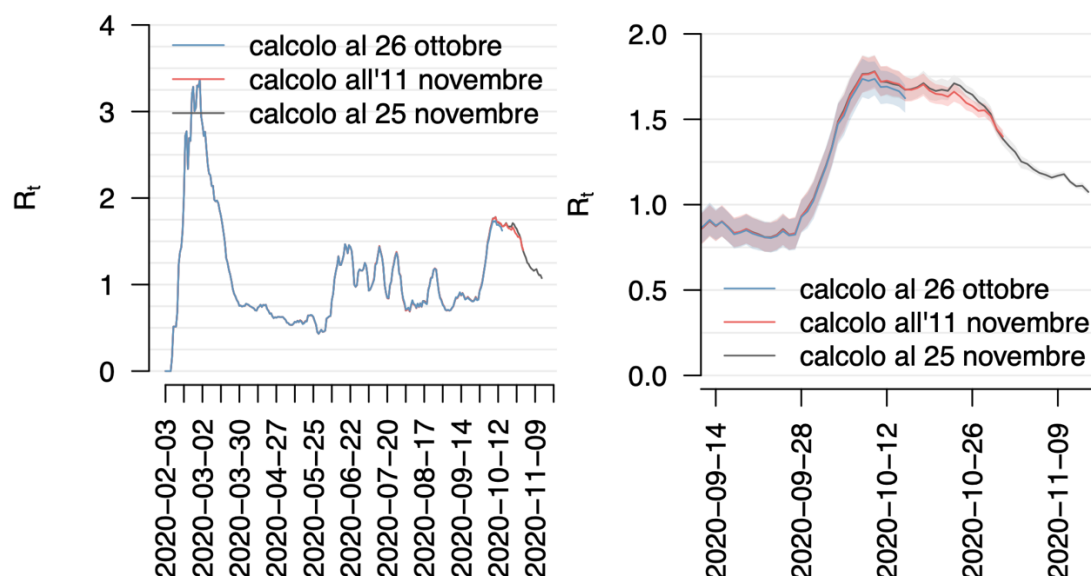


Figura 7. Confronto tra le stime della distribuzione di $R(t)$ in Emilia-Romagna utilizzando i dati del sistema di sorveglianza ISS a diverse date, da inizio epidemia (sinistra) e nel periodo più recente (destra). Le linee continue rappresentano i valori medi, le aree ombreggiate rappresentano i 95%CI. Si osserva che le stime fornite con il dato del sistema di sorveglianza ISS al 26 ottobre 2020 non sono modificate sostanzialmente dall'inserimento successivo di casi con data inizio sintomi risalente a prima del 26 ottobre. La stessa considerazione vale per le stime eseguite con il dato dell'11 novembre se confrontate con quelle effettuate col dato del 25 novembre. Ovviamente sono possibili scostamenti maggiori se le Regioni effettuano aggiornamenti retrospettivi del database più pesanti.

Disponibilità dei risultati

I risultati di queste diverse stime ($R(t)$, $\bar{M}(t)$, $R^{\text{aug}}(t)$, $R^{\text{hosp}}(t)$) della trasmissibilità di SARS-CoV-2 sono sempre stati resi disponibili alle Regioni nei rapporti settimanali della cabina di regia. La stima $\bar{M}(t)$ è disponibile nei bollettini della sorveglianza ISS a partire da aprile 2020. La stima della distribuzione $R(t)$ è disponibile sul sito del Ministero della Salute a partire dal 25 ottobre 2020.

Software per il calcolo della trasmissibilità

Per il calcolo di $R(t)$, FBK utilizza come codice di base un algoritmo scritto in codice C e disponibile all'indirizzo <https://github.com/majelli/Rt>, distribuito pubblicamente come informazione supplementare di [6]. In data 8 giugno 2020, è stato distribuito a tutte le Regioni uno script in codice R che permette il calcolo di $R(t)$ a partire dall'algoritmo EpiEstim (<https://cran.r-project.org/web/packages/EpiEstim/index.html>), pubblicamente disponibile. Lo script distribuito alle Regioni è adesso disponibile come appendice a questo documento (Figura 8).

Lo stesso software può essere utilizzato per il calcolo di $R^{\text{aug}}(t)$, posto di conoscere i coefficienti c_i per $i=1, \dots, K$.

Le stime di $\bar{M}(t)$ possono essere derivate facilmente da quelle di $R(t)$.

Lo stesso software può essere utilizzato per il calcolo di $R^{\text{hosp}}(t)$.

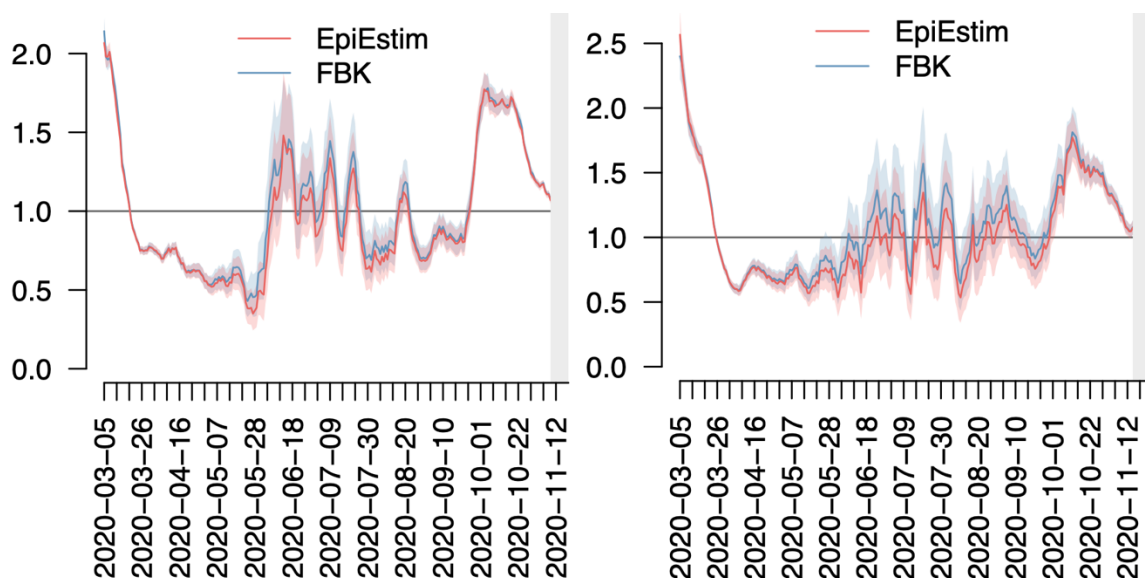


Figura 8. Confronto tra le stime della distribuzione della trasmissibilità da casi sintomatici $R(t)$ (sinistra) e da casi ospedalizzati $R^{\text{hosp}}(t)$ (destra) in Emilia-Romagna ottenute con l'algoritmo FBK (blu) e l'algoritmo EpiEstim distribuito alle Regioni (rosso). Le linee continue rappresentano i valori medi, le aree ombreggiate rappresentano i 95%CI. Le stime sono praticamente indistinguibili nei periodi caratterizzata da alta incidenza (prima e seconda onda). Piccole differenze sono invece osservabili nei picchi di trasmissione estiva. Bisogna comunque considerare che poiché il metodo di stima (MCMC) è stocastico, piccole differenze sono riscontrabili anche utilizzando lo stesso algoritmo con un seme iniziale diverso.

Proiezioni per le occupazioni di posti letto e terapia intensiva

Le proiezioni si basano sul metodo dell'equazione di rinnovo [2], che stima il numero di casi previsti sulla base del numero di casi osservati nei giorni precedenti, della distribuzione del tempo di generazione, e di ipotesi sull'andamento futuro del numero di riproduzione netto $R(t)$. Secondo l'equazione di rinnovo, il numero di nuovi casi sintomatici atteso al tempo t può essere stimato come:

$$C(t) = R(t) \sum_{s>0} C(t-s) \cdot l(s)$$

Per considerare la variabilità stocastica, la proiezione del numero di nuovi casi sintomatici a ogni giorno t viene campionato da una distribuzione di Poisson con frequenza media pari al valore di $C(t)$ dato dall'equazione di rinnovo. Il metodo di stima è robusto per proiezioni a breve termine e può essere usato per valutare le proiezioni con qualsiasi andamento ipotizzato di $R(t)$ nel futuro. Le proiezioni dei nuovi casi sintomatici fornite alle Regioni partono dai casi sintomatici alla data T e vengono calcolate fino ai 15 e 30 giorni successivi alla data di calcolo. $R(t)$ viene assunto costante su tutto il periodo di proiezione, e pari all'ultima stima disponibile di R^{hosp} . Il calcolo viene ripetuto $K=1000$ volte, campionando ogni volta un valore diverso della distribuzione a posteriori di R^{hosp} .

Per passare dalla proiezione del numero di casi sintomatici alle proiezioni di occupazione di posti letto, si assume che una certa frazione p di casi sintomatici venga ricoverata (in area medica o in terapia intensiva). Applicando al numero di casi sintomatici attesi la distribuzione del ritardo tra l'inizio dei sintomi e il ricovero, si ottiene una stima dell'evoluzione temporale degli ingressi in ospedale nell'ipotesi che tutti i casi sintomatici richiedessero un ricovero, $h(t)$. Applicando ad $h(t)$ la distribuzione della durata del ricovero, si ottiene la stima dell'evoluzione temporale delle uscite $q(t)$. La curva di occupazione effettiva dei posti letto $B(t)$ sarà data dalla differenza delle cumulate degli ingressi e delle uscite effettive nel tempo, ovvero:

$$B(t) = p \cdot \sum_{\tau=0}^t (h(\tau) - q(\tau))$$

Per l'occupazione di posti letto in area medica si usa una distribuzione del ritardo con media 2,7 giorni, mediana 1 giorno, intervallo interquartile (IQR) 0-3 giorni e una distribuzione di durata del ricovero con media 15,5 giorni, mediana 10 giorni e IQR 3-21 giorni. Per l'occupazione di posti letto in terapia intensiva si usa una distribuzione del ritardo con media 8,1 giorni, mediana 7 giorni, intervallo IQR 3-11 giorni e una distribuzione di durata del ricovero con media 14,6 giorni, mediana 11 giorni e IQR 6-19 giorni. Entrambe le distribuzioni sono basate su dati ospedalieri della Regione Lombardia su un campione ampio di ricoveri ($n > 40.000$ per l'area medica e $n > 3.900$ per le terapie intensive) [7].

I valori delle frazioni p di casi sintomatici osservati che vengono ricoverati in area medica (p_H) e in terapia intensiva (p_I) sono calcolati indipendentemente e settimanalmente per ciascuna Regione, per tenere conto dei diversi livelli di sottonotifica dei casi sintomatici rispetto al totale e dei criteri regionali di ammissione in ospedale e terapia intensiva. Il calcolo di p è fatto in maniera che la media sulle K iterazioni della curva stimata di $B(t)$ riproduca il dato medio di occupazione fornito dalla protezione civile negli ultimi Q giorni per cui il dato è disponibile. Q è definito per ogni Regione ed ha un valore diverso per le ospedalizzazioni in area medica (Q_H) e in terapia intensiva (Q_I); in entrambi i casi si considera il minimo valore di Q per cui la somma dei giorni-paziente nel dato superi 14. Per le Regioni che hanno pochi pazienti in terapia intensiva o in area medica, per cui risulti $Q_I > 60$ giorni o $Q_H > 5$ giorni, viene usato come valore di p_I o di p_H il valore medio dei coefficienti calcolati sulle Regioni rimanenti.

Nell'interpretare le proiezioni bisogna sempre tenere bene a mente le ipotesi su cui sono basate. In questo caso, le assunzioni importanti sono tre: $R^{\text{hosp}}(t)$ e i criteri di ammissione in area medica e terapia intensiva non cambiano nel periodo per cui si effettuano le proiezioni. Ovviamente, non c'è certezza che queste tre ipotesi siano vere per l'intero periodo delle proiezioni. $R^{\text{hosp}}(t)$ può cambiare a seguito di interventi di mitigazione o cambiamento del comportamento umano; i criteri di ammissione in area medica e terapia intensiva possono cambiare al variare dell'incidenza della malattia. Ad esempio, i criteri di ammissione possono diventare più restrittivi quando si va verso la saturazione dei posti in area medica o terapia intensiva.

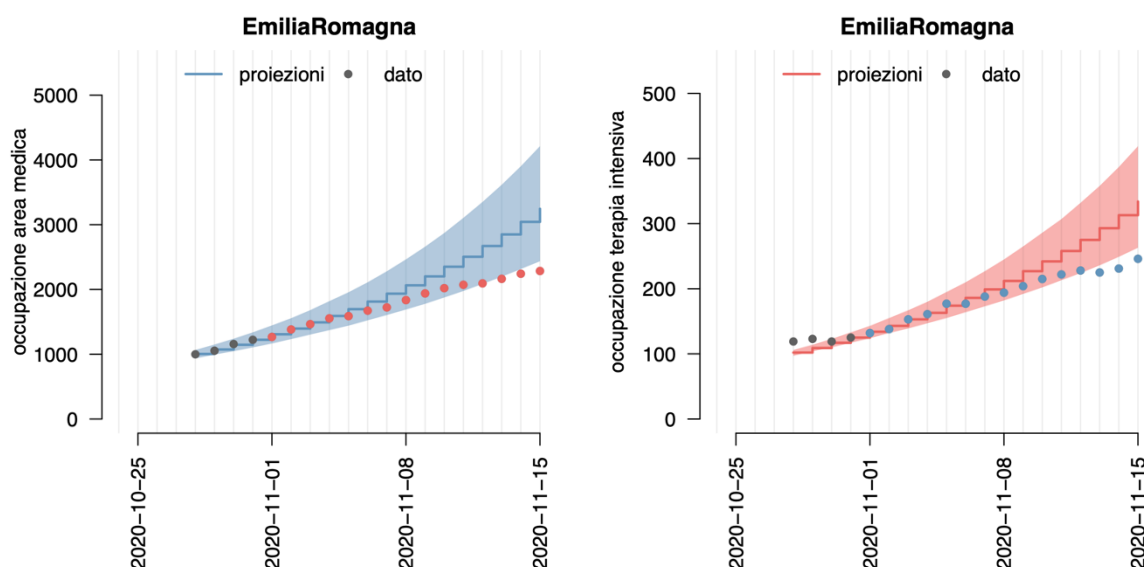


Figura 9. Sinistra: Confronto tra occupazione in area medica in Emilia-Romagna (punti) e proiezioni a 15 giorni con l'equazione di rinnovo [2] (linea continua: media; area ombreggiata: intervallo di predizione al 95%). Le proiezioni, effettuate il 31 ottobre, sono basate su una stima della trasmissibilità da ospedalizzazioni al 24 ottobre di $R^{hosp}(t)=1.49$ (95%CI: 1.39-1.59). Destra: come pannello di sinistra ma per l'occupazione di terapia intensiva. Le proiezioni risentono del sistematico calo di trasmissibilità $R^{hosp}(t)$, dovuto ai vari DPCM e ordinanze regionali, osservato durante tutto il periodo di proiezione. $R^{hosp}(t)$ ha infatti raggiunto valori vicini a 1, che significa che l'epidemia ha raggiunto il picco, intorno al 10 novembre.

Bibliografia

- [1] A Cori, NM Ferguson, C Fraser, S Cauchemez. **A new framework and software to estimate time-varying reproduction numbers during epidemics.** Am J Epidemiol, 178(9):1505-1512, 2013
- [2] J Wallinga, M Lipsitch. **How Generation Intervals Shape the Relationship between Growth Rates and Reproductive Numbers.** Proceeding of the Royal Society B, 274:1609:599-604, 2007.
- [3] D Cereda, M Tirani, F Rovida, V Demicheli, M Ajelli, P Poletti, F Trentini, G Guzzetta, V Marziano, A Barone, M Magoni, S Deandrea, G Diurno, M Lombardo, M Faccini, A Pan, R Bruno, E Pariani, G Grasselli, A Piatti, M Gramegna, F Baldanti, A Melegaro, S Merler. **The early phase of the COVID-19 outbreak in Lombardy, Italy.** ArXiv:2003.09320v1, 2020.
- [4] G Guzzetta, F Riccardo, V Marziano, P Poletti, F Trentini, A Bella, X Andrianou, M Del Manso, M Fabiani, S Bellino, S Boros, AM Urdiales, MF Vescio, A Piccioli, S Brusaferrò, G Rezza, P Pezzotti, M Ajelli, S Merler. **Impact of a Nationwide Lockdown on SARS-CoV-2 Transmissibility, Italy.** Emerging Infectious Disease, 27:1, 2020.

- [5] F Riccardo, M Ajelli, XD Andrianou, A Bella, M Del Manso, M Fabiani, S Bellino, S Boros, AM Urdiales, V Marziano, MC Rota, A Filia, FP D'Ancona, A Siddu, O Punzo, F Trentini, G Guzzetta, P Poletti, P Stefanelli, MR Castrucci, A Ciervo, C Di Benedetto, M Tallon, A Piccioli, S Brusaferrò, G Rezza, S Merler, P Pezzotti, for the COVID-19 working group. **Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020.** Eurosurveillance, 2020.
- [6] J Zhang, M Litvinova, W Wang, Y Wang, X Deng, X Chen, M Li, W Zheng, L Yi, X Chen, Q Wu, Y Liang, X Wang, J Yang, K Sun, IM Longini, ME Halloran, P Wu, B Cowling, S Merler, C Viboud, A Vespignani, M Ajelli, H Yu. **Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study.** The Lancet Infectious Diseases, 20:793-802, 2020.
- [7] F Trentini, V Marziano, G Guzzetta, M Tirani, D Cereda, P Poletti, R Piccarreta, A Barone, G Preziosi, F Arduini, PG Della Valle, A Zanella, F Grosso, G del Castillo, A Castrofino, G Grasselli, A Melegaro, A Piatti, A Andreassi, M Gramegna, M Ajelli, S Merler. **Healthcare strain and intensive care during the COVID-19 outbreak in the Lombardy region: a retrospective observational study on 43,538 hospitalized patients.** Medrxiv preprint. <https://doi.org/10.1101/2020.11.06.20149690>