

Guida alla lettura e all'interpretazione dei dati COVID-19

Fiammetta Menchetti, Silvia Noirjean *

Dipartimento di Statistica, Informatica, Applicazioni “Giuseppe Parenti”

Università degli Studi di Firenze

10 aprile 2020

1 Oggetto

Obiettivo di questa breve guida è quello di fornire uno strumento utile a comprendere i dati disponibili sulla pandemia COVID-19 in Italia e il loro contenuto informativo, fruibile da chiunque ne abbia necessità. Infatti, prima di iniziare qualsiasi tipo di analisi occorre comprendere i limiti dei dati di cui disponiamo. Affronteremo innanzitutto alcune problematiche relative alla fonte dei dati sulla diffusione della pandemia in Italia. Prenderemo poi in esame due grandezze di maggiore interesse (contagi e decessi), mettendo in luce i limiti fondamentali delle loro rilevazioni che dipendono dalle politiche scelte per il monitoraggio dell'epidemia stessa. Infine, ove possibile, proponiamo correzioni o chiavi di lettura che possano aiutare alla comprensione del fenomeno.

2 Fonti

2.1 Protezione Civile

La principale fonte di dati riguardanti la diffusione della pandemia da COVID-19 in Italia è la Protezione Civile, che giornalmente a partire dal 24 febbraio pubblica i relativi dati sulla repository GitHub¹.

- Ricoverati con sintomi (R_s): numero di soggetti ricoverati non in terapia intensiva.
- Terapia intensiva (R_i): numero di ricoverati in terapia intensiva.
- Totale ospedalizzati (T_o): totale dei soggetti ricoverati in ospedale, $T_o = R_s + R_i$.
- Isolamento domiciliare (I_d): totale dei positivi in isolamento presso le proprie abitazioni.
- Variazione totale positivi (ΔT_p): variazione del totale positivi rispetto al giorno precedente, $\Delta T_p = T_{p_n} - T_{p_{n-1}}$.
- Dimessi guariti (G): persone dimesse guarite.
- Deceduti (D): persone decedute positive al virus.

*Desideriamo ringraziare tutto il gruppo di ricerca sul COVID-19 del DISIA, e in particolare Fabrizia Mealli, Anna Gottard, Michela Baccini, Carla Rampichini, Giulia Cereda, Alessandra Mattei, Giulia Vannucci, Cecilia Viscardi e Matteo Pedone per l'interesse dedicato a questo lavoro e per i preziosi consigli.

¹[Link](#) alla repository GitHub.

- Totale positivi (Tp): totale dei soggetti “attualmente positivi”, dati dai casi totali (Tc) meno i guariti (G) meno i deceduti (D); comprende i ricoverati con sintomi, i ricoverati in terapia intensiva ed i soggetti in isolamento domiciliare, $Tp = To + Id$.
- Totale casi (Tc): comprende i soggetti attualmente positivi, i guariti ed i decessi, $Tc = Tp + G + D$.
- Nuovi positivi (ΔTc): variazione giornaliera dei casi totali, $\Delta Tc = Tc_n - Tc_{n-1}$.
- Tamponi (T): numero di tamponi effettuati.

Questi dati non sono tuttavia privi di errori, che possono essere attribuiti sia a errori di trascrizione sia al processo di raccolta (quindi a comunicazioni tardive o parziali da parte delle regioni). Per quanto riguarda i primi, si riscontra ad esempio che le serie cumulative di guariti, deceduti, totale casi e tamponi, che dovrebbero essere strettamente crescenti, alcune volte non lo sono; la Tabella 1 riporta alcune delle incongruenze di questo tipo riscontrate nel dataset relativo alle regioni². Per quanto riguarda le comunicazioni parziali, invece, sulla repository dove sono comunicati i dati c'è una sezione dedicata agli avvisi, dove vengono comunicati i problemi di questo tipo e la tipologia di dato a cui si riferiscono.

Tabella 1: Incongruenze riscontrate nel dataset chiamato “dati-regioni”.

Regione	Incongruenze
Calabria	guariti al 12/03; decessi al giorno 23/03
Campania	guariti al 21/03
Emilia-Romagna	tamponi al 30/03
Friuli Venezia Giulia	guariti al 17/03, al 18/03, al 24/03; tamponi al 19/03
Liguria	guariti al 15/03; casi totali al 01/03 e al 02/03
Lombardia	tamponi al 26/02
Piemonte	casi totali al 27/02 e al 09/03
Puglia	guariti al 14/03
Sicilia	casi totali al 02/03
Valle d'Aosta	tamponi al 15/03

2.2 Istat

Una fonte complementare di dati è l'Istat. Da una parte infatti la Protezione Civile fornisce dati relativi alla pandemia da COVID-19, dall'altra l'Istat fornisce annualmente i dati relativi ai decessi e alle cause di morte. Questi dati sono disponibili a livello nazionale, regionale e provinciale³. Sfortunatamente ad oggi le informazioni più recenti disponibili sono state rese note a dicembre 2019 e si riferiscono ai decessi avvenuti nel 2017. Questo ritardo è dovuto a come il processo di indagine è articolato. Esso infatti inizia con la comunicazione dell'avvenuto decesso da parte del medico per mezzo della compilazione di una scheda di morte, nella quale egli è tenuto a indicare il processo che ha condotto il soggetto esaminato alla morte e le sue eventuali comorbidità. Dopodiché il modello viene inviato al Comune di in cui è avvenuto il decesso per la comunicazione delle informazioni sociodemografiche da parte dell'ufficiale di stato civile ed infine giunge all'Istat [1]. Come si legge sul sito dell'Istituto,

²Dati disponibili a questo [link](#).

³Dati disponibili a questo [link](#).

sono adesso in corso le fasi di raccolta, codifica e correzione dei dati relativi al 2019 e al 2020; dunque, seguendo queste tempistiche, potremo presumibilmente conoscere il report ufficiale sulla mortalità di quest'anno, includente anche il COVID-19 come causa di morte, entro fine 2022 ⁴.

L'istituto tuttavia, conscio probabilmente dell'importanza di questi dati nella situazione attuale, il 1 aprile 2020 ha anticipato la pubblicazione dei dati riguardanti la mortalità fino al 21 marzo 2020⁵. Questi dati, sebbene abbiano un livello di dettaglio molto buono (comunale) e abbiano una scansione temporale in alcuni casi anche giornaliera, purtroppo presentano degli svantaggi rilevanti. Innanzitutto, riguardano soltanto 1084 comuni aderenti all'Anagrafe Nazionale della Popolazione Residente (ANPR), che include 5866 comuni sui 7904 totali presenti in Italia. Poichè non tutti i comuni che aderiscono all'ANPR hanno comunicato tempestivamente i dati relativi alla mortalità 2020, i 1084 comuni in realtà sono stati selezionati a partire da un sottoinsieme dei 5866 comuni che ha fornito dati completi. In particolare, questa selezione è stata operata prendendo soltanto "i comuni con un numero di decessi che, nel periodo 1 gennaio 21 marzo 2020, è risultato superiore o uguale a 10 unità e che nel mese di marzo del 2020 hanno presentato, rispetto alla corrispondente media del quinquennio 2015-2019, un incremento della mortalità pari ad almeno il 20%". Infine, questi dati non contengono informazioni sulle cause dei decessi ma soltanto sul numero complessivo dei decessi; per questo motivo i decessi riportati nell'anno 2020 saranno dovuti sia al COVID-19 sia ad altre cause.

2.3 Possibili correzioni

Per quanto riguarda le incongruenze riportate in Tabella 1, una possibile soluzione consiste nell'interpolare il dato errato. Ad esempio, i decessi riportati dalla Calabria nei giorni 22/03 e 24/03 sono 8 e 10 rispettivamente. Tuttavia, al giorno 23/03 il dato riportato è 7, e poiché si tratta di una serie storica cumulata, il dato riportato è chiaramente un errore. Per correggerlo ed ottenere quindi un dato più plausibile potremmo utilizzare un'interpolazione lineare ed imputare al 23/03 un numero di decessi pari a 9, o almeno riportare per quel giorno il valore osservato il giorno precedente, ovvero 8. Invece, per quanto riguarda il problema delle comunicazioni non tempestive o parziali, in parte possiamo fare riferimento agli avvisi pubblicati sulla repository GitHub dove talvolta viene segnalata nel dettaglio la correzione da applicare. In assenza di questa informazione è comunque possibile interpolare. Ad esempio, si apprende dagli avvisi che il 18/03 la Campania non ha comunicato nessun dato; in questi casi, il provider dei dati assegna al giorno mancante i valori del giorno prima, ma per ottenere un dato più realistico possiamo ricorrere ancora una volta all'interpolazione lineare.

Occorre specificare che vi sono alcuni casi dubbi nei quali, soprattutto per i tamponi, osserviamo valori costanti in due giorni consecutivi senza però che vi sia comunicazione di dati parziali da parte del provider. Ad esempio, per il Lazio si osserva che il totale di tamponi effettuati resta costante a 18371 per il 23/03 ed il 24/03, salvo poi passare a 20669 il giorno successivo. Un'evenienza di questo tipo potrebbe verificarsi negli stati iniziali della pandemia, ma appare improbabile nella fase più acuta. Tuttavia, sembra che il problema riguardi solo i tamponi, e non essendoci comunicazioni di alcun tipo, la scelta se lasciare il dato inalterato oppure operare una correzione spetta all'analista.

⁴Per ulteriori informazioni e aggiornamenti consultare a questa [pagina](#).

⁵Dati anticipatori Istat sulla mortalità 2020 disponibili a questa [pagina](#).

3 Contagi

3.1 Problematiche generali

Il dato dei contagi risente inevitabilmente del numero di tamponi effettuati, e questo accade per due motivi fondamentali. In primo luogo, perchè tendenzialmente al crescere del numero di tamponi aumenta il numero di contagiati che vengono scoperti. Tale andamento è visibile osservando le Figure 1 e 2 che rappresentano le serie storiche del numero di contagiati e di tamponi effettuati (in valore assoluto) in tutte le regioni italiane a partire dal 24 febbraio 2020. In Figura 3 è possibile vedere più nel dettaglio l'andamento dei tamponi e dei contagi per quattro regioni del nord e del centro Italia. Inoltre, anziché rappresentare il dato assoluto, i grafici sono relativi alle variazioni percentuali giornaliere, che ancor meglio evidenziano come tamponi e contagi si stiano muovendo insieme. In secondo luogo, il dato dei contagi risulta fortemente legato a quello dei tamponi poichè le Regioni non hanno seguito regole uniformi nella somministrazione dei test nel corso dell'epidemia. Un caso particolarmente emblematico in tal senso è il Veneto, che successivamente alla scoperta di un focolaio nel paese di Vò ha proceduto alla somministrazione di un doppio tampone a tutti gli abitanti. Inoltre, nonostante una circolare del Ministero della Salute stabilisca di sottoporre al test solo i sintomatici che nelle due settimane precedenti siano stati a contatto con un soggetto risultato positivo, il 14 marzo il governatore del Veneto ha annunciato che i tamponi verranno estesi anche agli asintomatici seguendo una politica di controlli "a tappeto", che idealmente dovrebbe andare a sottoporre a tampone 25000 persone al giorno [2]. Seguendo l'esempio del Veneto, anche Emilia Romagna e Toscana hanno annunciato in seguito di voler intraprendere la stessa politica.

La dipendenza intrinseca dei contagi rispetto ai tamponi e la mancata uniformità nelle politiche di somministrazione dei test, rendono difficoltose le analisi sul dato dei contagi: ad esempio, l'aumento dei contagi che osserviamo in Veneto è dato dall'incremento del numero di tamponi effettuati oppure da un effettiva estensione dell'epidemia? Oltre che per una più corretta descrizione del fenomeno, occorre tenere in considerazione queste problematiche anche prima e durante la calibrazione di un qualsiasi modello previsionale basato sui contagi.

3.2 Possibili correzioni

Prima di procedere con le analisi può risultare utile operare una normalizzazione per il numero di tamponi effettuati. In altre parole, si consiglia di dividere il differenziale dei contagiati in un certo arco temporale per la differenza del numero dei tamponi effettuati nello stesso arco temporale [3]. Anzichè concentrarsi sul numero dei contagi, ciò consente di modellare una sorta di probabilità di trovare un nuovo contagiato:

$$\rho_k(n) = \frac{C(n) - C(n - k)}{T(n) - T(n - k)}$$

Dove con $C(n)$ e $T(n)$ si indicano, rispettivamente, il numero di contagi e dei tamponi al tempo n . E' importante anche sottolineare che $k > 1$, poichè spesso occorre più di un giorno di tempo per conoscere l'esito del test. Scegliendo quindi $k = 3$ e $k = 5$ e applicando questa soluzione ai dati delle quattro regioni sopra considerate, otteniamo quanto riportato in Figura 4. Come è possibile notare, a differenza dell'andamento crescente che abbiamo visto in Figura 1, a partire da giorno $n = 20$ si comincia ad osservare un'inversione di tendenza, cioè, un decremento nella propensione di trovare nuovi contagiati dato il numero di tamponi effettuati.

Specifichiamo che i grafici riportati nelle Figure 1, 2, 3 e 4 sono stati creati a partire dai dati così come riportati dalla Protezione Civile, senza l'aggiunta di correzioni o interpolazioni; risulta evidente l'impatto di errori di trascrizione e trasmissione. Ad esempio, in Figura 2 possiamo osservare che la curva relativa ai tamponi effettuati in Emilia Romagna comincia a decrescere il 30 marzo; questo ha un impatto anche nella Figura 4, in cui si osserva uno "spike" negli ultimi due giorni considerati, che però dipende esclusivamente dal denominatore falsato.

Figura 1: Contagi totali cumulati, non escludendo guariti e decessi, per ogni regione italiana a partire dal 24 febbraio 2020

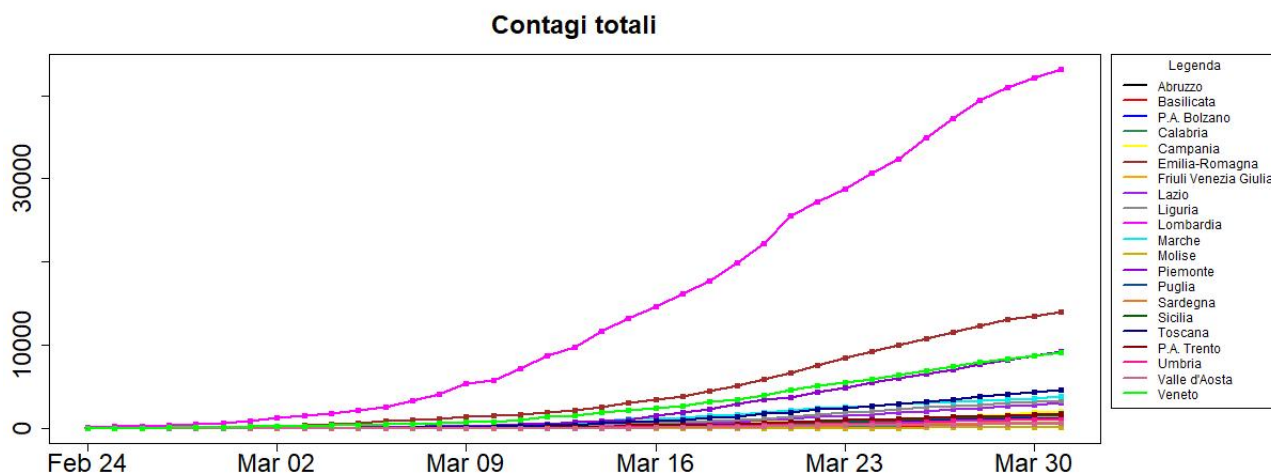


Figura 2: Tamponi effettuati per ogni regione italiana a partire dal 24 febbraio 2020

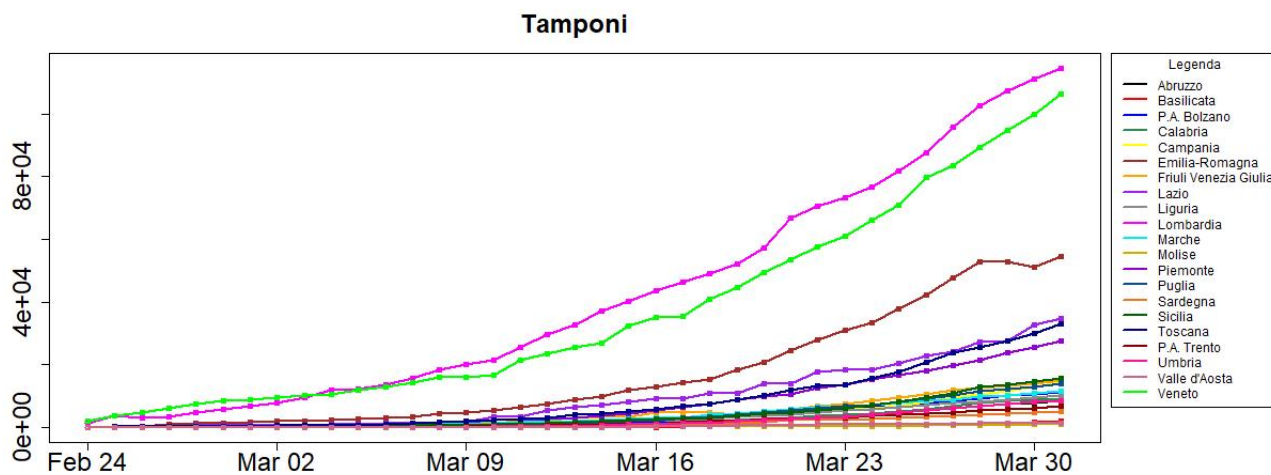


Figura 3: Variazione giornaliera percentuale del numero di tamponi e di contagi per quattro Regioni del nord e centro Italia (Lombardia, Veneto, Toscana, Emilia Romagna).

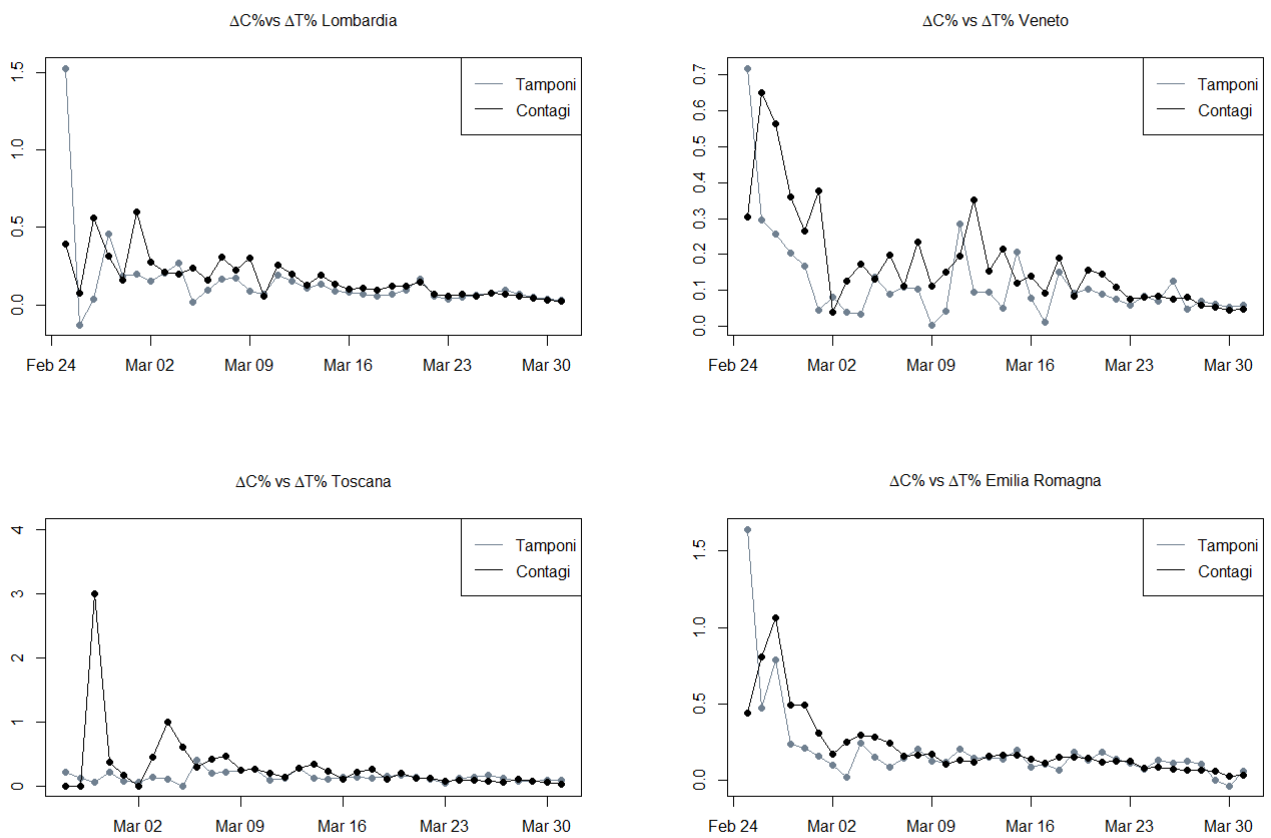
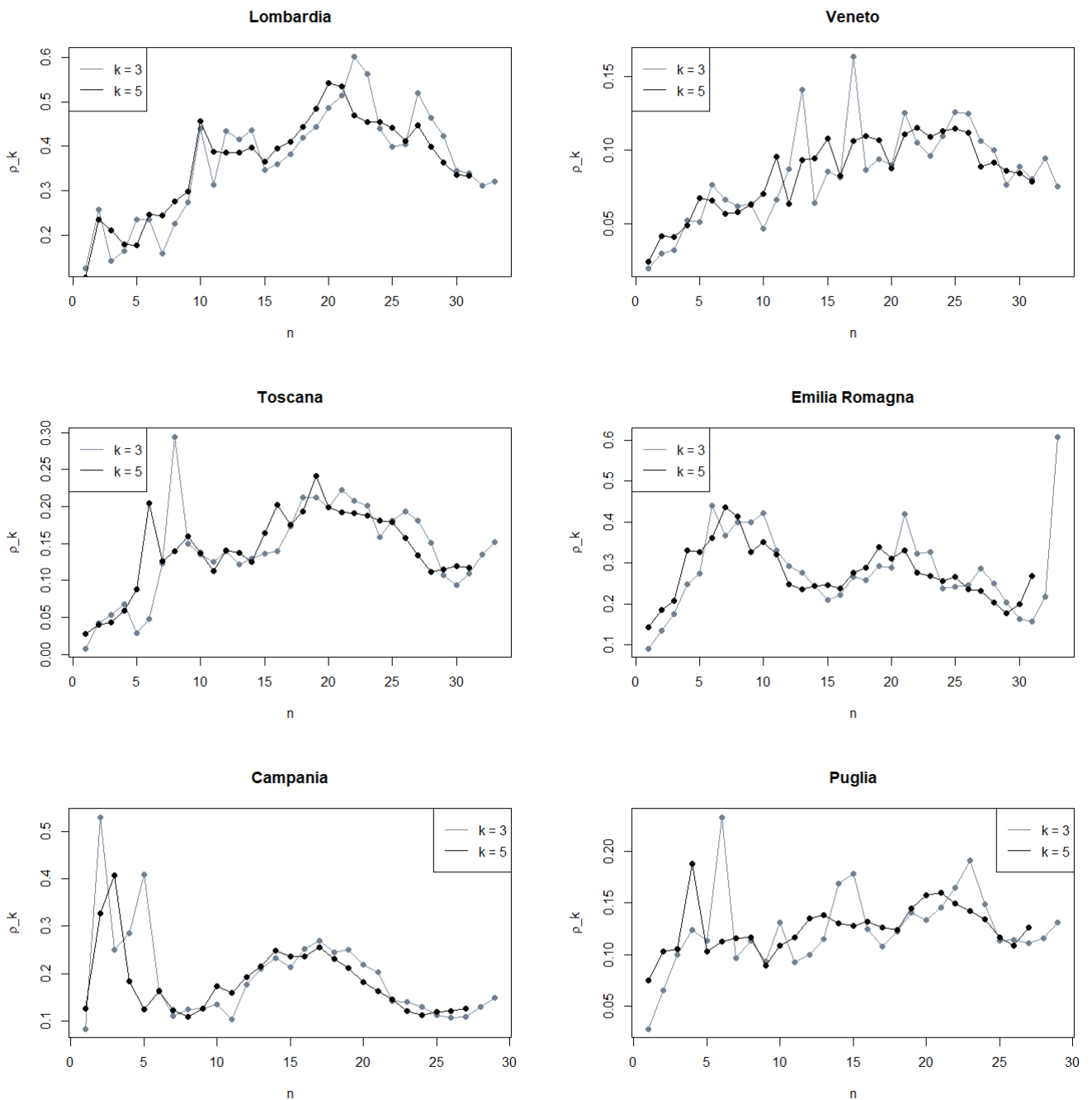


Figura 4: Contagi normalizzati per il numero di tamponi in due diversi periodi temporali in quattro regioni italiane (Lombardia, Veneto, Toscana, Emilia Romagna); lo zero sull'asse delle ascisse si riferisce al 24 febbraio 2020.



4 Decessi

4.1 Problematiche generali

Abbiamo visto che vi sono numerose problematiche legate al dato sui contagi e ognuna concorre alla sottovalutazione della reale dimensione del fenomeno. Diversamente da ciò che avviene per i contagi, il dato sul numero di morti risente essenzialmente di due problematiche che portano ad effetti di segno opposto. Il primo vizio è ancora una volta collegato al numero di tamponi effettuati e alle politiche di somministrazione dei test. Infatti, se è vero che gli infetti sono sottostimati rispetto al numero reale perché il tampone non viene somministrato a tutti, a maggior ragione è possibile che una parte degli infetti non testati si aggravi e muoia senza dunque essere riconosciuta nel novero dei decessi causati dal virus. Questo perché non ci sono politiche uniformi che impongano di effettuare tamponi post mortem a tutti i deceduti, e dunque il dato dei decessi potrebbe essere sottostimato. D'altro canto, come più volte ricordato dal capo della protezione civile nel corso dei bolletini di aggiornamento effettuati quotidianamente, il dato comunicato si riferisce alle persone decedute “con” il virus, ossia alle persone che al momento del decesso risultavano positive, e non distingue invece per quanti di questi decessi il virus sia stato la causa di morte primaria. In altre parole, il dato comunicato dalla protezione civile non distingue tra decessi avvenuti “con” il virus e “per” il virus e questo potrebbe portare ad una sopravvalutazione del dato reale.

Anche il tasso di letalità dell' 11.39% (decessi positivi sul totale dei contagiati) potrebbe dunque essere sopravvalutato, sia perché i contagi sono sottostimati sia perché non tutti i decessi considerati potrebbero essersi verificati per il COVID-19.

4.2 Possibili correzioni

Una possibile soluzione alla prima problematica, relativa ai decessi positivi a cui non è stato effettuato il tampone, è quella di calcolare la mortalità in eccesso di quest'anno rispetto agli anni scorsi; questa eccedenza infatti è verosimilmente attribuibile, direttamente o indirettamente, al COVID-19⁶. La differenza così calcolata andrà poi confrontata con i dati forniti dalla Protezione Civile sui decessi positivi. Uno scostamento tra queste quantità infatti può dare un'indicazione sul numero di decessi positivi al virus a cui però non è stato effettuato il tampone.

Illustriamo adesso più nel dettaglio la procedura proposta. Al fine di rendere più agevole la spiegazione, la si accompagna con un esempio di calcolo effettuato sui dati Istat relativi alla mortalità 2020, i quali però, come descritto nella Sezione 2.2, presentano notevoli limitazioni. Queste limitazioni, relative al modo in cui i comuni che fanno parte del campione sono stati selezionati, potrebbero portare a trarre delle conclusioni errate circa la mortalità attribuibile al COVID-19. Per questo motivo, l'esercizio di stima proposto dovrà essere sicuramente rivisto quando saranno disponibili dati relativi a tutti i comuni o almeno a comuni selezionati in modo casuale.

E' necessario calcolare innanzitutto la mortalità in eccesso di quest'anno. E' di particolare interesse la variazione del numero di decessi per tutte le cause dal 23/02 al 21/03 rispetto allo stesso periodo degli anni passati; questo confronto è reso possibile dai dati Istat sui decessi settimanali dal 01/01 al 21/03

⁶Come sarà più chiaro in seguito, con “decessi attribuibili” al virus si intendono quelli che non sarebbero avvenuti se non ci fosse stata l'epidemia. Per una parte di questi il virus è causa primaria mentre per la restante parte, invece, è causa indiretta di morte. Con causa “indiretta” di morte si intende che è plausibile che alcuni positivi siano deceduti poiché il virus ha contribuito ad aggravare condizioni di salute precaria pre-esistenti, e anche che alcuni negativi siano deceduti perché non hanno potuto o poiché hanno preferito non recarsi al pronto soccorso quando ne avevano necessità.

per gli anni dal 2015 al 2020⁷. Ad esempio, nei comuni della Lombardia che fanno parte del campione Istat, dal 2015 al 2019 si sono verificati mediamente 3330 decessi nel periodo considerato, mentre nelle stesse settimane nel 2020 se ne sono verificati 6035, ovvero un 81.2% in più. Poichè si osserva che il dato dei decessi è abbastanza stabile, oscillando da un minimo di 3214 nel 2017 ad un massimo di 3442 nel 2019, l'eccesso di mortalità di quest'anno è verosimilmente attribuibile all'epidemia di COVID-19. Come già ampiamente discusso però, sfortunatamente i dati Istat a cui si fa riferimento sono relativi soltanto ad alcuni comuni italiani. Per trarre delle conclusioni a livello regionale dobbiamo dunque assumere che l'eccesso di mortalità riscontrato nei comuni della regione di interesse considerati dall'Istat valga per l'intera regione (da qui è evidente quanto sarebbe rilevante disporre di un campione casuale). Tornando all'esempio della Lombardia, si ha che la media di decessi nell'intera regione dal 2015 al 2019 è di 7808⁸. Abbiamo visto che il dato dei decessi è abbastanza costante da un anno all'altro, perciò ci aspetteremmo di osservare anche per il 2020 un numero simile, che rappresenta dunque il numero di deceduti se l'epidemia non ci fosse stata. Tuttavia, assumendo che per l'intera Lombardia valga l'eccesso di mortalità riscontrato nei comuni lombardi inclusi nell'indagine Istat, è possibile ottenere una stima di quello che potrebbe essere il numero reale di decessi, che risulterebbe dunque pari a 14149, di cui 6340 direttamente o indirettamente causati dal virus (ovvero l'81.2% di 7808).

Infine, è necessario confrontare il dato che abbiamo trovato (decessi attribuibili al COVID-19) con quanto comunicato dalla Protezione Civile. Per la Lombardia ad esempio, dal 24/02 al 21/03 si sono registrati ufficialmente 3095 decessi positivi. Dunque, è possibile dedurre che ci siano circa 3245 decessi includenti sia le persone positive sfuggite ai controlli sia le persone che, probabilmente per timore di contrarre il virus, hanno deciso di non recarsi al pronto soccorso quando avrebbero avuto necessità di farlo (a partire dall'inizio dell'epidemia infatti si è riscontrata una diminuzione degli accessi al pronto soccorso in varie regioni italiane [4] [5]).

Se reputiamo valide le assunzioni alla base, questa procedura potrebbe costituire un valido metodo per avere un'idea dei numeri "sommersi" del fenomeno, sia con riferimento ai decessi che ai contagi. Infatti, se stimiamo che il numero di deceduti positivi è il doppio di quello riscontrato dai numeri ufficiali (il 105% in più), anche i contagi potrebbero essere il doppio (assumendo che il tasso di letalità sia costante). Applicando la stessa procedura a tutte le regioni italiane si ottengono i risultati riportati in Tabella 2.

Per quanto invece riguarda la seconda problematica relativa ai decessi, ossia la distinzione tra i morti "con" e "per" COVID-19, si rende necessaria una correzione successiva e non meno importante. Infatti, la procedura appena proposta consente di ottenere una stima dei decessi in qualche modo attribuibili al virus; tuttavia, occorre sottolineare che l'ultimo report dell'ISS riporta che su 3200 pazienti positivi deceduti, il 51.7% presentava 3 o più patologie ed il 59% aveva più di 80 anni. Da questo è possibile dedurre che il virus in molti casi possa aver portato ad effetti di anticipazione della mortalità (fenomeno noto in epidemiologia sotto il nome di "harvesting"); in sostanza, una parte della mortalità in eccesso che osserviamo nel mese di marzo sarebbe stata osservata nei restanti mesi dell'anno se non ci fosse stata la pandemia. D'altro canto, è anche possibile che l'eccesso osservato sia dovuto ad un ritardo nella mortalità, che potrebbe essersi verificato in relazione ad una stagione invernale molto mite. Come riportato dal Sistema nazionale di Sorveglianza della Mortalità Giornaliera (SiSMG) infatti, nell'inverno 2019-2020 la mortalità osservata nei mesi precedenti la pandemia è stata inferiore a quella

⁷Dati scaricabili al seguente [link](#).

⁸I dati di mortalità riferiti all'intera regione per gli anni 2015-2019 sono scaricabili al seguente [link](#).

Tabella 2: L'intera tabella si riferisce al periodo dal 23/02 al 21/03. Nella colonna "Media decessi" riportano le medie dei decessi osservati (per tutte le cause) per gli anni dal 2015 al 2019 (inclusi), mentre nella colonna "Eccesso mortalità" si riporta l'eccesso di mortalità osservata nel 2020 rispetto a queste medie. In aggiunta, nella colonna "COVID-19" si riportano i decessi del 2020 attribuibili, in maniera diretta o indiretta, al virus. A destra invece, nella colonna "PC", è possibile osservare il numero di deceduti positivi secondo la Protezione Civile. Infine, nella colonna "Diff" si riportano le differenze tra le ultime due colonne.

Regione	Media decessi	Eccesso mortalità %	COVID-19	PC	Diff
Abruzzo	1188.0	51.2	608	22	586
Basilicata	526.6	75.9	400	0	400
Calabria	1683.6	80.6	1356	5	1351
Campania	4434.2	57.6	2556	22	2534
Emilia-Romagna	4066.8	57.2	2326	715	1611
Friuli-Venezia Giulia	1185.4	22.0	261	42	219
Lazio	4597.2	11.6	534	50	484
Liguria	1734.6	18.9	328	152	176
Lombardia	7808.4	81.2	6340	3095	3245
Marche	1435.2	43.3	622	154	468
Molise	305.8	100.0	306	7	299
Piemonte	4220.8	38.0	1605	238	1367
Puglia	3148.4	34.3	1079	29	1050
Sardegna	1360.4	54.1	735	4	731
Sicilia	4515.0	30.4	1374	6	1368
Toscana	3536.0	31.5	1115	72	1043
Trentino-Alto Adige	762.6	28.8	220	48	172
Umbria	854.2	33.0	282.00	10	272
Valle d'Aosta	112.8	4.3	5	8	-3
Veneto	3962.6	35.6	1412	146	1266
Italia	51438.6	45.6	23464	4825	18639

attesa, con una riduzione stimata del 6% circa, attribuibile al minore impatto dell'influenza stagionale e alle temperature particolarmente miti dell'inverno passato⁹. Questo, come si legge nel report, ha determinato un incremento del pool di soggetti più fragili (età avanzata, malattie croniche) che si sono trovati esposti al COVID-19; questo fenomeno può essere almeno in parte responsabile dell'aumento della mortalità osservato in questa popolazione [6]. Il metodo che proponiamo di seguito permette di tenere conto di questi effetti, almeno per quanto riguarda i positivi, poichè permette di scindere i deceduti positivi in morti "per" e morti "con" il COVID-19, e i decessi avvenuti per le cause di cui sopra (anticipazione o posticipazione della mortalità) sono attribuiti alla parte "con". Questa scissione è possibile effettuando un'analisi basata sull'indagine per cause di morte Istat; in particolare, un'opzione potrebbe essere quella di calcolare la proporzione di decessi avvenuti negli ultimi anni per le malattie che sono risultate maggiormente associate al COVID-19, ed applicare questa proporzione ai deceduti risultati positivi. Come risultato otterremo quindi la proporzione dei positivi deceduti a causa delle comorbidità; la parte restante costituisce invece un lower bound della stima dei decessi direttamente causati dal virus.

Vediamo adesso il procedimento più nei dettagli. Le comorbidità maggiormente riscontrate nei deceduti positivi al virus sono riportate nel report bisettimanale rilasciato dell'Istituto Superiore di Sanità

⁹Il Sistema nazionale di Sorveglianza della Mortalità Giornaliera (SiSMG) è coordinato dal Dipartimento di Epidemiologia SSR Lazio ed è stato costituito nel 2004 al fine di monitorare su base giornaliera il numero di decessi e segnalare tempestivamente eccessi di mortalità al fine di attivare risposte all'emergenza in tempi brevi.

(ISS) ¹⁰; il più recente di questi report, risalente al 20 marzo, riporta che le patologie maggiormente riscontrate nei soggetti positivi deceduti sono quelle nella Tabella 3. Andando quindi a ricercare nelle

Tabella 3: Patologie croniche più comuni pre-esistenti nei pazienti deceduti, così come riscontrate tramite l’analisi di 481 cartelle cliniche su 3200 decessi (pari al 15 % del campione complessivo).

Patologie	Numero	%
Cardiopatía ischemica	145	30.1
Fibrillazione atriale	106	22.0
Ictus	54	11.2
Ipertensione arteriosa	355	73.8
Diabete mellito	163	33.9
Demenza	57	11.9
BPCO	66	13.7
Cancro attivo negli ultimi 5 anni	94	19.5
Epatopatia cronica	18	3.7
Insufficienza renale cronica	97	20.2

indagini Istat più recenti (quelle relative agli anni dal 2015 al 2017 inclusi) le cause di morte che comprendono queste patologie, possiamo calcolarne la percentuale sul totale dei decessi, escludendo quelli avvenuti per cause traumatiche. Infatti, non avrebbe senso assumere che tra i positivi deceduti ve ne siano alcuni morti per cause incidentali o traumatiche.

Occorre notare che l’intero modus operandi si regge chiaramente sull’assunzione che la proporzione dei decessi per comorbidità del COVID-19 si mantenga più o meno costante negli anni; visto che, nel nostro caso, consideriamo soltanto gli ultimi tre anni disponibili, e che ne facciamo anche la media, riteniamo che tale assunzione sia plausibile. Un’altra assunzione implicita che facciamo è supporre che decessi delle persone risultate positive siano dovuti o al COVID-19 o alle comorbidità. Questo implica un’assunzione anche sui decessi delle persone negative. Infatti, se ad esempio il 75% dei decessi per cause non traumatiche avviene per le comorbidità, vuol dire che il restante 25% avviene per altre malattie. Tuttavia, quando applichiamo il 25% ai malati COVID-19 e lo attribuiamo al virus, stiamo implicitamente assumendo che le altre malattie possono essere causa di decesso solo nelle persone negative¹¹. Possiamo reputare valida anche questa assunzione, poiché il numero di persone positive decedute nè per il virus nè per le comorbidità, bensì per malattie diverse da quelle riportate dall’ISS, è da ritenersi esiguo.

A titolo di esempio, la Tabella 4 riporta i numeri relativi alla Lombardia ed il calcolo delle percentuali. Il metodo proposto può essere applicato non solo al totale dei decessi positivi comunicato dalla Protezione Civile ma anche alla stima precedentemente ottenuta dei decessi attribuibili al virus e non rilevati dalle stime ufficiali¹². Sommando le stime così ottenute è dunque possibile avere un’indicazione circa numero totale dei decessi avvenuti “per” il COVID-19.

Applicando questo procedimento a tutte le regioni italiane, otteniamo quanto riportato in Tabella 5, dove si vede che il totale dei deceduti “per” COVID-19 in Italia al 21 marzo risulterebbe 3608¹³.

¹⁰I report bisettimanali comunicati dall’ISS sono disponibili al seguente [link](#).

¹¹In altre parole, stiamo assumendo che i positivi siano morti o per il virus o per le comorbidità riportate dall’ISS, mentre i negativi siano morti o per incidenti e altre cause traumatiche, o per le comorbidità o per altre malattie. Anche se per le nostre finalità non è necessario conoscere il numero dei decessi negativi al COVID-19, questa specificazione appare doverosa in quanto occorre comprendere e rendere note tutte le assunzioni alla base delle nostre stime.

¹²Con riferimento a questi ultimi occorre però specificare che i risultati così ottenuti potrebbero essere sovrastimati, in quanto, come ricordato in precedenza, di questo gruppo potrebbero far parte anche soggetti negativi deceduti per

Tabella 4: Cause di morte legate alle patologie dei soggetti deceduti al covid-19 così come riportate dall'indagine Istat degli ultimi tre anni con riferimento alla regione Lombardia

Cause di morte Istat	2017	2016	2015
malattie del sistema circolatorio	32512	30730	33331
tumori	31118	31110	30851
diabete mellito	2400	2305	2369
demenza	3518	3031	2984
malattie croniche delle basse vie respiratorie	3613	3265	3200
altre malattie del sistema respiratorio	1926	1774	2116
altre malattie dell'apparato digerente	2544	2474	2521
malattie dell'apparato genitourinario	1834	1714	2030
(A) Totale intermedio	79465	76403	79402
(B) Cause esterne di traumatismo	3478	3342	3421
(C) Totale (tutte le cause)	98784	94363	98211
(A/C-B)%	83.4%	83.9%	83.8%
Media			83.7%

Tabella 5: L'intera tabella si riferisce al periodo dal 23/02 al 21/03. Nella colonna "PC" si riporta il numero di decessi positivi secondo la Protezione Civile; nella colonna "Diff" invece si riporta la stima del numero di decessi dovuti (direttamente o indirettamente) al virus in aggiunta al numero comunicato dalla Protezione Civile. Nella colonna "%" si riportano le percentuali di decessi per il COVID-19. Nelle colonne "PC PER" e "Diff PER" si riportano, rispettivamente, le stime dei decessi per COVID-19 ottenute applicando la metodologia proposta al numero di decessi positivi comunicato dalla Protezione Civile e al numero stimato di decessi sfuggiti dai conteggi ufficiali. Nell'ultima colonna infine, si riporta il totale stimato dei decessi per COVID-19, ottenuto come somma delle due colonne precedenti.

Regione	PC	Diff	%	PC PER	Diff PER	Totale
Abruzzo	22	586	14.9	3	88	91
Basilicata	0	400	13.8	0	55	55
Calabria	5	1351	13.4	1	182	183
Campania	22	2534	12.2	3	309	312
Emilia-Romagna	715	1611	16.2	116	261	377
Friuli Venezia Giulia	42	219	15.3	6	34	40
Lazio	50	484	14.6	7	70	77
Liguria	152	176	17.7	27	31	58
Lombardia	3095	3245	16.3	505	529	1034
Marche	154	468	16.0	25	75	100
Molise	7	299	13.9	1	42	43
Piemonte	238	1367	16.4	39	224	263
Puglia	29	1050	15.0	4	157	161
Sardegna	4	731	17.1	1	125	126
Sicilia	6	1368	13.9	1	189	190
Toscana	72	1043	17.5	13	183	196
Trentino Alto Adige	48	172	16.2	8	28	36
Umbria	10	272	16.1	2	44	46
Valle d'Aosta	8	-3	15.9	1	0	1
Veneto	146	1266	15.5	23	196	219
Italia	4825					3608

l'impossibilità di accedere al pronto soccorso.

¹³Nella Tabella 5 si riporta un valore di "Diff" negativo per la Valle d'Aosta; ciò è da attribuirsi al fatto che in questa regione non si è osservato un rilevante incremento della mortalità rispetto agli anni passati.

In conclusione, vista la rilevanza dell'impatto che le diverse problematiche descritte potrebbero potenzialmente avere sui numeri riportati dalla Protezione Civile, suggeriamo di leggerli ed analizzarli con cautela.

Riferimenti

- [1] Istat. Informazioni sulla rilevazione - indagine su decessi e cause di morte. <https://www.istat.it/it/archivio/4216>, 24 marzo 2020.
- [2] Rai News. Veneto, Zaia: tamponi a tappeto. <http://www.rainews.it/dl/rainews/articoli/zaia-veneto-tamponi-a0107990-269a-4cce-80e6-e8851a56a45e.html>, 17 marzo 2020.
- [3] Duccio Fanelli Francesco Piazza. Analysis and forecast of covid-19 spreading. <https://www.researchgate.net/project/Analysis-and-forecast-of-COVID-19-spreading>, 12 marzo 2020.
- [4] Sky TG24. Coronavirus, Gallera: “Diminuzione degli accessi in tutti i pronto soccorso lombardi”. <https://tg24.sky.it/milano/2020/03/27/coronavirus-lombardia-conferenza-stampa-gallera.html>, 27 marzo 2020.
- [5] Il Tirreno. Effetto coronavirus: accessi in calo del 20% al pronto soccorso di Prato. <https://iltirreno.gelocal.it/prato/cronaca/2020/02/25/news/effetto-coronavirus-accessi-in-calo-del-20-al-pronto-soccorso-1.38515883>, 25 febbraio 2020.
- [6] DEP Lazio. Andamento della mortalità giornaliera (SiSMG) nelle città italiane in relazione all’epidemia di Covid-19, 1 febbraio - 28 marzo (secondo rapporto). <https://tinyurl.com/vclrxjg>, 3 aprile 2020.