

Analisi spaziale per l'identificazione di cluster di casi durante l'emergenza COVID-19 a Roma e nel Lazio

Spatial analysis for detecting clusters of cases during the COVID-19 emergency in Rome and in the Lazio Region (Central Italy)

Chiara Badaloni,^{1*} Federica Asta,^{1*} Paola Michelozzi,¹ Francesca Mataloni,¹ Enrico Di Rosa,² Paola Scognamiglio,³ Francesco Vairo,³ Marina Davoli,¹ Michela Leone⁴

¹ Dipartimento di epidemiologia del Servizio sanitario regionale, ASL Roma 1, Regione Lazio, Roma

² Dipartimento di prevenzione, ASL Roma 1, Roma

³ Servizio regionale per l'epidemiologia, sorveglianza e controllo delle malattie infettive, Istituto nazionale malattie infettive "Lazzaro Spallanzani" IRCCS, Roma

⁴ UOS Controllo di gestione, ASL Frosinone, Frosinone

*Gli autori hanno contribuito in ugual misura al lavoro

Corrispondenza: Federica Asta; f.asta@deplazio.it

RIASSUNTO

INTRODUZIONE: in Italia, uno dei Paesi europei più colpiti dall'epidemia di COVID-19, i dati mostrano la forte eterogeneità geografica dell'epidemia. Gli studi che valutano la diffusione spaziale della pandemia hanno analizzato, finora, vaste aree geografiche.

OBIETTIVI: proporre una strategia di analisi per accertare la natura non casuale della diffusione spaziale dei casi di infezione da COVID-19 e identificare eventuali aggregazioni territoriali, al fine di potenziare le attività di *contact tracing* in specifiche aree del territorio regionale e di una grande area urbana.

METODI: sono stati considerati tutti i casi di COVID-19 della Regione Lazio notificati al Servizio regionale per l'epidemiologia, sorveglianza e controllo delle malattie infettive (Seresmi) con aggiornamento giornaliero dall'inizio dell'epidemia al 27.04.2020. Le analisi sono state svolte considerando due periodi di indagine (il primo dall'inizio dell'epidemia al 6 aprile e il secondo dall'inizio dell'epidemia al 27 aprile) e due diversi livelli di aggregazione: l'intera regione Lazio escluso il comune di Roma, dove i 377 comuni rappresentano le unità areali, e il comune di Roma, dove le unità areali in studio sono le 155 zone urbanistiche (ZUR). La statistica Scan di Kulldorff è stata utilizzata per accertare la natura non casuale della diffusione spaziale dei casi infetti e identificare eventuali aggregazioni territoriali dei casi di infezione da COVID-19, utilizzando un'analisi spaziale retrospettiva in due tempi sovrapposti.

RISULTATI: l'analisi condotta a livello regionale nei due periodi di indagine ha messo in evidenza la presenza di 7 *cluster* localizzati. Nel comune di Roma, nel primo periodo è stato identificato un unico *cluster* (Centro Storico) che include 7 zone urbanistiche, mentre nel secondo periodo sono stati osservati due cluster distinti (Omo e Farnesina).

CONCLUSIONI: la statistica Scan è un prezioso strumento di sorveglianza per monitorare i focolai di malattia durante la fase attiva dell'epidemia e un utile contributo alla sorveglianza epidemiologica durante l'epidemia COVID-19 in un determinato territorio.

Parole chiave: COVID-19, pandemia, *clustering* spaziale, sorveglianza epidemiologica

COSA SI SAPEVA GIÀ

- Finora gli studi che hanno valutato la diffusione spaziale della pandemia COVID-19 sono stati limitati e hanno analizzato vaste aree geografiche.
- L'analisi di cluster si colloca nell'ambito delle analisi spaziali, dove il livello di aggregazione territoriale in studio diventa un elemento fondamentale.

COSA SI AGGIUNGE DI NUOVO

- Questo è il primo studio italiano che utilizza la statistica Scan per identificare cluster da COVID-19 in una regione italiana.
- Viene suggerita un'analisi spaziale retrospettiva in due tempi sovrapposti per monitorare l'andamento dei cluster epidemici e per individuare con tempestività nuovi possibili focolai.
- L'approccio di indagine proposto è tanto più efficiente quanto più i nuovi casi COVID-19 sono resi disponibili tempestivamente, consentendo di verificare se le misure di contenimento e di isolamento sono state efficaci nel ridurre la diffusione dell'epidemia.

ABSTRACT

BACKGROUND: one of the most affected European countries by the COVID-19 epidemic is Italy; data show the strong geographical heterogeneity of the epidemic.

OBJECTIVES: to propose an analysis strategy to ascertain the non-random nature of the spatial spread of COVID-19 cases infection and identify any territorial aggregations, in order to enhance contact tracing activities in specific areas of the Lazio Region (Central Italy) and a large urban area as Rome.

METHODS: all cases of COVID-19 of the Lazio Region notified to the Regional Service for Epidemiology, Surveillance, and Control of Infectious Diseases (Seresmi) with daily updates from the beginning of the epidemic to April 27, 2020 were considered. The analyses were carried out considering two periods (the first from the beginning of the epidemic to April 6 and the second from the beginning of the epidemic to April 27) and two different levels of spatial aggregation: the entire Lazio region excluding the Municipality of Rome, where the 377 municipalities represent the area units, and the Municipality of Rome, where the area units under study are the 155 urban areas (ZUR). The Scan statistic of Kulldorff was used to ascertain the non-random nature of the spatial spread of infected cases and to identify any territorial aggregations of cases of COVID-19 infection, using a retrospective spatial analysis in two overlapping periods.

RESULTS: analysis was conducted at regional level in the two survey periods and revealed the presence of 7 localized clusters. In the Municipality of Rome, a single cluster (Historic Centre) was identified in the first period which includes 7 urban areas, while in the second period two distinct clusters (Omo and Farnesina) were observed.

CONCLUSIONS: Scan statistics are an important surveillance tool for monitoring disease outbreaks during the active phase of the epidemic and a useful contribution to epidemiological surveillance during the COVID-19 epidemic in a specific territory.

Keywords: COVID-19, pandemic, spatial clustering, disease surveillance

INTRODUZIONE

Al 04.04.2020, in base ai dati pubblicati dal Centro europeo per la prevenzione e il controllo delle malattie (ECDC), sono stati notificati complessivamente, in tutto il mondo, 3.469.241 casi positivi al COVID-19, di cui 247.158 decessi (7,1%). In Europa, nei Paesi dell'Unione europea (UE), dell'area economica europea (EEA) e del Regno Unito sono stati registrati 1.512.466 casi e 143.820 decessi (9,5%).¹ L'Italia è il quarto Paese in Europa per numero di casi totali, dopo Russia, Inghilterra e Spagna, e il secondo per numero di decessi (dopo la Spagna), con un totale di 210.717 casi positivi di coronavirus e 28.884 persone decedute (13,7%). Una delle evidenze chiare che emerge dai dati della sorveglianza è la forte eterogeneità geografica della diffusione dell'epidemia nel Paese: con le regioni del Nord (Lombardia, Piemonte, Emilia-Romagna e Veneto) più colpite delle regioni del Centro-Sud. Utilizzando una suddivisione geografica più fine, i dati per provincia e per comune mostrano la forte eterogeneità della diffusione dell'epidemia.²

Molte organizzazioni e università, tra cui l'Istituto superiore di sanità (ISS), l'Organizzazione mondiale della sanità (OMS) e la Johns Hopkins University (JHU), hanno messo a disposizione dati e visualizzazioni che presentano la diffusione geografica del Coronavirus,³⁻⁵ ma finora gli studi che hanno valutato la diffusione spaziale della pandemia COVID-19 sono stati limitati e hanno analizzato vaste aree geografiche.⁶⁻⁹ Negli Stati Uniti, è stata utilizzata una statistica di scansione prospettica spazio-temporale per rilevare i *cluster* emergenti di COVID-19 a livello di contea, mettendo in luce come questo tipo di analisi possa essere utile per integrare le mappe di base dei casi e dei tassi d'incidenza di malattia e per comprendere meglio la progressione del rischio per tutta la durata della pandemia.⁶ Anche alcuni ricercatori cinesi hanno utilizzato un approccio geografico, sottolineando una significativa dipendenza spaziale del numero di nuovi casi e confermando che l'esame della diffusione spaziale, soprattutto nelle prime fasi dell'epidemia, è molto importante per prevenire ulteriori trasmissioni.¹⁰ In Corea, l'analisi dei *cluster* ha reso possibile l'identificazione del focolaio di Daegu e l'adozione di misure restrittive di contenimento allo scopo di ridurre il numero di nuovi casi e il numero di decessi.⁹

In Italia, pochi studi hanno analizzato la diffusione spaziale della pandemia COVID-19.¹¹ Nonostante i metodi di analisi spaziale abbiano avuto finora un utilizzo limitato, procedure standardizzate applicate tempestivamente ai dati della

sorveglianza possono rappresentare uno strumento utile per l'identificazione di aggregazioni territoriali di casi (*cluster*). Queste metodologie rappresentano un utile strumento per mettere in atto strategie di contenimento della diffusione dell'infezione e ridurre i contatti, potenziando gli interventi di *contact tracing* in specifiche aree del territorio.

La metodologia Scan che viene proposta è usata per identificare *cluster* basati sulla densità spaziale dei casi di una determinata malattia e può essere uno strumento utilizzato per la sorveglianza anche durante il corso di un'epidemia.^{12,13} Questo metodo è in grado di valutare se specifiche aggregazioni spaziali sono dovute al caso o se rappresentino, invece, aggregati anomali di casi di malattia.^{6,14-16} La statistica Scan considera cerchi o ellissi quali finestre di scansione, che si spostano sistematicamente attraverso la specifica area in studio per identificare aggregati di casi. Se applicata a uno specifico periodo di studio, la statistica Scan è in grado di rilevare i *cluster* attivi e, a mano a mano che aumentano i casi confermati, può essere riprodotta al fine di confermare nello spazio e nel tempo i gruppi di casi già individuati oltre che rilevare nuovi *cluster* emergenti. Questo tipo di approccio offre un contributo tempestivo al monitoraggio dell'epidemia in una specifica area e può essere utile per mirare gli interventi di sanità pubblica.

Il presente lavoro propone una strategia di analisi spaziale per accertare la natura non casuale della diffusione dei casi infetti e identificare eventuali aggregazioni territoriali dei casi di infezione da COVID-19, utilizzando la statistica Scan di Kulldorff.^{13,17} L'analisi considera due diversi livelli di aggregazione: la regione Lazio (dove le unità areali sono i singoli comuni) e il comune di Roma (considerando come unità areali le zone urbanistiche) ed è stata eseguita in due diversi periodi temporali: dall'inizio dell'epidemia fino al 6 aprile e dall'inizio dell'epidemia fino al 27 aprile. Nella regione Lazio, i dati sui casi di COVID-19 vengono aggiornati quotidianamente attraverso un monitoraggio continuo e, quindi, sfruttando la potenza dei metodi *cluster* di tipo Scan, è possibile individuare con tempestività focolai di epidemia, assegnando maggiori risorse nelle aree più critiche e implementando quarantene più o meno severe.

MATERIALI E METODI

POPOLAZIONE IN STUDIO

Sono stati considerati tutti i casi di COVID-19 della Regione Lazio notificati al Servizio regionale per l'epidemiologia, sorveglianza e controllo delle malattie infettive (Serresmi) dell'Istituto nazionale malattie infettive Lazzaro

Spallanzani. Il Seresmi, nell'ambito del Sistema di sorveglianza integrata nazionale, riceve la notifica di tutti i casi di COVID-19 con una conferma di laboratorio del virus che causa il COVID-19.¹⁸ Dall'inizio dell'epidemia, con aggiornamento giornaliero, è stata fornita al Dipartimento di epidemiologia del Servizio sanitario regionale del Lazio la lista dei nuovi contagi identificati tramite un codice anonimo univoco. Per ogni caso è stata resa disponibile l'informazione sull'indirizzo di residenza o domicilio (qualora la residenza fosse fuori Regione) e il Comune di residenza o domicilio; altre informazioni, quali età e genere del caso, non sono disponibili.

A partire dalle informazioni sugli indirizzi e il comune sono stati esclusi tutti i casi con indirizzo e/o comune mancante. I restanti indirizzi sono stati geolocalizzati (usando la procedura di geocodifica di Google Maps) e, sulla base delle coordinate ottenute, per gli indirizzi relativi al Comune di Roma è stata attribuita la corrispondente zona urbanistica (ZUR). A seguito di questa procedura, sono stati ulteriormente esclusi gli indirizzi per i quali è stato riscontrato un errore nella geocodifica.

ANALISI STATISTICA

Le analisi sono state svolte considerando due diversi livelli di aggregazione: il primo riguarda l'intera Regione Lazio escluso il comune di Roma, dove i 377 comuni rappresentano le unità areali; mentre il secondo riguarda l'analisi del solo comune di Roma, considerando quali unità areali le 155 ZUR. Per l'identificazione dei *cluster*, considerando entrambe le aggregazioni territoriali, si è deciso di prendere in esame arbitrariamente due intervalli temporali distinti: il primo comprende tutti i casi incidenti cumulati dall'inizio dell'epidemia al 6 aprile (periodo 1) e il secondo dall'inizio dell'epidemia fino al 27 aprile (periodo 2). È da notare che il primo periodo coincide con la fase di crescita esponenziale dei casi, mentre il secondo comprende anche la fase discendente, iniziata i primi di aprile nella regione Lazio.¹⁸

Sono stati analizzati sia i test per il *clustering* generalizzato¹⁹ sia quelli per il *clustering* localizzato,^{12,13,16} seguendo la metodologia riportata in dettaglio nell'articolo di Schifano et al. (2016).²⁰

METODI DI CLUSTER SPAZIALE

I metodi per la ricerca di *cluster* spaziali generalizzati verificano l'esistenza di una tendenza generale all'aggregazione spaziale (*clustering*) dei casi sul territorio oggetto di studio; il test di Tango¹⁹ valuta se gli eccessi di rischio sono più simili tra zone limitrofe rispetto ad aree più lontane. In generale, in tutti i test per il *clustering* generalizzato, la statistica test è una somma dei casi nelle diverse unità areali e la sua dipendenza dalle distanze geografiche tra i casi rende il test più potente nelle aree urbanizzate che non in quelle rurali, perché meno popolate.^{19,21,22}

I metodi per l'identificazione di *cluster* localizzati si pro-

pongono di scansionare piccole aree all'interno dell'intera area di studio al fine di valutare se specifiche aggregazioni spaziali siano dovute al caso o se si tratti di casi anomali, ovvero gruppi significativi di casi di malattia (eventuali *cluster*).^{6,14-16} Si è scelto di utilizzare il test di Kulldorff e Nagarwalla (K&N),^{12,17} uno tra i test più versatili e potenti. La procedura di K&N si basa sulla costruzione di finestre circolari la cui lunghezza del raggio è determinata in modo tale che l'area includa un numero di soggetti residenti corrispondente a una percentuale definita a priori della popolazione dell'intera area in studio. Per il test di K&N, l'elemento critico è la scelta della percentuale di popolazione da includere nella finestra. Tale criticità è stata superata utilizzando la metodologia *Scan statistics*⁶ implementata nell'applicativo SpatialEpiApp dell'omonimo *package* del software R-gui che utilizza il pacchetto statistico implementato in SatScan (<http://www.satscan.org>). L'applicativo permette di individuare *cluster* spaziali e di ottimizzare il valore relativo alla percentuale di popolazione da includere nella finestra circolare, utilizzando finestre da 0 fino a un limite massimo del 50% della popolazione totale in studio.¹³ È stato utilizzato un modello spaziale di Poisson, considerando come unità spaziale i comuni del Lazio o le zone urbanistiche del comune di Roma e come popolazione sottostante quella residente al 2018 selezionata da fonte Istat.

Per confermare i *cluster* spaziali individuati con la procedura di K&N e rendere più solidi i risultati, si è scelto di condurre un'analisi di sensibilità utilizzando l'altro metodo maggiormente utilizzato in letteratura, cioè il test di Besag & Newell (B&N).²³ Dal momento che entrambi i test di *cluster* localizzato presentano criticità, si è ritenuto utile eseguire entrambi i test, dato che non è noto se un test sia superiore all'altro. Anche la procedura di B&N si basa sulla costruzione di finestre circolari, ma in questo caso la lunghezza del raggio si estende fino a includere un numero minimo di casi definito a priori (k). Per il test di B&N, l'elemento critico è quindi la scelta del k . Si è scelto di individuare i *cluster* per diversi valori di k (10, 20, 30, 50).

Sono stati definiti *cluster* solo i comuni o le ZUR più significative (p -value <0,01 e 1.000 simulazioni MCMC). La scelta di un p -value inferiore ad =0,05 è giustificata dalla necessità di indicare come *cluster* solo quelli con maggiore probabilità di essere non casuali, al fine di non creare falsi allarmi, dato il gran numero di test condotti.²⁰

In questo studio si presentano i risultati del test di K&N confermati anche dal test di B&N sia attraverso una visualizzazione grafica sia attraverso una tabella riassuntiva contenente il centroide del *cluster*, i comuni o le ZUR che compongono il *cluster*, il numero totale di casi osservati nel *cluster*, la popolazione contenuta nel *cluster* e i p -value di ogni singolo *cluster*. Per centroide di *cluster* si intende il centro geografico dell'area dell'intero *cluster* rappresentato da un comune che per lo studio rappresenta l'unità areale.

RISULTATI

Dall'inizio dell'epidemia al 27 aprile 2020, sono stati considerati per l'analisi 3.514 (circa il 95% del totale dei casi notificati) casi di COVID-19 in tutta la regione Lazio escluso il comune di Roma; 2.203 casi nel comune di Roma. Il restante 5% non è stato considerato, poiché non riportava il comune/indirizzo di residenza o perché la procedura di geocodifica non ha avuto successo.

Il numero dei casi per comune varia tra 0 e 254 casi (comune di Civitavecchia); mentre il *range* per il comune di Roma varia tra 0 e un massimo di 71 casi (ZUR di Torre Angela). La distribuzione dei casi sul territorio sembra mettere in evidenza alcune zone di aggregazione spaziale ben definite. Nei materiali aggiuntivi on-line sono riportate le mappe che rappresentano la distribuzione del numero di casi di COVID-19 notificati per comune di residenza/domicilio al 27 aprile (figura S1) e per ZUR (figura S2). Il test di Tango per entrambe le analisi, considerando sia i Comuni sia le ZUR, mette in luce la possibilità di un'aggregazione non casuale dei casi su tutto il territorio analizzato, quindi una tendenza generale alla clusterizzazione nelle aree in studio ($p\text{-value}=0.005$).

I risultati distinti dei singoli test per i due intervalli temporali sono riportati nelle figure da S3 a S10. Per il test di B&N, si è scelto di presentare i risultati per valori di k maggiormente utilizzati in letteratura²⁰ e, nello specifico, $k=50$ per l'analisi a livello comunale e $k=30$ per l'analisi a livello di ZUR (per gli altri valori di k , i risultati non vengono riportati); il valore di α è stato fissato a 0,01. Il test di K&N è stato ottimizzato da una frazione di popolazione pari al 50% e anche in questo caso α è stato fissato a 0,01. Di seguito sono riportati i *cluster* localizzati per le due aggregazioni territoriali e per i due intervalli temporali analizzati considerando le aree identificate da K&N consolidate dal metodo di B&N (figura 1 2, 3 e 4).

Nella figura 1 sono rappresentati i *cluster* individuati nei due intervalli di tempo considerati (periodo 1 e periodo 2) nell'analisi condotta a livello regionale, considerando i comuni come unità areali.

In entrambi i periodi, si osservano 7 *cluster*, ma di composizione differente. Si nota la contrazione del *cluster* di Tarquinia: da 5 comuni (Monte Romano, Tarquinia, Tuscania, Allumiere, Civitavecchia) all'inizio del mese di aprile si riduce a un solo comune (Allumiere) alla fine dello stesso mese; si riporta anche l'espansione del *cluster* di Grottaferrata, che ha incluso successivamente anche il comune di Rocca di Papa. Il *cluster* di Contigliano e di Vico nel Lazio rimangono stabili, mentre i *cluster* di Fondi, Nerola e Bracciano cessano di essere tali. Infine, nel periodo più recente, si osservano nuovi *cluster* a Celleno, Campagnano di Roma e Fiano Romano. Da notare che il *cluster* di Campagnano di Roma è spazialmente vicino a quello di Bracciano, che potrebbe fare supporre un'origine in comune.

I risultati relativi ai due periodi in studio sono riportati in dettaglio in tabella 1.

Nella figura 2 sono rappresentati i *cluster* individuati nell'analisi condotta a livello di zona urbanistica nel Comune di Roma per i due periodi di tempo in studio.

Nel primo periodo viene identificato un solo *cluster* localizzato al centro del comune di Roma che comprende 7 zone urbanistiche (tabella 2). Nel secondo periodo, invece, si osservano due *cluster* localizzati di dimensioni più ridotte: uno composto solo dalla ZUR Omo e l'altro da due ZUR: Farnesina e Foro Italico (tabella 2).

DISCUSSIONE

I casi di COVID-19 analizzati nella Regione Lazio (escluso il comune di Roma) dall'inizio dell'epidemia al 27.04.2020 sono stati 3.514, mentre 2.203 casi sono quelli nel comune di Roma. La distribuzione territoriale dei casi sembra mettere in evidenza alcune zone di aggregazione spaziale ben definite sia a livello regionale sia nel comune di Roma e suggerisce la presenza di diversi focolai, soprattutto al di fuori del Comune di Roma. Nel caso di un'epidemia, l'aggregazione non casuale di casi suggerisce, infatti, la possibile origine comune a seguito di contatti originati da uno o più casi. Una spiegazione alternativa risiede nelle politiche di esecuzione dei test diagnostici che in Italia sono state molto eterogenee tra Regioni e anche all'interno della stessa regione e che influenzano il numero di casi notificati in un territorio specifico. Questa seconda ipotesi è difficilmente verificabile in mancanza di serie giornaliere dei numeri di tamponi effettuati per comune/ZUR.

Si osserva, inoltre, una tendenza generale alla clusterizzazione nelle aree in studio ($p\text{-value}=0,005$). I metodi per l'identificazione di *cluster* localizzati di casi di COVID-19 consentono di identificare aggregati spaziali in Lazio e per il comune di Roma per due diversi intervalli temporali. Il primo periodo cerca di cogliere uno dei momenti di massima espansione dell'epidemia, il secondo include il periodo da inizio aprile in poi nella Regione Lazio in cui la curva epidemica cominciava a presentare una flessione. Pertanto, i due periodi differiscono anche in termini di numero di casi giornalieri, con una minore potenza associata al periodo di contrazione della curva epidemica nel mese di aprile (secondo periodo). Entrambe le statistiche utilizzate hanno riportato la presenza di 7 *cluster* localizzati per l'analisi condotta a livello regionale nei due periodi di indagine. Rimangono stabili alcuni *cluster* (Contigliano e Vico nel Lazio), mentre si spengono altri *cluster* (Fondi, Nerola e Bracciano). Nel secondo periodo, si mettono in evidenza, inoltre, nuovi *cluster* (Celleno, Campagnano di Roma e Fiano Romano). È da notare che, mentre per alcuni di questi *cluster* (Fondi dal 19 marzo, Nerola dal 25 marzo e Contigliano dal 30 marzo) è stata adottata una chiusura totale (cosiddetta zona rossa) con divieto di ingresso e allontanamento dal comune, i *cluster* di Bracciano e Vico nel Lazio, estesi su diversi comuni, non sono stati riconosciuti come zona rossa e potrebbero in realtà risentire di un *detection bias* dovuto a una maggiore prevalenza di test diagnostici

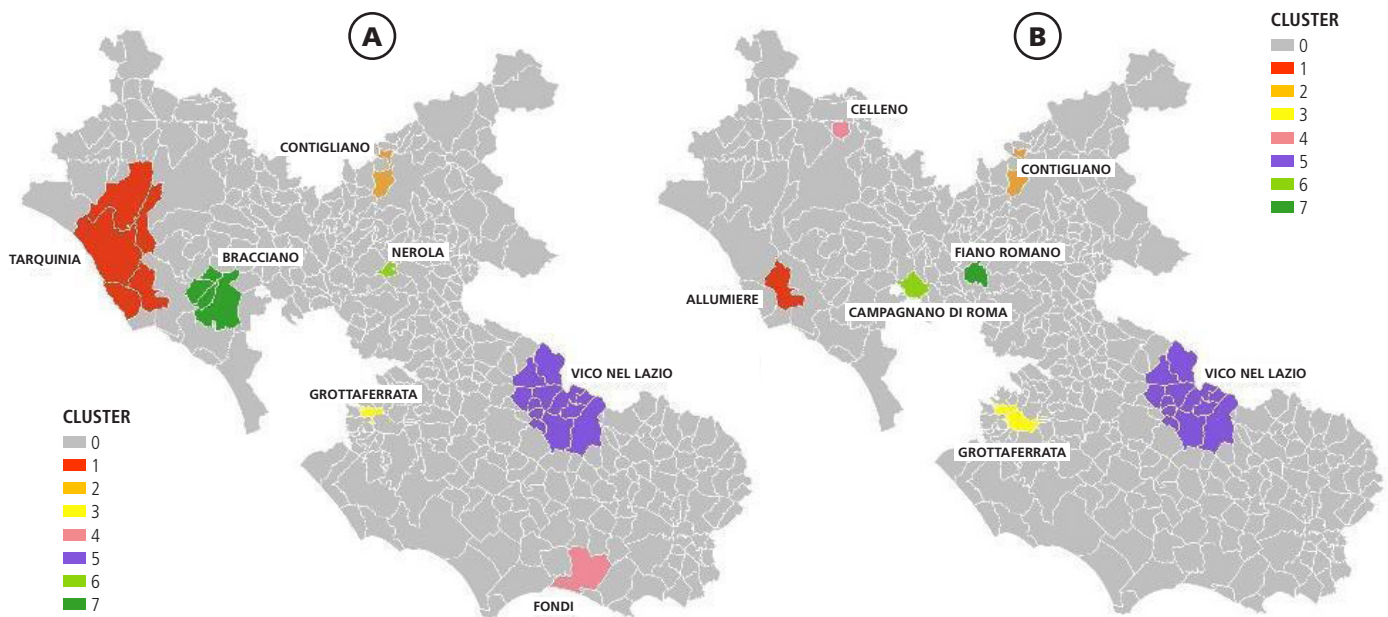


Figura 1. Distribuzione dei cluster per la regione Lazio esclusa Roma: (A) periodo 1 (da inizio epidemia a 6 aprile 2020) e (B) periodo 2 (da inizio epidemia a 27 aprile 2020).
Figure 1. Clusters distribution for the Lazio Region (Central Italy), excluding Rome: (A) period 1 (from the outbreak of epidemic to 6 April 2020) (B) and period 2 (from the outbreak of epidemic to 27 April 2020).

CLUSTER	AREA CENTRALE	N. AREE	COMUNI INCLUSI	POPOLAZIONE 2018	CASI OSSERVATI	P-VALUE
PERIODO 1						
1	Tarquinia	5	Monte Romano, Tarquinia, Tuscania, Allumiere, Civitavecchia	83.330	232	<0,001
2	Contigliano	1	Contigliano	3.884	51	<0,001
3	Grottaferrata	1	Grottaferrata	20.460	77	<0,001
4	Fondi	1	Fondi	39.779	86	<0,001
5	Vico Nel Lazio	11	Alatri, Colleparado, Filetino, Fiuggi, Fumone, Guarcino, Torre Cajetani, Trevi Nel Lazio, Trivigliano, Veroli, Vico Nel Lazio	71.835	121	<0,001
6	Nerola	1	Nerola	1.969	13	<0,001
7	Bracciano	4	Oriolo Romano, Bracciano, Canale Monterano, Manziana	34.902	60	<0,001
PERIODO 2						
1	Allumiere*	1	Allumiere	4.059	13	<0,001
2	Contigliano	1	Contigliano	3.884	68	<0,001
3	Grottaferrata*	2	Grottaferrata, Rocca Di Papa	37.661	159	<0,001
4	Celleno	1	Celleno	1.336	51	<0,001
5	Vico Nel Lazio	11	Alatri, Colleparado, Filetino, Fiuggi, Fumone, Guarcino, Torre Cajetani, Trevi Nel Lazio, Trivigliano, Veroli, Vico Nel Lazio	71.835	161	<0,001
6	Campagnano Di Roma	1	Campagnano Di Roma	11.561	78	<0,001
7	Fiano Romano	1	Fiano Romano	15.688	50	<0,001

Il metodo di Kulldroff e Naragwalla aveva identificato anche un cluster isolato a Sacrofano, questo però non è stato convalidato dall'analisi di sensibilità e, quindi, non viene considerato come cluster. / The Kulldroff and Naragwalla's method had also identified an isolated cluster in Sacrofano, but this was not validated by the sensitivity analysis and therefore it was not considered.
* Cluster incerto: il numero dei comuni inclusi è stato modificato rispetto all'originale in seguito all'analisi di sensibilità. / Uncertain cluster: the number of municipalities included has been changed after the sensitivity analysis.

Tabella 1. Cluster identificati da entrambi i test, periodo 1 (da inizio epidemia a 6 aprile 2020) e periodo 2 (da inizio epidemia a 27 aprile 2020), popolazione residente, casi COVID-19 e p-value per la regione Lazio, esclusa Roma.
Table 1. Clusters identified by both tests, period 1 (from the outbreak of epidemic to 6 April 2020) and period 2 (from the outbreak of epidemic to 27 April 2020), resident population, COVID-19 cases and p-value for the Lazio region excluding Rome.

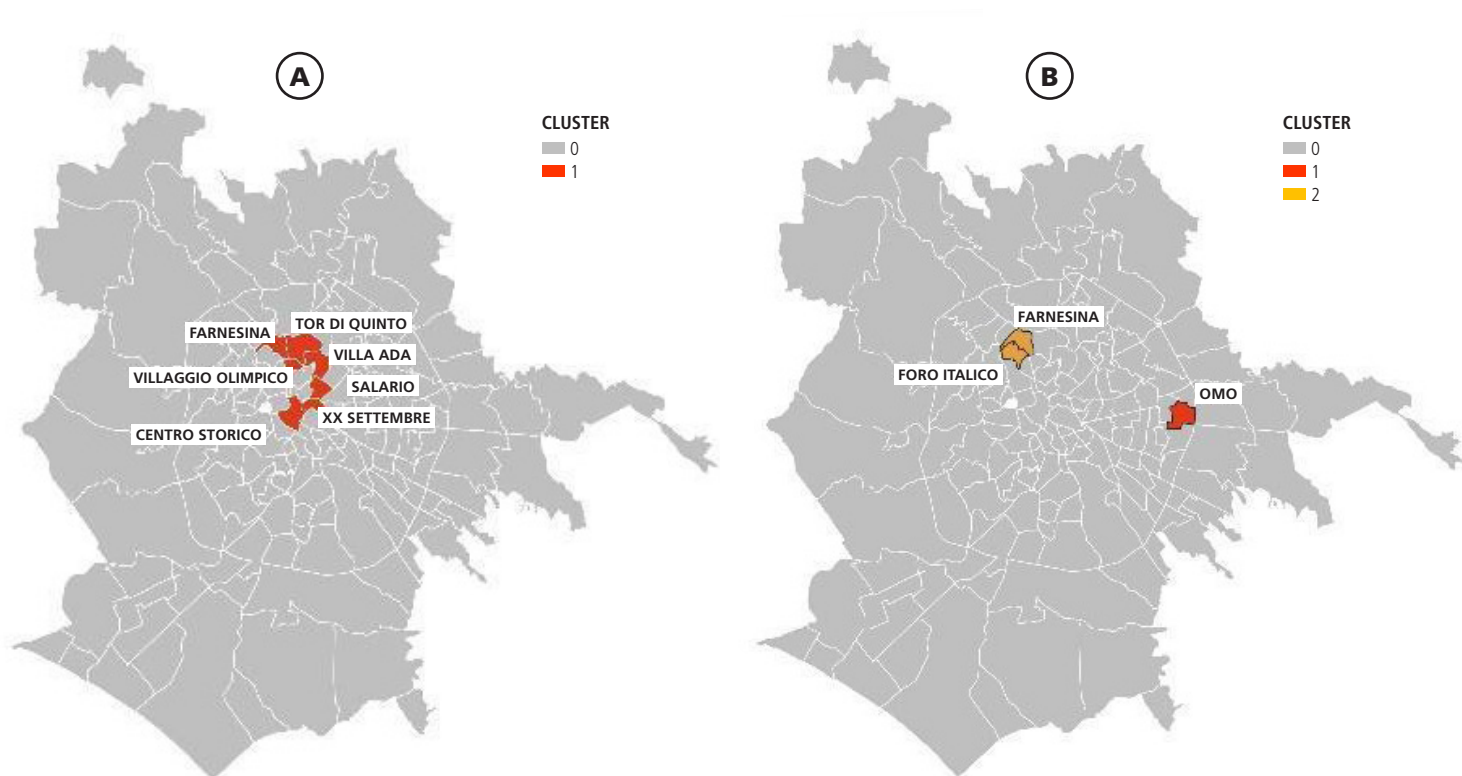


Figura 2. Distribuzione dei cluster per zone urbanistiche, Roma: (A) periodo 1 (da inizio epidemia a 6 aprile 2020) e (B) periodo 2 (da inizio epidemia a 27 aprile 2020).
Figure 2. Clusters distribution by urban areas, Rome: (A) period 1 (from the outbreak of epidemic to 6 April 2020) (B) and period 2 (from the outbreak of epidemic to 27 April 2020).

CLUSTER	AREA CENTRALE	N. AREE	ZUR INCLUDE	POPOLAZIONE 2018	CASI OSSERVATI	P-VALUE
PERIODO 1						
1	Centro storico	7	Centro storico, XX settembre, Villaggio olimpico, Salario, Villa Ada, Tor di quinto, Farnesina	94.474	103	<0,001
PERIODO 2						
1	Omo	1	Omo	1.975	23	<0,001
2	Farnesina	2	Farnesina, Foro italico	19.922	37	<0,001

Tabella 2. Cluster identificati da entrambi i test, periodo 1 (da inizio epidemia a 6 aprile 2020) e periodo 2 (da inizio epidemia a 27 aprile 2020), popolazione residente, casi COVID-19 e p-value per il comune di Roma.

Table 2. Clusters identified by both tests, period 1 (from the outbreak of epidemic to 6 April 2020) and period 2 (from the outbreak of epidemic to 27 April 2020), resident population, COVID-19 cases and p-value for the Municipality of Rome.

rispetto alle zone circostanti, per esempio per la presenza di casi positivi in diverse RSA del territorio.

Nel Comune di Roma, nel primo periodo è stato identificato un unico *cluster* (centro storico) che include 7 ZUR, mentre nel secondo periodo sono stati osservati due *cluster* distinti (Omo e Farnesina), uno localizzato nella periferia Est, dove la prevalenza di immigrati, soprattutto provenienti da Paesi asiatici (per esempio il Bangladesh), è elevata, e l'altro situato in un'area più centrale a Nord della città. Quest'ultimo *cluster* è contiguo al *cluster* Centro storico del primo periodo, in quanto entrambi comprendono la ZUR Farnesina e, quindi, potrebbe essere riconducibile a un'evoluzione temporale del primo *cluster*.

Nel Lazio sono diversi i *cluster* di infezione notificati nelle RSA, in centri di riabilitazione e lungodegenza e case di riposo. Secondo una *survey* condotta dall'ISS, nelle RSA le

principali cause dell'elevato numero di casi di COVID-19 sono stati attribuiti alla mancanza di dispositivi di protezione individuale, alla mancanza di personale sanitario o alla sua scarsa formazione, alla difficoltà nel trasferire i residenti affetti da COVID-19 in strutture ospedaliere, all'impossibilità di far eseguire i tamponi.²⁴ L'approccio proposto consente, solo in parte, di osservare *cluster* nei comuni in cui è noto che sono localizzate alcune di queste strutture (per esempio, *cluster* di Contigliano e Rocca di Papa e Veroli e Nerola).

Questo è tra i primi studi italiani che utilizza la statistica Scan per identificare *cluster* di casi di COVID-19 a livello locale in una regione italiana e in cui viene monitorato l'andamento dei *cluster* epidemici e sono individuati nuovi possibili focolai. Questo approccio è tanto più efficace quanto più i dati sono aggiornati tempestivamente. È importante ricordare che è stato utilizzato un metodo strin-

gente per l'individuazione dei *cluster*, scegliendo un *p-value* molto piccolo, in modo da ridurre il rischio di falsi positivi. Tuttavia, l'accuratezza del metodo si basa su una serie di assunzioni, tra cui la validità dell'informazione sulla residenza/domicilio dei soggetti, che potrebbe non essere aggiornata nel caso dei soggetti delle residenze sanitarie. Inoltre, il metodo utilizza casi di COVID-19 alla diagnosi, assumendo che i giorni precedenti non abbiano influito sulla distribuzione spaziale, ipotesi difficilmente accettabile. In questa applicazione del metodo Scan, la mancata informazione della data di diagnosi per il primo periodo temporale (fino al 26 marzo) non ha consentito di modellare adeguatamente l'interazione spazio-tempo. L'approccio di indagine proposto potrebbe essere più efficiente se nel prossimo futuro saranno resi disponibili tempestivamente i nuovi casi di COVID-19, consentendo di restituire mappe di rischio aggiornate della situazione dell'epidemia nell'area in studio, identificando nuovi *cluster* e monitorando allo stesso tempo i *cluster* identificati precedentemente. La disponibilità di dati di incidenza in tempo reale ha la potenzialità di permettere di seguire l'evoluzione temporale dell'epidemia attraverso le variazioni nella localizzazione e nell'estensione dei *cluster* spaziali. Questo tipo di approccio applicato in modo prospettico potrebbe offrire uno strumento di monitoraggio *real-time* in una specifica area a supporto dei dipartimenti di prevenzione per pianificare il tracciamento dei contatti e altre attività di prevenzione. Inoltre, l'approccio prospettico potrebbe rappresentare la naturale evoluzione del metodo qui utilizzato, consentendo di individuare situazioni critiche, come incrementi rapidi nell'incidenza, focolai localizzati e verificare se le misure di contenimento e di isolamento locali siano state efficaci nel ridurre la diffusione dell'epidemia.⁶ Come suggerito da Rothman,²⁵ la natura descrittiva degli studi di *cluster* non ne permette l'utilizzo per testare ipotesi eziologiche, mentre potrebbero rappresentare la base per pianificare uno studio coorte o caso-controllo.

Uno dei principali problemi nell'interpretazione dei risultati di questo studio è che dipendono dalla capacità del sistema di sorveglianza di identificare i casi di infezione da

COVID-19, riflettendo le diverse strategie messe in atto a livello locale (ASL) per l'identificazione dei casi. La ridotta disponibilità dei test diagnostici, in concomitanza alla rapida diffusione del virus, con conseguente progressivo incremento delle richieste di tamponi faringei, può avere determinato all'inizio dell'epidemia la necessità di individuare criteri di priorità per l'esecuzione dei tamponi ed è necessario tener conto che tali criteri possono essersi modificati nel tempo. Si sottolinea che il numero dei casi positivi notificati rappresenta molto probabilmente una sottostima dell'effettiva ampiezza del contagio da Coronavirus.

Secondo il National Center For Environmental Health (CDC) non esiste ancora un test ottimale per individuare possibili aggregazioni spaziali, per questa ragione sono state applicate due tecniche Scan per verificare la presenza di *cluster* localizzati sul territorio, scegliendo tra le più versatili e potenti: Kulldorff&Nagarwalla e Besag&Newell.^{26,27} Questa considerazione ha portato ad assumere un approccio più conservativo, al fine di ottenere risultati più consistenti, considerando la presenza di un *cluster* localizzato laddove il test di K&N viene confermato anche dall'analisi di sensibilità effettuata con il test di B&N.²⁰

Oltre all'analisi spaziale, quella dei contatti tra individui rappresenta sicuramente un elemento importante per identificare le aree e i sottogruppi di popolazione a elevato rischio di diffusione dell'epidemia. È fondamentale che all'analisi di *cluster* spaziali si affianchi anche un'analisi in grado di monitorare e analizzare il tipo di interazioni sociali che hanno portato alla diffusione del COVID-19 sul territorio, come per esempio una *social network analysis*. L'interazione di questi diversi approcci di analisi può rappresentare un valido supporto alla sanità pubblica, consentendo di disporre di strumenti efficienti per il monitoraggio a livello regionale e locale della diffusione del rischio di contagio, al fine di rafforzare la capacità di gestione della fase due dell'emergenza.

Conflitti di interesse dichiarati: nessuno.

Data di sottomissione: 18.06.2020.

Data di accettazione: 11.08.2020.

BIBLIOGRAFIA

1. European Centre for Disease Prevention and Control. Covid-19 pandemic. Stoccolma, ECDC, 2020. Disponibile all'indirizzo: <https://www.ecdc.europa.eu/en/covid-19-pandemic>
2. Dipartimento della protezione civile. Emergenza Coronavirus: la risposta nazionale. Protezione civile, Roma, 2020. Disponibile all'indirizzo: <http://www.protezionecivile.gov.it/attivita-rischi/rischio-sanitario/emergenze/coronavirus>
3. Istituto Superiore di Sanità. Coronavirus - SARS-CoV-2. Roma, ISS, 2020. Disponibile all'indirizzo: www.epicentro.iss.it/coronavirus
4. Organizzazione Mondiale della Sanità. WHO coronavirus (Covid-19) dashboard. Ginevra, OMS, 2020. Disponibile all'indirizzo: <https://who.sprinklr.com/>
5. Johns Hopkins University. Coronavirus resource center. Baltimore, Maryland, USA, JHU, 2020. Disponibile all'indirizzo: <https://coronavirus.jhu.edu/>
6. Paula Moraga. SpatialEpiApp: a shiny web application for the analysis of spatial and spatio-temporal disease data. Lancaster, Lancaster University, UK, 2020. Disponibile all'indirizzo: <https://cran.r-project.org/web/packages/SpatialEpiApp/vignettes/manual.pdf>
7. Zoumpikas T. Covid-19 cluster analysis - Towards data science. Medium 2020. Disponibile all'indirizzo: <https://towardsdatascience.com/covid-19-cluster-analysis-405ebbd10049>
8. Yong SEF, Anderson DE, Wei WE et al. Connecting clusters of Covid-19: an epidemiological and serological investigation. *Lancet Infect Dis* 2020;20(7):809-15.
9. Park JY. Spatial visualization of cluster-specific Covid-19 transmission network in South Korea during the early epidemic phase. *MedRxiv* 2020. doi: 10.1101/2020.03.18.20038638
10. Kang D, Choi H, Kim JH, Choi J. Spatial epidemic dynamics of the Covid-19 outbreak in China. *Int J Infect Dis* 2020;94:96-102.
11. Mizumoto K, Dahal S, Chowell G. Spatial variability in the risk of death from Covid-19 in Italy. *Int J Tuberc Lung Dis* 2020;24(8):829-37.
12. Kulldorff M. A spatial Scan statistic. *Commun Statist Theory Meth* 1997;26(6):1481-96.
13. Kulldorff M. Prospective time-periodic geographical disease surveillance using a scan statistic. *J R Stat Soc Series A* 2001;164(1):61-72.

14. Desjardins MR, Whiteman A, Casas I, Delmelle E. Space-time cluster and co-occurrence of Chikungunya and Dengue Fever in Colombia from 2015 to 2016. *Acta Trop* 2018;185:77-85.
15. Owusu C, Desjardins MR, Baker KM, Delmelle E. Residential mobility impacts relative risk estimates of space-time clusters of Chlamydia in Kalamazoo county, Michigan. *Geospat Health* 2019;14(2).
16. Whiteman A, Desjardins MR, Eskildsen GA, Loaiza JR. Detecting space-time clusters of Dengue fever in Panama after adjusting for vector surveillance data. *PLoS Negl Trop Dis* 2019;13(9):e0007266.
17. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Stat Med* 1995;14(8):799-810.
18. Riccardo F, Andrianou X, Bella A et al. Epidemia Covid-19: aggiornamento nazionale al 21 luglio 2020. Roma, ISS, 2020. Disponibile all'indirizzo: https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-Covid-19_21-luglio-2020_appendix.pdf
19. Tango T. *Statistical methods for disease clustering - Statistics for biology and health*. New York, Springer, 2010.
20. Schifano P, Leone M, Michelozzi P. Cluster Statistical Analysis in Epidemiology. *Epidemiol Prev* 2016;40(5) Suppl 2:55-64.
21. Kulldorff M, Tango T, Park P. Power comparisons for disease clustering tests. *Comput Stat Data Anal* 2003;42(4):665-84.
22. Tango T. A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Stat Med* 1995;14(21-22):2323-34.
23. Besag J, Newell J. The detection of clusters in rare diseases. *J R Stat Soc Series A* 1991;154(1):143-55.
24. Istituto superiore di sanità. *Survey nazionale sul contagio Covid-19 nelle strutture residenziali e sociosanitarie*. Roma, ISS, 2020. Disponibile nell'indirizzo: <https://www.epicentro.iss.it/coronavirus/pdf/sars-cov-2-survey-rsa-rapporto-3.pdf>
25. Rothman KJ. A sobering start for the cluster busters' conference. *Am J Epidemiol* 1990;132 (Suppl 1):S6-13.
26. National center for environmental health. Guidelines for investigating clusters of health events. *MMWR Recomm Rep* 1990;39(RR-11):1-23.
27. National center for environmental health. Investigating suspected cancer clusters and responding to community concerns. Guidelines from CDC and the Council of State and territorial epidemiologists. *MMWR Recomm Rep* 2013;62(RR-8):1-28.