

PROPOSTA DI UNA INDAGINE A CAMPIONE PER UNA STIMA AFFIDABILE DEI PARAMETRI FONDAMENTALI DELLA EPIDEMIA DA SARS-CoV-2

1. PREMESSA E MOTIVAZIONE

L'attuale emergenza per il controllo della diffusione del SARS-CoV-2 impone una analisi accurata delle fonti di dati sulle quali basare la stima dei principali parametri epidemiologici e quindi la previsione delle future evoluzioni per poter suggerire interventi di politica sanitaria efficaci.

Proprio in questa fase di grande incertezza sulla dinamica di un fenomeno nuovo risulta cruciale la sua misura in modo quanto più possibile accurato. L'incertezza rimarrà, è ineludibile, ma il livello di imprecisione nel monitoraggio dell'evoluzione può e deve essere ridotto e tenuto sotto controllo.

I dati sui tamponi finora raccolti sono basati su un campionamento 'di convenienza' che ha privilegiato l'esame dei casi che manifestavano sintomi. Come è noto, un campione di convenienza non consente di produrre una stima probabilistica non distorta e con un prefissato livello di accuratezza. Come è stato messo in evidenza da diversi studi, nei dati raccolti finora risulta sottostimato il numero di contagiati essendo assenti gli asintomatici, con una conseguente sovrastima del tasso di letalità. Anche i dati raccolti a tappeto di recente dalla regione Veneto non rappresentano un campione probabilistico. Raccogliendo le osservazioni davanti ai supermercati, infatti, si sovra-includono le persone sane e si sotto-includono quelle che, manifestando sintomi o essendo più vulnerabili, sono confinate a casa.

Una conoscenza chiara delle dinamiche in atto ci sembra necessaria per le misure da assumere da parte dei responsabili politici e sanitari oltre che per la consapevolezza e i comportamenti da adottare da parte della popolazione. E' necessario un impianto di rilevazione che consenta di produrre stime non distorte e di fare confronti statisticamente significativi nel tempo e tra diverse aree geografiche, tenendo conto dei differenti contesti economici, demografici, sociali, ambientali e culturali.

Non ci si può limitare a misurare in modo accurato il numero dei contagi, dei ricoveri e dei loro esiti e i risultati dei tamponi, ma occorre fare luce sulle caratteristiche individuali, familiari e di contesto che possano favorire o ostacolare l'infezione e stimare correttamente l'effetto degli interventi adottati per modificare l'evoluzione del fenomeno. Un lavoro da fare insieme, epidemiologi, virologi, responsabili di strutture sanitarie, esperti di progettazione di indagini e di modelli statistici di analisi ed esperti di valutazione di politiche pubbliche.

Per ricavare stime affidabili e quindi realmente utili riteniamo sia indispensabile, compatibilmente con l'impegno di forze attualmente dispiegato per il controllo dell'emergenza, progettare e realizzare un protocollo di osservazione a campione riferito all'intera popolazione italiana, seguendo un disegno campionario rigoroso. Sul campione dovrebbe essere effettuato il tampone.

Questo consentirebbe di valorizzare la informazione preziosa raccolta fino ad ora da parte delle strutture sanitarie secondo le indicazioni delle istituzioni preposte, integrandola all'interno di un quadro complessivo che contenga la stima accurata delle variabili d'interesse in domini temporali e spaziali individuati opportunamente, in modo da poter monitorare efficientemente l'evoluzione del contagio e valutare gli effetti di azioni e interventi.

L'uso integrato di diverse fonti – fra le quali, in primo luogo, quelle amministrative nel corso dell'accoglienza e della cura delle persone che si rivolgono al sistema sanitario, quelle statistiche progettate per misurare con accuratezza la diffusione del contagio, quelle fornite da nuove fonti idonee a tracciare gli spostamenti delle persone e i loro contatti - rappresenta una necessità per poter fornire dati di qualità, non distorti e non fuorvianti, sui quali fondare decisioni importanti per la salute pubblica. È una sfida metodologica, tecnologica e organizzativa alla quale la comunità degli statistici, e certamente degli statistici ufficiali, può dare un contributo utile.

* * * * *

Quanto illustrato di seguito ha, dunque, la finalità di proporre un protocollo osservazionale che consenta di stimare i contagiati da SARS-CoV-2 secondo diverse categorie di gravità, con affidabilità definita a livello temporale e territoriale.

A partire dalla popolazione di cui si è accertato l'avvenuto contagio (conclamati) l'obiettivo è quello di stimare la popolazione contagiata ma non ancora diagnostica (asintomatica).

Ai fini della procedura proposta, in ciascun momento temporale, vengono preliminarmente individuati due sottoinsiemi di individui che denominiamo: *target A* e *target B*:

Target A. È il sottoinsieme che comprende le persone il cui stato di infezione sia stato conclamato (che possono essere ricoverati o in quarantena coatta) e tutte le persone che hanno avuto contatti con loro risalenti fino a 14 giorni prima; si tratta, pertanto, dell'insieme di persone prevedibilmente contagiate, e non solo di quelle per le quali il contagio sia già stato determinato. In sostanza, sia la dimensione emersa sia quella sommersa del fenomeno.

Target B. È il sottoinsieme residuo di tutte le persone non entrate in contatto con persone del *Target A*. Questo sottoinsieme comprende quindi persone sane, i cosiddetti *silenti* e le persone in fase di incubazione per le quali i sintomi si manifesteranno successivamente, nell'arco di massimo 14 giorni.

Le stime relative ai due sottoinsiemi possono essere ottenute sulla base di un'osservazione continua nel tempo e seguendo due metodologie distinte, entrambe basate sul campionamento *indiretto*, lo stesso che si usa per la stima di popolazioni *rare* o *elusive* (o per le popolazioni animali).

Qui di seguito diamo alcune idee su come realizzare l'osservazione campionaria per ciascuno dei due gruppi.

2. CAMPIONAMENTO PER IL TARGET A

Lo schema di campionamento, continuo nel tempo, si articola nelle seguenti fasi:

- A1. selezione per ogni unità di tempo (giorno, settimana, altro) di un campione di conclamati;
- A2. ricostruzione di tutti i contatti, risalenti fino a 14 giorni prima, delle persone selezionate al punto A1; per semplificare questa fase, si potrebbero utilizzare apposite App disponibili su smartphone per il tracciamento degli individui;
- A3. somministrazione dei tamponi a una quota prestabilita delle persone che hanno avuto contatti con il campione selezionato al punto A1.

Il campionamento nella fase A1 dovrebbe essere realizzato tenendo conto delle dimensioni temporali e spaziali.

- ✓ **Campionamento spaziale.** Si realizza selezionando un campione panel di strutture sanitarie (in base alla loro dimensione e altre caratteristiche). L'utilizzo di tecniche di campionamento spaziale permetterebbe di avere un campione ottimale da vari punti di vista;
- ✓ **Campionamento temporale.** Nelle strutture sanitarie selezionate nel campionamento spaziale, sarebbe successivamente selezionato un campione di persone conclamate di cui ricostruire i contatti. L'individuazione del campione potrebbe essere realizzata per intervalli di tempo definiti di accesso alla struttura (ricovero, altro). Ad esempio potrebbe essere selezionata una persona tra i conclamati che in un intervallo di due ore hanno avuto accesso alla struttura sanitaria.

Ai fini della stima statistica del numero di persone contagiate in un dato dominio di stima *territoriale* (territorio nazionale/specifica area geografica come, ad esempio, una regione) e *temporale* (settimana/giorno) potrebbe essere sufficiente coinvolgere nell'indagine dello specifico dominio (*spazio-temporale*) circa 1.000 persone tra i contatti delle persone conclamate su cui effettuare i tamponi. Tale dimensione campionaria, assicurerebbe una stima affidabile con un errore relativo di campionamento inferiore al 5% qualora la proporzione dei contagiati per il target fosse intorno al 20%.

Presupponendo circa 25 contatti per ogni conclamato, le persone a cui somministrare i 1.000 tamponi potrebbero essere individuate, selezionando:

- ✓ circa 200 persone conclamate (di cui si ricostruirebbero $5.000 = 200 \times 25$ contatti);
- ✓ campionando una quota di circa il 20% dei contatti di cui sopra.

Nota 1. È chiaro che ai fini sanitari, per la prevenzione del contagio sarebbe opportuno seguire tutti i contagiati. Tuttavia, da un punto di vista statistico, per ottenere stime di elevata qualità sul numero di persone contagiate questo non è necessario.

3. CAMPIONAMENTO PER IL GRUPPO B

Per la stima del numero di contagiati al di fuori dei contatti delle persone conclamate deve essere selezionato un *panel* di individui che dovranno essere seguiti in modo continuo nel tempo. Le fasi operative da eseguire sono le seguenti:

- B1. il panel viene sottoposto regolarmente al tampone (ad esempio una volta a settimana);

B2. se una persona del panel risulta positiva, sono ricostruiti tutti i suoi contatti negli ultimi 14 giorni;

B3. Il tampone viene sottoposto anche a un campione di questi contatti.

Il numero di persone da coinvolgere nel *panel* potrebbe essere di circa 1.000 (per circa 1.200 tamponi) per un dato dominio di stima *territoriale* e *temporale*, garantendo una stima affidabile con un errore relativo di campionamento inferiore al 5% qualora la proporzione dei contagiati per il target fosse intorno al 4%.

Il *panel* potrebbe essere formato con una metodologia in due fasi, in modo da selezionare mediante un *pre-screening*:

- ✓ una quota di persone che continuano a spostarsi (e quindi maggiormente soggette a contagio);
- ✓ una quota di persone con pochi contatti che seguono essenzialmente le medesime prescrizioni di quanti posti in quarantena.

La stima delle intersezioni tra A e B può essere ottenuta sulla base delle sovrapposizioni tra i campioni dei due target.

Roma, 26 marzo 2020

Giorgio Alleva, ordinario di statistica, Sapienza Università di Roma, ex Presidente dell'Istat

Giuseppe Arbia, ordinario di statistica economica, Università Cattolica di Roma, esperto di analisi spaziale dei dati

Piero Demetrio Falorsi, ex direttore della direzione metodologica dell'Istat, esperto di disegni di campionamento

Guido Pellegrini, ordinario di statistica economica, Sapienza Università di Roma, esperto di valutazione di politiche pubbliche e Presidente della Commissione di garanzia dell'informazione statistica

Alberto Zuliani, professore emerito di statistica, già Presidente dell'Istat