



FORUM
DISUGUAGLIANZE
DIVERSITÀ

Progetto di rilevazione campionaria dinamica per affrontare il Sars-Cov-2 e il Covid-19

di Gianluigi Bovini, Manlio Calzaroni, Piero De Chiara, Mauro Masselli*

Sintesi

La nota presenta una bozza di progetto per una rilevazione campionaria dinamica che aiuti a contrastare il Sars-Cov-2 (il virus) e la Covid-19 (l'infezione polmonare) misurando: la diffusione virale, le conseguenze sanitarie e l'efficacia delle misure di contenimento.

In molti paesi europei sono state avanzate proposte per affiancare ai dati puntuali raccolti dalle autorità sanitarie un panel rappresentativo dell'intera popolazione, per stimare con test ripetuti l'effettiva diffusione del contagio anche presso i non sintomatici e chi presenta sintomi trascurabili e non viene analizzato con le informazioni oggi disponibili.

Anche in Italia, dopo una prima proposta del Forum Disuguaglianze Diversità, si sta coagulando la spinta in questa direzione, che ha trovato forza grazie in particolare alla "Proposta per un'indagine campione" (Alleva, Arbia, Falorsi, Pellegrini, Zuliani). Nell'auspicio che gli sforzi convergano, questa nota costituisce un ulteriore approfondimento della proposta succitata; gli obiettivi sono sostanzialmente gli stessi mentre le metodologie differiscono leggermente, ma essendo compatibili possono trovare una sintesi proficua che tenga conto degli "obiettivi" fissati dalla Autorità Politica e del confronto con gli esperti dei differenti settori chiamati in causa dalle proposte.

Nella prima parte si distinguono gli obiettivi della rilevazione, tra quelli di breve periodo (contributo al contenimento del contagio) e quelli di medio periodo (correlazione con le misure adottate, informazioni per la scelta e il monitoraggio di riaperture selettive nel tempo, sentinella sulla ricomparsa di focolai, strumentazione per future emergenze.)

Nella seconda parte si affrontano problemi tecnici quali: il reclutamento, la numerosità, rappresentatività e rotazione del campione; la modalità e frequenza dei test ripetuti sugli stessi soggetti.

Nella terza parte si fa riferimento ad altre indagini sinergiche: mortalità differenziale rispetto agli anni precedenti (che è già in via di espletamento da parte ISTAT_ISS) e indagini su sub-popolazioni particolari. Si valuta infine la opportunità di integrare, sulla base delle normative vigenti, le fonti

amministrative con altre fonti private, compresi i dati comportamentali in possesso dei principali operatori internet per analisi mirate sugli affetti da Covid-19.

Premessa

L'attuale, elevata incertezza sui dati relativi alla reale consistenza delle persone contagiate dal Sars-Cov-2 rappresenta un grave problema per assumere decisioni politiche, che tengano in considerazione in modo adeguato nel breve e medio periodo i trade-off fra le conseguenze sanitarie dell'epidemia e gli effetti sociali ed economici delle misure adottate. Per ridurre questa incertezza appare indispensabile realizzare un sistema informativo, incentrato su una indagine campionaria panel, in grado di produrre dati sulla diffusione virale, sulle conseguenze sanitarie e sull'efficacia delle misure di contenimento.

La lotta contro Sars-Cov-2 e Covid-19, infatti, non può prescindere dall'utilizzo di informazioni tempestive, complete ed affidabili. Mentre per il gruppo degli "ospedalizzati" cioè delle persone colpite dall'infezione Covid-19 le informazioni di base disponibili possono ritenersi soddisfacenti anche se ulteriormente migliorabili, quelle relative allo stato di salute dell'intera popolazione, ovvero delle persone non colpite dal virus Sars-Cov-2, degli asintomatici e dei pauci sintomatici non sono disponibili, se non talvolta a livello locale o utilizzando stime non fondate su dati certi. Questa mancanza di informazione limita l'uso dei dati degli ospedalizzati per calcolare indicatori significativi per l'intera popolazione e la velocità e la dimensione della diffusione del virus.

Conoscere e comunicare questi dati in tempo quasi reale è il primo servizio che lo Stato deve fornire in questa situazione, per ridurre l'ansia e il rischio di misure e comportamenti non informati, troppo blandi o troppo duri.

L'analisi dei soli dati oggi disponibili, inoltre, è soggetta a un elevato numero di anomalie e sbalzi inspiegabili. Mentre è impedita la confrontabilità con le dinamiche e le esperienze degli altri Paesi, a causa delle diverse pratiche nella somministrazione di tamponi. A tale riguardo un obiettivo, anche se non a breve termine, è prevedere la condivisione su scala almeno europea di un progetto di rilevazione campionaria, con metodologie analoghe che rendano possibile la comparazione dei dati.

In una prima fase è però utile partire dall'Italia, per le evidenti ragioni di urgenza.

Per ovviare alla mancanza informativa si propongono allora un'indagine campionaria panel (con rotazione) sulla popolazione ed elaborazioni che, insieme alle fonti già esistenti o attivabili e agli studi epidemiologici, possano costituire un "sistema informativo integrato" in grado di fornire dati per il governo dell'epidemia.

Per quanto riguarda la rilevazione, la numerosità del campione e la periodicità della rilevazione, esse dovranno essere calibrate non solo sull'affidabilità attesa delle stime, ma anche sulle risorse organizzative necessarie e sull'intervallo di tempo per la replicazione dell'indagine. Saranno inoltre condizionate dall'incidenza reale della sottopopolazione dei contagiati sul totale della popolazione, che ad oggi è incognita ed è quindi il principale parametro target dello studio. Sarà infine decisiva la scelta di quale test utilizzare per pervenire alla misurazione delle variabili chiave.

La Bozza di progetto qui elaborata dovrà evidentemente trovare il confronto con le Autorità sanitarie, al fine di risolvere diverse questioni aperte e valutare la sostenibilità dell'intervento nelle attuali condizioni.

1. Obiettivi

L'indagine unitamente alle altre fonti svolgerà il ruolo sia di "sentinella" sia di base informativa per il monitoraggio della dimensione e della diffusione del virus. Gli obiettivi possono essere così riassunti:

- 1) Stimare le modalità di diffusione
- 2) Stimare il numero effettivo dei contaminati e dei tassi di guarigione/asintomaticità e letalità
- 3) Valutazione dell'impatto di misure restrittive alternative
- 4) Valutare i tempi di attenuazione/rimozione dei provvedimenti
- 5) Prepararsi alla prossima epidemia/pandemia

- 1) Un primo obiettivo è *stimare le modalità di diffusione*. Per quanto il campione sia limitato a poche decine di migliaia di individui, esso può fornire informazioni di interesse generale sulle modalità di diffusione, perché consente di avere informazioni sulle caratteristiche socio demografiche, livelli territoriali e sui tempi del contagio delle differenti "sub-popolazioni" di interesse che permetteranno di identificare situazioni e comportamenti a differenti livelli di rischio. Il campionamento consente anche di porre obiettivi sanitari più complessi. A titolo esemplificativo, si pensi alla possibilità di individuazione di soggetti negativi al tampone, in precedenza asintomatici, che hanno contratto il virus e sono guariti in modo inconsapevole. Questa condizione si potrebbe rilevare con un ulteriore esame (quello degli anticorpi), che potrebbe eventualmente essere proposto a una frazione del campione risultata negativa al test, per potere stimare l'effettiva diffusione del contagio e la conseguente, progressiva formazione dell'indennità di gregge.
- 2) Secondo obiettivo strategico è la *stima del numero effettivo dei contaminati e dei tassi di guarigione/asintomaticità e letalità*. I test effettuati solo sui pluri-sintomatici e sulle categorie a rischio, per quanto applicati con criteri più o meno ampi e stabili, non consentono questa stima, che può essere garantita solo con test universali o, come si propone, su un campione di alcune decine di migliaia di individui rappresentativi dell'intera popolazione.
- 3) Terzo obiettivo è indagare le correlazioni con le misure restrittive adottate e fornire elementi di *valutazione dell'impatto di misure restrittive tra loro alternative*. Il concetto di distanziamento sociale è una nebulosa composta da provvedimenti indipendenti ed eterogenei. La chiusura delle scuole che distanzia coetanei, i funerali proibiti o consentiti in forma privata, le fabbriche e gli uffici chiusi o aperti, le code ai supermercati, le passeggiate di salute o con i cani, la frequenza di treni e metropolitane, hanno probabilmente effetti diversi sulla diffusione del contagio. Non siamo oggi in grado di valutarlo, ma con una indagine campionaria possiamo avere una base per farlo, per applicare strumenti predittivi (anche di Intelligenza artificiale, purché spiegabili) e per consentire discussioni e decisioni più informate.
- 4) Le considerazioni di cui al punto precedente possono sembrare premature nella fase di picco che stiamo attraversando, in cui prevale la spinta a un indurimento progressivo delle

restrizioni. Ma una volta superata la fase acuta, ciò che presumibilmente avverrà in modo asincrono nel territorio nazionale, le misurazioni proposte potranno consentire di *valutare i tempi di attenuazione/rimozione dei provvedimenti*. Ciò potrà su queste basi avvenire in modo diversificato per fascia di età e zona geografica ed anche su provenienza internazionale, in modo da tenere sotto controllo eventuali ritorni del virus reintrodotto da altri Paesi in periodi successivi. Inoltre, all'uscita definitiva dalla crisi sanitaria, il campione può essere ridimensionato, ma riutilizzato per predire e misurare gli effetti di alcune delle misure economiche e sociali da usare nella ricostruzione.

- 5) Infine non illudiamoci che questa sarà l'ultima epidemia. Gli errori e gli strumenti che avremo sperimentato durante in il Covid-19, compreso questo Campione, saranno *patrimonio prezioso per affrontare la prossima*.

2. Rilevazione campionaria panel con rotazione

Com'è noto i vantaggi dell'utilizzo di un campione probabilistico consistono nell'ottenere informazioni con costi relativamente limitati e "rappresentativi" dell'intera popolazione. Il campione consente di testare la diffusione del virus sull'intera popolazione, mentre gli attuali dati risentono del fatto che i tamponi sono fatti in prevalenza ai sintomatici o alle categorie più esposte. Basandosi solo sui dati puntuali disponibili scompaiono gli immuni, gli asintomatici e la maggior parte dei pauci sintomatici. Il numero dei contagiati risulta fortemente sottostimato e il tasso di morbilità e letalità limitato¹ alla sola popolazione colpita dall'infezione Covid-19 dando quindi una informazione distorta a livello di comunicazione di massa. Gli svantaggi dell'analisi campionaria consistono in possibili distorsioni nella composizione del campione e nell'errore statistico che è inversamente proporzionale all'ampiezza della base campionaria.

Obiettivo generale dell'indagine campionaria è la produzione corrente di stime relative agli aggregati relativi ai contagiati sintomatici, ai pauci sintomatici, agli asintomatici e quindi del calcolo del tasso di letalità e dei tassi di morbilità per genere, età e territorio; le principali categorie socio-demografiche dovrebbero essere rappresentative almeno su base nazionale e pluriregionale (Nord-ovest, Nord-est, Centro, Sud e isole).

Il disegno della rilevazione proposta è un'indagine campionaria "panel" ovvero la rilevazione ripetuta in più "occasioni" di indagine sulle stesse persone ad intervalli di tempo predeterminati; una quota delle persone rilevate viene sostituita ad intervalli prestabiliti in modo da rinnovare il campione iniziale (rotazione del campione), limitare il peso della rilevazione sulle unità "intervistate" e ridurre il rischio di distorsioni dovuti a comportamenti derivanti dall'appartenenza al panel. Tale disegno campionario è utilizzato e particolarmente adatto per la stima delle variazioni fra tempi successivi poiché, rispetto a campioni indipendenti, riduce l'errore di campionamento delle stime delle variazioni e permette la loro analisi anche a livello individuale invece che solo a livello aggregato (uno schema semplificato è riportato in APPENDICE).

La popolazione di riferimento è quella definita dalla presenza nell'anagrafe della popolazione (che contiene anche informazioni su età e sesso degli individui, l'indirizzo e la composizione della

¹ https://www.scienzainrete.it/articolo/ma-quant-sono-gli-ammalati-di-covid-19-italia/simone-franchini-enzo-marinari-giorgio?fbclid=IwAR06SpK6Nt-fsk1K0dqUvRh4R-2Uubub_qiLvwbDT-GvQMxzAXJTISsXaQ7Q

famiglia). I registri anagrafici assicurano la copertura della popolazione di nazionalità italiana e di quella di nazionalità estera “regolare”.

L’unità di rilevazione è l’individuo; a parità di dimensione campionaria permette una maggiore dispersione territoriale ed evita il fattore cluster ².

Diversamente da altre indagini il reclutamento del campione non pare particolarmente difficile. È verosimile che l’adesione volontaria sia molto ampia, sia per motivi altruistici, sia per il beneficio percepito di essere sottoposti con frequenza a test e monitorati. Non è quindi necessario ipotizzare alcun coinvolgimento obbligatorio; ci si può anzi attendere che la disponibilità sia tanto ampia da consentire ai selezionatori la formazione di una riserva che consente di applicare criteri di selezione e rotazione ancor più cautelativi di quelli raccomandati dall’Istat. Inoltre al momento dell’adesione volontaria al panel è più agevole raccogliere il consenso relativo al trattamento dei dati per fini e tempi limitati. Questo consenso informato è strumento necessario e sufficiente per un efficace utilizzo dei dati anche sulla base della normativa sulla privacy vigente, a maggior ragione quando si otterrà l’accesso alla portabilità dei dati comportamentali in possesso dei principali operatori internet, per gli utilizzi indicati alla fine di questa proposta.

L’auspicato alto tasso di adesione semplifica anche problemi classici quali il reperimento effettivo presso la residenza ufficiale delle persone selezionate (in un numero non trascurabile di casi, soprattutto nella fascia più giovane, non esiste coincidenza fra il domicilio effettivo e la residenza anagrafica) e consente la formazione di adeguate liste di sostituzione, nel rispetto delle quote campionarie, con modalità che non compromettano la tempestività della rilevazione.

Nella formazione del campione verranno incluse persone che sono già state in precedenza sottoposte al test. Nella parte sanitaria del questionario bisognerà rilevare anche questa notizia, con la precisazione degli esiti del precedente test.

La rilevazione delle informazioni viene effettuata mediante due strumenti:

- il test che può essere effettuato o mediante tampone, come indicato nella Proposta Alleva e altri o mediante prelievo di sangue per l’individuazione di anti-corpi del Sars-Cov-2. Dal punto di vista organizzativo e statistico, la scelta tra metodi andrebbe effettuata considerando il livello di precisione, ovvero la quota di falsi positivi e negativi prodotti dai due sistemi, la velocità e la semplicità sia nella somministrazione sia nelle analisi di laboratorio da effettuare, il contenuto informativo del test (ad esempio se in grado di individuare un soggetto esposto al virus nei mesi precedenti – “i guariti”). Questi parametri devono essere attentamente valutati; ad esempio può essere conveniente optare per uno strumento che consenta una maggiore facilità di somministrazione, una maggiore velocità nei risultati e la possibilità di non gravare solo sulle strutture pubbliche coinvolgendo anche le strutture accreditate dal SSN più capillarmente diffuse sul territorio, seppure con un livello di precisione leggermente inferiore. In questo senso, considerando gli obiettivi conoscitivi dell’indagine, potrebbe essere preferito l’utilizzo di test basati sugli anticorpi. In ogni caso le scelte tecniche devono essere connesse alla realtà e alle

² Tuttavia potrebbe essere ipotizzato la rilevazione dei componenti della famiglia in particolari condizioni; ad esempio nel caso di individuo asintomatico oppure nel caso di presenza di “badanti” spesso non incluse nella scheda anagrafica ma che vivono a stretto contatto con persone anziane campionate.

indicazioni delle Istituzioni che devono essere direttamente coinvolte nel disegno delle indagini".

- una indagine CATI (a zero rischio di epidemia) per individuare le possibili cause di contagio (viaggi, partecipazione a eventi di massa, rispetto o meno delle misure restrittive, ecc.) non limitato all'individuo ma anche agli altri componenti della famiglia. L'indagine CATI è uno strumento flessibile i cui contenuti possono essere cambiati per adeguarli allo stato dell'epidemia.

Le unità positive ad una rilevazione non verranno sottoposte al test nelle rilevazioni successive. Tale sotto-campione viene considerato nel panel per la procedura di stima e potrà essere utilizzato per ulteriori indagini mediante CATI (ad esempio sui comportamenti) o per programmare studi epidemiologici (ad es tentando di ricostruire i contatti nel periodo di incubazione e sottoporre a test tutti o un campione di contatti). Su tali unità inoltre sarà possibile aggiungere informazioni rilevanti tratte dai Big Data, ad esempio gli spostamenti mediante i gestori telefonici, analisi comportamentali da Google e Facebook etc. (vedi più avanti il § 3),

La numerosità del campione deve essere calibrata non solo sull'affidabilità attesa delle stime ma anche soprattutto sulle risorse organizzative necessarie:

- la disponibilità dei "rilevatori" o delle strutture dove effettuare il test;
- la disponibilità dei laboratori di analisi del test;
- l'intervallo di tempo per la replicazione dell'indagine a sua volta dipendente anche da considerazioni sanitarie;
- la disponibilità di una struttura di supporto a livello centrale e locale;
- il livello di significatività territoriale che si vuole ottenere.

L' IFW di Kiel ha proposto una base campionaria di 10.000 individui con test ripetuti ogni 5 giorni³. La dimensione campionaria dipende dall'incidenza "reale" della sottopopolazione dei contagiati sul totale della popolazione che, come evidenziato, è non conosciuto (e che quindi è il principale parametro target dello studio). Ad esempio se ipotizzassimo una percentuale "vera" pari al 4% in un campione di 20.000 unità si avrebbero tendenzialmente 800 casi; la dimensione di circa 20.000 partecipanti al panel, coerente con quella indicata nella Proposta di Alleva e altri, è quindi il livello minimo di unità da considerare

Il periodo intercorrente tra due rilevazioni e la durata del panel (ovvero il numero di prelievi per unità campionaria) dipendono da:

- *risorse disponibili* (rilevatori, laboratori, struttura di supporto); in particolare si dovrebbe tenere conto della capacità di analisi dei test ovvero del numero dei laboratori disponibili, per evitare che essi si trasformino in un collo di bottiglia della procedura, rendendo la disponibilità delle informazioni relative ai singoli panel caotica e random;
- *"supportabilità" dei prelievi da parte del singolo individuo*;
- *tempi di diffusione del virus*, ad esempio utilizzando la distribuzione di frequenza dei giorni di trasmissione del contagio tra persona infetta e sana (il numero di giorni medio o mediano perché si verifichi il contagio o i giorni corrispondenti ad una data percentuale di

³ <https://www.ifw-kiel.de/publications/media-information/2020/a-new-approach-to-coronavirus-testing-necessary-to-contain-uncertainty/>

contagiati), massimizzando in questo modo la probabilità di avere un numero sufficiente di variazioni di stato della popolazione e minimizzando i costi dell'informazione "utili" al processo decisionale;

- *tempestività delle informazioni* necessarie alla valutazione dell'introduzione di misure restrittive o della loro cancellazione;

La fissazione del periodo intercorrente tra due rilevazioni consecutive non è quindi un problema tecnico ma piuttosto l'esito di un processo decisionale che vede coinvolti esperti di settore e management della crisi.

Un periodo più lungo dei 5 giorni proposti da Kiel assicurerebbe comunque gran parte degli obiettivi individuati, esaltando il ruolo di "sentinella" e di produzione di indicatori per l'analisi di diffusione del virus e per la predisposizione delle misure necessarie al suo contrasto.

I test implicano una interazione tra i partecipanti al panel e il personale sanitario, per la quale è necessario rinviare alle osservazioni della autorità sanitarie:

- test presso il domicilio o presso strutture specializzate, con le opportune cautele relative ai contatti e agli spostamenti
- comunicazione o meno del risultato del test ai positivi del tutto asintomatici. (dal punto di vista statistico è preferibile non comunicarlo per non alterare il comportamento del soggetto; dal punto di vista sanitario tutt'altro discorso
- possibile fornitura, in alcuni casi, di materiale per l'autodiagnostica
- altro

Nel caso di un test "domiciliare" i "rilevatori" dovrebbero essere comunque persone già formate ovvero personale infermieristico o paramedico. L'effettuazione del test comporta l'utilizzo di dotazioni individuali di sicurezza per evitare il rischio di contagio ai rilevatori e alle persone sottoposte a trattamento. Il numero dei rilevatori dipende dal tempo necessario per spostamenti e somministrazione del test nonché dall'organizzazione di supporto necessaria per fissare gli appuntamenti e la rotazione del campione. In linea generale possiamo ipotizzare al massimo 10 prelievi al giorno per rilevatore.

3. Altre proposte

Elaborazione di dati sulla mortalità differenziale. Si tratta di un'elaborazione già in via di effettuazione da parte di ISS e ISTAT e mira al confronto dei dati sulla mortalità tra i mesi del corrente anno e i corrispettivi del o degli anni precedenti, secondo il sesso e l'età. Questa misurazione potrà fornire, nell'ottica di un sistema informativo integrato, non solo un contributo ad una stima più precisa del numero di decessi da Covid-19, ma anche un indicatore di efficacia delle misure di restrizione adottate nel nostro Paese.

Indagini ad hoc su subpopolazioni marginali e/o particolari. Un particolare problema è rappresentato dal coinvolgimento nella rilevazione delle persone presenti sul territorio nazionale, ma non iscritte in anagrafe. Come sappiamo si tratta probabilmente di più di mezzo milione di persone (in larga prevalenza stranieri di giovane età), che si concentrano in alcune grandi città. Nell'attuale situazione di emergenza questo problema viene poco evidenziato, ma la fragilità delle loro condizioni di vita e l'impossibilità in molti casi di permanere in un'abitazione rendono questo potenziale fattore di contagio non secondario. In questo caso non si può escludere una estrazione

da altri registri (ad es. Caritas diocesane) e un coinvolgimento obbligatorio, eventualmente premiato con corsie preferenziali di iscrizione all'anagrafe.

Per quanto riguarda altre sub-popolazioni problematiche già citate, possiamo ipotizzare che il raggiungimento di individui appartenenti ad alcune categorie sia quanto meno problematico per l'orario in cui essere disponibile per il test; ad esempio il personale impegnato nel settore della distribuzione e del trasporto pubblico, le forze di polizia etc. Un problema a sé è rappresentato dal personale medico che viene sottoposto correntemente al test e che dovesse risultare nel campione. Una possibile soluzione è sostituire tali unità e considerare tale personale una distinta sub-popolazione le cui informazioni dovrebbero essere accessibili (almeno a livello macro) e aggiunte alle stime prodotte dall'indagine. Infine, campionamenti particolari dovrebbero essere effettuati sulla popolazione residente in convivenze (case di cura e riposo, monasteri, prigioni...) le cui liste sono comunque in gran parte disponibili presso le anagrafi.

Elaborazioni ed analisi mediante fonti esistenti: Istat e operatori digitali

Un forte miglioramento informativo, facilmente organizzabile nel breve periodo e a costi estremamente ridotti se non nulli, potrebbe derivare dall'elaborazione di dati esistenti presso l'Istat. A tale scopo sarebbe necessario che la Protezione Civile, in deroga alle norme sulla privacy, fornisse all'Istat gli identificativi (codice fiscale) di contagiati, ospedalizzati, guariti e deceduti; l'Istat potrebbe così procedere a agganciare le informazioni in proprio possesso a tali identificativi (sia quelle derivanti da archivi amministrativi – anagrafi, INPS etc. - sia quelle ottenute cumulando indagini campionarie). Questo consentirebbe di:

1. analizzare l'individuo ospedalizzato in relazione alla sua famiglia ovvero di altri individui per i quali sono disponibili altre informazioni socio-demografiche;
2. geolocalizzare tutti i casi fornendo aggregazioni a livello, ad esempio, di sezione di censimento e permettendo la stima di modelli spazio-temporali con elevata capacità predittiva;
3. individuare le situazioni maggiormente fonte di contagio, conoscendo luogo di residenza, di lavoro e di studio e alcune caratteristiche socio demografico.

Tali analisi sarebbero rese più esplicative e precise se per le singole unità fosse possibile disporre, sempre con l'assenso dell'Authority per la privacy, le informazioni sui loro spostamenti archiviate presso i gestori telefonici.

Su tali individui/famiglie, inoltre, potrebbero inoltre essere pianificate e eseguite indagini CATI (a zero rischio) per individuare le possibili cause di contagio (viaggi, partecipazione a eventi di massa, rispetto o meno delle misure restrittive, ecc.). Ovviamente quanto sopra non permetterebbe di stimare quantità come il tasso reale di contagiati, distinguendoli in sintomatici e asintomatici ma permetterebbe di ottenere fattori di espansione, non "complessivi" ma legati a specifiche situazioni di rischio, da utilizzare nel calcolo dei contagiati reali sulla base dei rilevati. Fattori già oggi disponibili peraltro calcolati utilizzando modelli basati sul numero dei deceduti.

Ricorrendo a fonti private è possibile elaborare anche una stratificazione comportamentale utilizzando le informazioni in possesso di operatori quali Google e Facebook⁴. L'acquisizione di

⁴ In base al GDPR, Regolamento UE 2016/679, art. 20, ogni singolo partecipante al campione ha il diritto di trasmettere tali dati a un altro titolare del trattamento senza impedimenti da parte del titolare del trattamento cui li

questi dati comportamentali può implicare tempi più lunghi rispetto a quelli necessari per l'avvio del campione socio-demografico. Tuttavia l'iter e l'acquisizione del consenso individuale dei partecipanti al campione vanno avviati da subito per l'importanza evidente delle correlazioni di alcune caratteristiche socio demografiche (ad es. sesso ed età, professione) con alcuni pattern comportamentali (intensità e modalità degli spostamenti; intensità dei rapporti familiari; partecipazione ad eventi sportivi, culturali, politici; l'interesse manifestato per le informazioni medico-scientifiche; ecc.). Inoltre l'utilizzo di strumenti di IA può far emergere correlazioni significative (positive o negative) con pattern imprevedibili e talvolta inspiegabili, il cui utilizzo si è dimostrato efficace in molti campi, quali i mercati finanziari e la comunicazione pubblicitaria. È verosimile che tali correlazioni saranno utilizzate da soggetti privati e pubblici in occasione delle prossime crisi epidemiche; tanto vale iniziare a sperimentarle pubblicamente in occasione di quella attuale.

** Gianluigi Bovini è statistico. Membro del segretariato di ASviS (Alleanza italiana per lo sviluppo sostenibile). In precedenza Capo Area Programmazione, Controlli e Statistica del Comune di Bologna.*

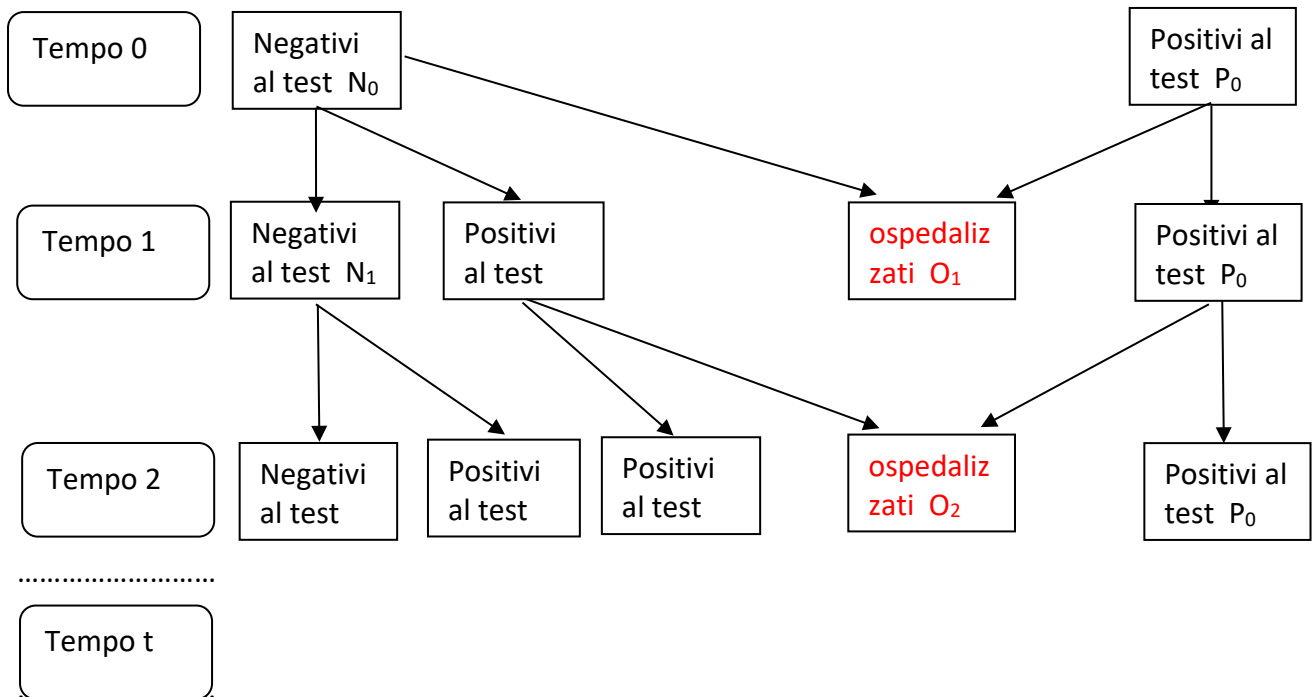
Manlio Calzaroni è statistico. Dal 1980 al 2017 ha lavorato presso l'Istituto Nazionale di Statistica, ricoprendo il ruolo di Direttore in diverse aree: censimenti, registri statistici delle unità economiche e della popolazione; uso di dati amministrativi per fini statistici. Oggi è consulente statistico in ambito nazionale ed internazionale.

Piero De Chiara è stato dirigente in OLIVETTI, Telecom Italia e LA7. E' stato Consigliere in AGCOM e vice-presidente del Comitato nazionale Italia digitale

Mauro Masselli, esperto statistico senior, collabora a progetti statistico-metodologici dell'Eurostat; negli anni precedenti Dirigente di Ricerca Metodologico presso ISTAT, Componente del Nucleo di Valutazione e Verifica degli Investimenti Pubblici del Ministero dell'Economia e Direttore dell'Unità Tecnica della Rete dei Nuclei di Valutazione Verifica degli Investimenti Pubblici delle Regioni e dei Ministeri.

ha forniti. L'attivazione di questo diritto alla portabilità può essere una delle condizioni previste per la partecipazione al panel, aprendo la strada per l'acquisizione di dati comportamentali di rilevante interesse pubblico.

APPENDICE: schema dello sviluppo del panel e calcolabilità dei principali indicatori



INDICATORI PRINCIPALI

Indicatori al tempo 0	N_0/N			P_0/N	
Indicatori al tempo 1	N_1/N	N_1/N_0	P_1/N_0	$(P_1+P_0)/N$	O_1/N
Indicatori al tempo 2	N_2/N	N_2/N_1	P_2/N_1	$(P_2+ P_1+P_0)/N$	$(O_2+O_1)/N$
.....					
Indicatori al tempo t	N_t/N	N_t/N_{t-1}	P_t/N_{t-1}	$(P_t+.....+P_2+ P_1+P_0)/N$	$(O_t+.....+ O_2+O_1)/N$

NB

- 1) I positivi sono in realtà distinti in asintomatici e pauci sintomatici; questa distinzione non è derivabile dal test ma è possibile sulla base di domande fatte dal “rilevatore” e/o mediante CATI. Tali aggregati e gli specifici indicatori non sono stati inseriti nel diagramma per semplicità.
- 2) Il numero di ospedalizzati è anche reperibile su base censuaria dalla Protezione Civile o dall’ISS mentre i singoli ospedalizzati sono registrati presso il Data base dell’ISS. L’accesso al DB o la fornitura di elaborazioni dai dati del DB dovrebbe essere garantiti. Quindi l’indicatore “ospedalizzati/N” è calcolabile così come altri più specifici indicatori di derivazione degli ospedalizzati dai negativi o dai positivi.