

La pandemia dei dati: intervista a Giorgio Alleva e Alberto Zuliani



Martina Patone

Follow

Nov 16 · 18 min read



Photo by [camiah](#) on [Unsplash](#)

L'appuntamento con i dati del giorno è ormai diventato una rubrica quotidiana di molti giornali e media nell'Italia del coronavirus: c'è chi li spiega, chi li mette in chiaro, chi li legge e chi li analizza. Eppure, i dati non sono solo numeri la cui esistenza è indipendente dalla realtà umana: siamo noi a decidere come raccogliarli, analizzarli e

interpretarli. In ognuna di queste fasi, se non si opera con metodo, si rischia di avere un'immagine distorta della realtà (o non averla affatto), e dunque di compromettere tutte le azioni che da questa conoscenza — erronea o assente — derivano.

L'economista Giorgio Alleva e lo statistico Alberto Zuliani, presidenti dell'Istat dal 2014 al 2018, il primo, e dal 1993 al 2001, il secondo, in una loro lettera aperta al Corriere parlano delle criticità dei dati epidemici a nostra disposizione e avanzano una proposta che potrebbe aiutare a far luce sulla situazione reale del Paese, partendo dalla raccolta di dati fatta con metodi statistici.

Abbiamo avuto il piacere e l'onore di intervistare i due professori, che gentilmente e nel dettaglio ci hanno raccontato cosa c'è che non va nei dati di oggi e come si potrebbe fare di meglio, con solo poche risorse in più.

Domanda: Il problema dei dati lo avevamo già incontrato durante la prima ondata. All'epoca i tamponi scarseggiavano e venivano fatti solo ai sintomatici gravi. Sapevamo che il numero di positivi registrato dai bollettini quotidiani della Protezione Civile era sottostimato, così come il numero di decessi. Così è stato poi confermato dai risultati delle indagini sierologiche e di mortalità dell'Istat. Oggi questo problema sembra non esserci più, perché è vero che facciamo più tamponi (26.000 circa fatti a marzo contro i 220.000 circa fatti oggi giornalmente), eppure, come faceste all'epoca, oggi scrivete che il problema dei dati è ancora attuale e che c'è bisogno di un metodo di raccolta 'statistico'. *Qual è dunque il problema? E perché non possiamo fidarci dei dati che abbiamo oggi?*

Risposta: Non è che non possiamo fidarci dei dati che vengono comunicati attraverso i media. Non sono sufficienti a dare risposta alle domande che ogni persona si fa. Inoltre non viene reso disponibile un 'data-set' di dati individuali resi anonimi perché i ricercatori di qualsiasi estrazione disciplinare — e anche statistici — possano dare il proprio contributo alla migliore conoscenza della pandemia e quindi aiutare a fronteggiarla. Sarebbe una cosa ottima se operassero in équipe con una genuina disponibilità alla interdisciplinarietà. Ecco le domande che i cittadini si pongono: qual è la probabilità di contagiarsi; se contagiati in modo lieve, di trovare risposta nella medicina territoriale; se contagiati seriamente, di trovare accoglienza nelle strutture ospedaliere, per terapie sub-intensive e intensive; se entrati in ospedale, di uscirne guariti; in quali luoghi e contesti c'è rischio maggiore di contagio; infine, sulla base di quali informazioni vengono prese le decisioni circa le misure restrittive da adottare. Non è citando insieme, giorno per giorno, il numero di tamponi effettuati, di casi positivi e di decessi che possiamo capire cosa stia accadendo. Neppure il bollettino settimanale dell'Istituto superiore di sanità, che pure presenta una quantità maggiore di dati, dà risposte esaurienti. Il problema, anzi i problemi, sono: la mancanza di alcuni dati

fondamentali; l'assenza di collegamento fra quelli disponibili; la scarsa iniziativa per l'approfondimento di problemi/opportunità emergenti; la limitata produzione di informazione adeguata a sostenere i governi — ai diversi livelli — per le decisioni da prendere, a convincere i cittadini ad assumere comportamenti adeguati e le imprese ad attivare misure di contrasto nei luoghi di lavoro; infine, la scarsa “sensibilità statistica” nella definizione, raccolta, elaborazione e comunicazione dei dati.

Domanda: *Esistono esempi per le problematiche elencate?*

Risposta: A otto mesi dall'inizio dei contagi ancora non sappiamo quale sia la prevalenza nella popolazione, cioè quanti sono positivi al tampone su 100 residenti puntualmente o nell'arco di pochi giorni, eventualmente da una certa età in poi (l'*Office for National Statistics* del Regno Unito considera la popolazione da 16 anni in poi); ovviamente non se ne segue l'evoluzione nel tempo. A noi sembrano dati essenziali. Mancano stime delle probabilità di passaggio fra stati delle persone che si imbattono nella patologia: da asintomatico a eventualmente sintomatico, al trattamento domiciliare, al ricovero in terapia sub-intensiva e intensiva, alla guarigione o malauguratamente alla morte, informazioni fondamentali per consentire la lettura delle conseguenze sulla salute delle persone e una previsione migliore dell'impatto sul sistema sanitario. È stata molto propagandata l'app *Immuni* per il tracciamento, ma le ASL, aziende sanitarie locali, non caricano regolarmente tutti i dati sulle persone contagiate. Soprattutto, non risulta che sia stato predisposto un piano di elaborazione della grande massa di dati che potrà essere raccolta in questo modo.

Domanda: *Che cosa è andato storto, quindi?*

Risposta: In generale, ci si è accontentati dei dati disponibili senza riflettere su quali fossero quelli realmente necessari per capire, attrezzandosi per raccoglierci. Non è maturato un atteggiamento di “conoscere per decidere e operare” (*knowledge for action*). Le misure del governo appaiono ispirate al buon senso, adottando un principio giustificato di precauzione (per la verità non soltanto in Italia); sono state spesso estese a tutto il territorio nazionale e in qualche caso continuano a esserlo, forse con l'aspettativa di aumentarne il rispetto; in effetti hanno contribuito a creare un clima di forte coesione nazionale nella prima fase acuta dell'epidemia a marzo e aprile. Da una settimana alcune misure sono state differenziate in relazione a valutazioni di rischio per le diverse regioni, ma poiché le valutazioni poggiano su una base informativa assai fragile si sta alimentando un forte contenzioso fra governo centrale e regioni.

Domanda: *In questo momento di che tipo di dati ci accontentiamo?*

Risposta: Sostanzialmente, ogni giorno, vengono diffusi tre dati: numero di tamponi, di nuovi contagiati e di decessi. L'aumento di contagiati e di morti a ritmi alquanto

sostenuti ha suscitato nelle ultime settimane forte allarme; ma i tre dati non sono sufficienti a far capire al paese come l'epidemia stia procedendo. Occorrerebbero trasparenza dei procedimenti di raccolta ed elaborazione, un quadro informativo convincente e una comunicazione chiara. C'è scarsa sensibilità statistica; basta riflettere su un indicatore, non sempre esplicitato, ma suggerito dall'accostamento fra numero di tamponi e numero di contagi giornalieri: il tasso di positività (numero di soggetti positivi al test diagnostico/numero di tamponi processati dai quali derivano) utilizzato per confronti nel tempo e nello spazio, senza curarsi dell'omogeneità dei dati, ad esempio riguardo a quale tipo di test sia stato eseguito e per quale motivo. Considerando quest'ultimo aspetto, si sa che i tamponi vengono effettuati attualmente in ragione di circa il 30% per ciascuno dei seguenti tre motivi: presenza di sintomi; tracciamento di contatti; attività di *screening*, con tassi di positività ovviamente differenti per i tre gruppi. Ma il mix non è rimasto invariato nel corso del tempo, non risulta omogeneo fra le regioni e nell'ultimo periodo è aumentata la quota di tamponi per i quali il motivo non è noto (14,1% fra il 5 e il 18 ottobre). Quindi, i confronti non possono essere proposti sensatamente. C'è una situazione emblematica che testimonia la scarsa sensibilità statistica, questa volta da parte del governo: il presidente dell'Istat non è stato indicato fra i 70 componenti della "Task force dati" che il decreto "Cura Italia" del marzo scorso ha previsto di istituire presso il Ministro dell'Innovazione.

Domanda: *In cosa consiste, in breve, la vostra proposta? Quali sono gli indici che vi proponete di stimare attraverso il vostro campione e quali nuove informazioni sull'epidemia potrebbero darci?*

Risposta: A fine marzo, in piena prima ondata dell'epidemia, quando si conosceva ancora molto poco di Covid-19, abbiamo avanzato la proposta di estrarre un campione della popolazione italiana, eventualmente a partire da una certa età, sul quale effettuare il test diagnostico per valutare la prevalenza e al quale proporre un questionario per raccogliere elementi utili a interpretare le dinamiche di contagiosità. La proposta era articolata; oltre a stimare la prevalenza puntuale mediante l'integrazione dei dati raccolti su un campione di contagiati e di loro contatti e su un campione di popolazione, era previsto di seguire longitudinalmente parte delle persone per poter valutare i passaggi di stato [da una condizione clinica all'altra, ndr]. Le stime ottenibili soddisfacevano le caratteristiche di non distorsione e presentavano un'efficienza di gran lunga superiore a quella di un campionamento casuale semplice ([l'articolo](#) che presenta la proposta metodologica è stato accettato per la pubblicazione da un'importante rivista internazionale di statistica) [*Journal of Official Statistics*, ndr]. Era previsto che la rilevazione fosse ripetuta ogni due settimane in modo da seguire l'evolvere dell'epidemia nel corso del tempo, modulando opportunamente i due campioni e il panel [Campione rappresentativo che resta immutato nel corso di indagini successive, ndr]. Allora non se ne fece nulla. Una giustificazione potrebbe essere rinvenuta nella

situazione di emergenza sanitaria che si stava vivendo, la quale suggeriva di destinare i tamponi (30–35.000 ogni giorno) ai soli sintomatici. Attualmente, si effettuano e si processano oltre 200.000 tamponi ogni giorno e destinarne il 5% una volta ogni due settimane alla stima della prevalenza sembra del tutto fattibile. Ovviamente, si tratta di un tassello, secondo noi molto utile, di un sistema articolato e integrato di monitoraggio.

Domanda: *Se questo è solo un tassello, in cosa consistono gli altri?*

Risposta: Altre componenti essenziali dovrebbero essere almeno le seguenti. Un campione di scuole e alunni per seguire l'evoluzione dei contagi nel sistema educativo; se si fosse in grado di farlo, il rilievo potrebbe essere totale. L'aggiornamento settimanale dalla rete dei medici di base (eventualmente da parte di un loro campione) sul numero di pazienti che hanno segnalato sintomi di sospetto Covid-19, sul risultato dei test eventualmente effettuati e sugli esiti successivi. La medicina territoriale è stata in larga misura assente nei mesi di marzo e aprile. Ancora oggi dà risposte parziali, disomogenee sul territorio, che non fanno argine al ricorso alle strutture ospedaliere le quali rischiano così una saturazione rapida senza reale necessità. L'aggiornamento giornaliero di informazioni principali dai reparti ospedalieri Covid e, in particolare, dalle terapie intensive e sub-intensive è un tassello essenziale del sistema informativo integrato; però, non può limitarsi al numero di nuovi ingressi fornito giornalmente; occorrono il numero di presenti alle ore 0.00, quello di ingressi e uscite nella giornata per esito e la storia dei ricoverati dal momento del loro ingresso all'uscita per dimissione o morte, in modo da poter valutare appropriatamente la pressione sui reparti specializzati; da dati come questi è possibile ottenere informazioni su tasso di saturazione giornaliero delle strutture, durata media della degenza, tasso di guarigione, tasso di letalità dei ricoverati. Molto utile è anche l'aggiornamento su contagi ed esiti per medici e paramedici nelle complessive strutture ospedaliere, ora episodico e disomogeneo. I dati dovrebbero essere integrati con quelli sanitari, orientati a individuare i focolai. I risultati riferiti alle persone dovrebbero essere resi disponibili per classi di età. Tutti dovrebbero essere dettagliati opportunamente a livello territoriale, oltre che per regione anche per le grandi città.

Domanda: *Oltre al comparto sanitario, chi altri potrebbe fare qualcosa per ottenere dati più specifici?*

Risposta: L'Istat potrebbe aggiungere stabilmente qualche domanda specifica ai questionari di rilevazione periodica sulle imprese (lo ha fatto una volta) e sulle forze di lavoro su presenza e intensità dello *smart working*, consentendo valutazioni sulla sua diffusione ed efficacia, nonché sull'adozione di misure igieniche e sul rispetto delle misure di contenimento via via adottate dal governo, dalle regioni e dai comuni - da parte della popolazione. Un impegno specifico dovrebbe essere dedicato

all'individuazione attraverso il sistema di tracciamento dei luoghi e dei contesti nei quali la trasmissione del SARS-CoV-2 è più frequente, predisponendo e aggiornando mappe di rischio e mettendo i decisori nelle condizioni di proporre interventi selettivi su aree territoriali, sociali ed economiche, ormai richiesti e attesi dalla collettività. Come ogni strumento efficace di monitoraggio, il sistema delineato dovrebbe essere in grado di produrre rapidamente stime affidabili su aspetti via via emergenti.

Domanda: Sono in molti a dubitare della validità di dati ottenuti da campioni: basti pensare alla puntata di Report dell'8 giugno 2020 in cui si metteva in dubbio la validità statistica di un campione di 150.000 cittadini a fronte di 60 milioni di italiani. Eppure, gli statistici sanno che la teoria del campionamento garantisce stime rappresentative della popolazione. *Come si fa a prendere un campione rappresentativo? Di quante persone abbiamo bisogno? Potete fare un confronto con altre indagini campionarie che l'Istat conduce regolarmente?*

Risposta: Ormai da oltre un secolo è stato dimostrato scientificamente che è possibile estendere i risultati da un campione probabilistico alla popolazione dalla quale è stato estratto. Le unità che ne fanno parte devono essere selezionate in modo casuale fra quelle appartenenti alla popolazione di interesse, eventualmente sfruttando le informazioni disponibili sulle loro caratteristiche. La selezione casuale consente di affiancare una misura di accuratezza a una stima indotta dai dati campionari relativamente a una caratteristica della popolazione. specularmente, si può stabilire 'a priori' la dimensione del campione in funzione del margine di errore che si è disposti a tollerare. Si può quindi maneggiare sapientemente il 'trade-off' [scambio, ndr] tra livello di precisione e fattibilità della rilevazione, essenzialmente in termini di costi, tempi e risorse organizzative da mettere in campo. La dimensione della popolazione influisce molto poco sulla numerosità del campione; assai più influente è la rarità o meno della caratteristica della popolazione che si intende stimare. D'altra parte, la qualità del campione dipende dalla qualità della lista dalla quale è possibile estrarlo, in termini di copertura rispetto alla popolazione di interesse, livello di aggiornamento e informazioni disponibili per le unità che vi sono comprese.

Domanda: *Come si può quindi selezionare un campione per monitorare la situazione epidemiologica?*

Risposta: Per la selezione del campione necessario a condurre una indagine per determinare la probabilità di contagio della popolazione residente a una certa data si deve disporre della lista anagrafica nominativa dei residenti dalla quale estrarre casualmente il numero di unità programmato (e anche una lista di riserva per eventuali sostituzioni), tenendo conto delle informazioni anagrafiche disponibili, ad esempio genere, età, cittadinanza, area geografica, tipo di comune di residenza. Secondo la

nostra proposta un campione periodico di 10.000 persone (cioè il 5% di 200.000, come abbiamo detto nelle precedenti risposte) permetterebbe di monitorare il tasso di contagio della popolazione a livello nazionale con un margine di errore accettabile. Approfondimenti regionali comporterebbero uno sforzo ulteriore. Attualmente, con oltre 200.000 accertamenti condotti quotidianamente dal sistema sanitario, si tratterebbe di una quota aggiuntiva marginale, compensata dal contributo informativo che potrebbe apportare. Per confronto, l'indagine campionaria periodica sull'occupazione condotta dall'Istat si basa su un campione di circa 22.000 persone al mese. Naturalmente non sarà sempre facile avere la disponibilità delle persone a partecipare a interviste personali o telefoniche e non si può sottovalutare la difficoltà di adesione volontaria a un accertamento diagnostico. Occorre una forte campagna promozionale sul valore collettivo dell'informazione da raccogliere. D'altra parte, nella fase che stiamo vivendo, conoscere le proprie condizioni di salute è importante e potrebbe essere un'opportunità da sfruttare.

Domanda: Sappiamo dunque che mancano dati 'statisticamente rappresentativi' della popolazione per poter stimare la reale prevalenza del virus (cioè la porzione della popolazione affetta dalla Covid-19 in uno specifico momento del tempo), e che la vostra proposta servirà a soddisfare questa necessità. Ultimamente sono molti gli scienziati che chiedono che vengano resi noti i dati del tracciamento, dati che potrebbero essere usati per capire quali sono i posti dove il rischio di contagio è più alto e quali sono le fasce di popolazione più a rischio, tenendo in considerazione eventuali differenze temporali. *Con i dati ottenuti dal campione che proponete è possibile ottenere queste stime?*

Risposta: La stima della probabilità di contrarre il virus e la sua evoluzione nel tempo, eventualmente con dettaglio territoriale, è solamente un tassello del quadro informativo necessario per la gestione dell'emergenza. Certamente, l'individuazione dei luoghi e dei contesti dove si manifesta più frequentemente il rischio di contagio è essenziale per mettere in atto misure mirate e fondate da parte delle istituzioni. Indagini campionarie su scala locale possono dare risposte utili. Ma riteniamo che la direzione da prendere sia quella dell'integrazione di tutti i dati raccolti attraverso il sistema sanitario con altri derivanti da indagini campionarie, all'interno di una strategia informativa rigorosa, efficace, cioè rispondente ai bisogni, messa a punto con il contributo anche delle competenze e delle istituzioni statistiche. In questo quadro, è di grande interesse l'elaborazione dei dati sui tracciamenti ottenuti attraverso le dichiarazioni dei contagiati o raccolti automaticamente dall'app *Immuni* ora utilizzati solamente per individuare persone alle quali comunicare disposizioni delle autorità sanitarie. Interviste ai contagiati sulle loro abitudini (ad esempio negli spostamenti) possono fornire informazioni importanti. La rappresentatività statistica potrebbe essere assicurata mediante un campionamento dei soggetti contagiati e dei loro contatti, uno degli approcci che abbiamo proposto. Sarebbe anche interessante raccogliere dati sul rispetto

delle norme igieniche anti-Covid che al momento non abbiamo o non conosciamo.

Domanda: *Attraverso un campione si potrebbero raccogliere queste informazioni, con questionari, così da stimare l'impatto che i comportamenti individuali hanno sul rischio di infezione?*

Risposta: Si potrebbero sfruttare le indagini correnti previste dai Regolamenti europei e condotte dagli Istituti nazionali di statistica sulle persone e sulle imprese. È questa la proposta del *Centre for European Policy Studies*, presentata fin dal mese di marzo da Daniel Gros [economista tedesco, attualmente direttore del *Centre for European Policy Studies*, ndr] per utilizzare l'ampio campione panel [Campione rappresentativo che resta immutato nel corso di indagini successive, ndr] europeo denominato '*EU statistics on income and living conditions*' (EU-SILC). In direzione analoga, l'Organizzazione internazionale del lavoro ha sollecitato gli Istituti nazionali di statistica a introdurre nelle rilevazioni sul lavoro un modulo destinato a valutare l'impatto del Covid-19, rimuovendo eventualmente sezioni con minore priorità. Una rilevazione via web estesa, con partecipazione volontaria, è stata sperimentata in Israele su circa 60.000 persone le quali hanno fornito informazioni relative a età, genere, stato di salute generale, sintomi legati a Covid-19 ed eventuale stato di isolamento. Lo scopo principale è stato quello di confrontare i risultati dei test e le situazioni di contagio per persone con differenti caratteristiche socio-economiche. L'impianto è stato realizzato dall'Istituto Weizmann. Per studiare in profondità fenomeni complessi come quelli connessi con la pandemia in corso occorrono disegni congiunti che sfruttino e valorizzino il potenziale informativo delle varie fonti. La nostra proposta è quella di realizzare una integrazione tra sorveglianza sanitaria e monitoraggio statistico, non di sostituire una fonte con l'altra.

Domanda: *Quale ruolo hanno i big data?*

Risposta: Le nuove fonti (*big data*) hanno potenzialità estremamente importanti e possono fornire un contributo prezioso. Pensiamo a sensori per rilevare il tasso di occupazione dei vettori del servizio pubblico di trasporto; ai dati della telefonia mobile per stimare la dimensioni dei flussi di spostamenti tra comuni nel corso della giornata e durante i giorni della settimana ma anche le presenze in specifici luoghi (gli "assembramenti"), agli *scanner data* per studiare come si siano modificate le abitudini di consumo nella grande distribuzione. Il *Daily Social Mood Index* che l'Istat ha sviluppato nel 2018 sulla base di un campione esteso di Twitter rappresenta un buon esempio di sperimentazione utile. È stato appena pubblicato su *Nature* un articolo che presenta i risultati di *dynamyc network models* per simulare la diffusione di SARS-CoV-2 in dieci nelle maggiori aree metropolitane degli Stati Uniti a un livello territoriale molto fine, proprio basato sui dati della telefonia mobile. Occorre costruire partnership strategiche tra le società che possiedono quei dati e le istituzioni pubbliche che hanno la

responsabilità della gestione sanitaria dell'epidemia in modo da trasformarli in bene pubblico. La *Covid Era* rappresenta un'occasione da non perdere.

Domanda: Uno degli aspetti che ha caratterizzato il 2020 è il crollo delle certezze che avevamo nella scienza. La maggior parte delle persone si è scontrata con la realtà che nella scienza si lavora con l'incertezza. Da una parte c'è l'incertezza derivante dalla natura dei dati, che ne complica l'interpretazione, come chi lavora con la statistica sa bene. E poi c'è l'incertezza intrinseca nei fenomeni del mondo, alla quale chi lavora con la statistica è abituato e sa che certe volte, situazioni 'fuori dalla media' possono verificarsi proprio perché esiste quell'incertezza e non può essere cancellata. È perfettamente possibile lanciare una moneta dieci volte e vedere uscire dieci volte testa. Forse è anomalo, ma non impossibile. Chi lavora con la statistica si abitua presto a questo tipo di ragionamento, ma non tutti si sentono a loro agio di fronte a un simile evento, e anzi lo reputano sbagliato. *Cosa ne pensate? Nella vostra esperienza come presidenti dell'Istat, avevate già notato questo "analfabetismo statistico"? Cosa si dovrebbe fare secondo voi per rimediare? Secondo voi, in questa pandemia avremmo avuto bisogno di più statistici?*

Risposta: In Italia il livello di acquisizione della 'numeracy' (la capacità di destreggiarsi in matematica) misurato dalle principali indagini internazionali sugli studenti pre-universitari è basso; nelle scuole si dà in generale poco spazio alle discipline scientifiche, soverchiate da quelle classiche, soprattutto in alcuni tipi. Riusciamo male anche perché i test internazionali includono non poche domande di natura statistica per rispondere alle quali i nostri studenti risultano poco attrezzati. L'analfabetismo statistico si esprime in due direzioni. Da una parte, manca la capacità critica nell'interpretazione dei dati; dall'altra, generalmente, non si valuta l'incertezza che è insita in ogni manifestazione della natura e anche della società. Quanto al primo aspetto, un modulo di formazione statistica nella scuola e nell'università potrebbe dare un contributo importante, abituando al metodo scientifico, difendendo dalle cialtronerie, dalle *fake news* veicolate da numeri (non dati e meno che mai informazioni statistiche) e dagli inganni. Se ne gioverebbero tutte le discipline poiché la statistica è un importante collante per l'interdisciplinarietà, frontiera della formazione avanzata e della ricerca. D'altra parte, l'incertezza è ancora scarsamente presente nella pratica educativa pre-universitaria e rimane in molti casi marginale in quella universitaria. Per questo anche alle persone istruite, soprattutto se adulte e anziane, risulta difficile convivere, capire che gli eventi futuri sono aleatori e che si può soltanto assegnare un certo grado di probabilità al loro verificarsi. Tornando al nostro tema, una maggiore consapevolezza statistica consentirebbe di capire che gli scienziati hanno appena intrapreso una marcia di avvicinamento al coronavirus e che sono ancora molte le congetture da verificare scientificamente. E quando una di esse risultasse più plausibile di altre, approfondimenti successivi potrebbero revocarla in dubbio, poiché la ricerca scientifica procede per gradi,

anche attraverso errori dei quali deve fare tesoro. Tuttavia, all'incertezza insita nei fenomeni, nel nostro caso anche per la variabilità biologica delle persone che comporta risposte differenti all'aggressione del virus e alle terapie messe in atto, si sommano quella derivante dalla difficoltà di individuare corrispettivi empirici soddisfacenti delle variabili di interesse e quella conseguente al mancato investimento in informazione aggiuntiva rilevante.

Domanda: Abbiamo capito che ogni fenomeno può verificarsi con un certo grado di probabilità, ma quando cerchiamo di “tradurre” questo concetto in dati da analizzare la situazione si complica. *Per parlare di Covid-19, le variabili in gioco sono moltissime. Qual è la difficoltà maggiore?*

Risposta: Un esempio di informazione insoddisfacente è, paradossalmente, il set di indicatori definito dal Ministero della salute per la valutazione del livello di rischio da Covid-19, che è alla base della classificazione delle regioni per intensità delle misure di contenimento da apprestare. Dei ventuno indicatori, sei riguardano la capacità di monitoraggio, sei la capacità di accertamento diagnostico e nove l'intensità di trasmissione del virus e la tenuta dei servizi sanitari. Dei primi sei, due sono opzionali in quanto “relativi a flussi di sorveglianza non attualmente attivi”; purtroppo, si tratta dei due indicatori riferiti alle residenze sanitarie assistite, focolai noti dell'epidemia. Gli altri quattro sono banali e poco rilevanti: la capacità di monitoraggio viene accertata sulla base della frazione di casi sintomatici notificati per i quali sia stata indicata la data di inizio dei sintomi; sulla frazione di casi ospedalizzati in reparti diversi dalla terapia intensiva per i quali sia stata indicata la data di ricovero; sulla frazione di casi ospedalizzati in reparti di terapia intensiva per i quali sia stata indicata la data di ricovero o trasferimento; infine, sulla frazione di casi notificati per i quali sia indicato il comune di domicilio o residenza. Non si percepisce alcuno sforzo di ricercare corrispettivi empirici della capacità di monitoraggio maggiormente pertinenti e probanti. Riguardo alla capacità di accertamento diagnostico, vengono indicati: il tempo medio intercorso fra la data di inizio dei sintomi e quella di isolamento (opzionale); il tempo medio intercorso fra la data di inizio dei sintomi e quella di diagnosi; la percentuale di tamponi positivi escludendo per quanto possibile (*sic!*) le attività di *screening* e il *re-testing*; numero, tipologia professionale e tempo/persona dedicati rispettivamente al *contact-tracing* e al prelievo/invio ai laboratori dei contatti stretti e dei casi in quarantena e isolamento; infine, frazione di casi confermati di infezione nella regione per i quali sia stata effettuata una regolare indagine epidemiologica con ricerca dei contatti stretti rispetto al totale di nuovi casi. Le definizioni sono alquanto approssimative (lontane dal rigore statistico) e alcune informazioni risultano di difficile accertamento (ad esempio il tempo/persona) cosicché la discrezionalità è prevedibilmente elevata. Riguardo, infine, all'intensità di trasmissione del virus e alla tenuta dei servizi sanitari, due dei nove indicatori sono opzionali, uno dei quali è il

numero di accessi al pronto soccorso con quadri sindromici riconducibili a Covid-19 (che sarebbe di interesse). Altri quattro sono riferiti a contagiati: negli ultimi quattordici giorni; per data di inizio dei sintomi o di diagnosi con cadenza giornaliera; non associati a catene di trasmissione note; presenza di focolai (con una definizione generica). Gli ultimi tre sono specifici e rilevanti: tasso di occupazione di posti letto in area medica per pazienti Covid-19; tasso di occupazione di posti letto in terapia intensiva per pazienti Covid-19; R_t [indice di riproduzione, ndr]. Gli indicatori prescindono dalla dimensione demografica dei territori regionali che, invece, in molti diversi, sarebbe necessario considerare.

Domanda: Chiudiamo con una riflessione sulla situazione italiana...

Risposta: Ci siamo fatti e continuiamo a farci due domande. C'è bisogno di una maggiore presenza della competenza statistica? Rispondiamo sì, senza spirito di corpo. Se ci sia bisogno di una discesa in campo della statistica ufficiale? Rispondiamo ancora sì. C'è lì un serbatoio di professionalità da utilizzare: in primo luogo, per eliminare le ambiguità del quadro definitorio attualmente presenti, addirittura riguardo ai contagiati accertati, poiché non viene specificato se si debba considerare soltanto i test molecolari o anche gli antigenici e l'arbitrarietà dell'interpretazione può provocare una forte disomogeneità dei dati che provengono dai differenti presidi, aziende sanitarie locali e, per aggregazione, regioni, non consentendo confronti corretti; in secondo luogo, per condurre indagini specifiche, ad esempio per stimare la prevalenza nella popolazione come suggerito da noi o per valutare il rischio di trasmissione del SARS-CoV-2 in specifici contesti territoriali e sociali mediante i dati derivanti dal sistema di Istat ^a Data ^b Campionamento ^c Articolo ^d New ^e zione dei dati raccolti correntemente. L'incertezza rimarrà, è ineludibile, ma l'imprecisione può e deve essere ridotta.

La redazione di questo articolo è a cura di Monica Murano.

[About](#) [Help](#) [Legal](#)

Get the Medium app

