

I dati sul CoViD19: maneggiare con prudenza

scritto da Federico de Luca | 22 Aprile 2020



Mai come in questo periodo si è vista una proliferazione di dati pubblicati quotidianamente su tutti i canali di comunicazione (formali o meno) a proposito della crisi per CoViD19. Se non altro, questo rappresenta un ottimo esempio di quanto siano importanti i dati e di come sia auspicabile che le scelte politiche facciano affidamento su di essi per ottenere i risultati migliori.

Ciononostante, però, nei dibattiti attuali si parla molto poco dei limiti che accompagnano questi dati, e questo potrebbe trasmettere un falso senso di sicurezza al momento della loro interpretazione e del loro uso. Questa nota vuole aiutare a colmare questa lacuna, condividendo alcune riflessioni sul significato e l'interpretazione dei dati sul CoViD19, e aiutando i lettori a capirne i limiti e a comprendere la conseguente incertezza che circonda le conclusioni che se ne possono trarre. Quest'incertezza rischia di essere particolarmente fuorviante ogni volta che cerchiamo di fare confronti tra Regioni e Stati diversi.

È bene chiarire che i limiti di questi dati non si sono manifestati per la prima volta con la pandemia di CoViD19. L'epidemiologia ha sviluppato metodi di analisi complessi per tenerne conto, e le loro conclusioni contengono un certo margine di incertezza con cui le autorità devono confrontarsi. Questa nota, invece, vorrebbe stimolare nei suoi lettori un pensiero critico verso questi dati e, in particolare, verso tutte quelle analisi che, unendo in maniera semplicistica qualche dato, mirano a fornire delle facili conclusioni che, però, non hanno alcuna validità.

I casi positivi

Iniziamo dall'andamento del numero di casi positivi, che forse è l'argomento più pubblicamente trattato dall'inizio della crisi. Purtroppo, nonostante a prima vista possa sembrare una misura molto semplice da gestire, questo dato presenta alcune gravi limitazioni che dovrebbero essere prese in considerazione quando lo si usa, perché:

- non equivale al numero totale dei contagiati, che è per definizione maggiore (a meno che i test non vengano svolti sulla totalità della popolazione e a intervalli regolari nel tempo);
- dipende dal numero e dal tipo di test effettuati, che spesso non è uniforme nelle diverse fasi temporali e luoghi dell'epidemia;
- dipende dalla tempistica con cui i risultati dei test effettuati diventano disponibili e da quella con cui, una volta disponibili, questi vengono comunicati alle autorità;
- dipende dal corretto funzionamento dei test, che in alcuni casi non si sono rivelati affidabili;
- dipende dalla completezza con cui vengono riportati i dati (può succedere, infatti, che in alcuni giorni qualche fonte riscontri un problema e non possa riportare i propri dati in maniera completa);
- dipende dalle caratteristiche delle persone sottoposte al test, perché questo influenza la probabilità che queste risultino positive (per esempio, somministrando i test solo a persone con sintomi si troverà una proporzione di casi certamente più elevata di quella rilevabile se i test venissero somministrati a tutta la popolazione);
- dipende dall'eventuale ripetizione dei test sugli stessi individui (ma in diversi momenti nel tempo);
- dipende dallo stadio di avanzamento dell'epidemia in ogni realtà specifica (1000 nuovi casi nella parte finale della pandemia avrebbero un significato certamente diverso da 1000 nuovi casi nel pieno della stessa).

I decessi

L'altro andamento citato nella maggior parte dei dibattiti è quello del numero di decessi, probabilmente quello col maggior carico emotivo in tutta la narrativa legata alla pandemia di CoViD19. Sfortunatamente, anche questo dato ha delle grosse debolezze, perché:

- dipende dalla capacità (o dalla possibilità) delle autorità di identificare correttamente la presenza del CoViD19 nelle persone decedute (ad esempio, una scarsità di risorse potrebbe costringere a concentrare test e le energie sui pazienti in vita invece che sui morti, oppure le persone decedute in casa potrebbero non essere sottoposte al test con la stessa regolarità dei decessi registrati in ospedale);
- dipende dalla completezza con cui vengono riportati i dati (può succedere, infatti, che in alcuni giorni qualche fonte riscontri un problema e non possa riportare i propri dati in maniera completa);
- dipende dalle caratteristiche della popolazione colpita dal CoViD19 (per esempio, una popolazione mediamente anziana – e quindi con un maggior numero di individui di salute già compromessa – registrerà

inevitabilmente un maggior numero di decessi);

- dipende dalle caratteristiche delle persone sottoposte al test (con o senza sintomi, a maggiore o minore rischio, per i più diversi fini diagnostici), perché questo aumenta le probabilità di una diagnosi precoce e di un'assistenza medica adeguata, e quindi una conseguente diminuzione dei decessi;
- dipende dalla preparazione e dalla capacità del sistema sanitario nazionale (o regionale) a far fronte alla pandemia (e.g. attrezzature, personale);
- dipende dallo stadio di avanzamento dell'epidemia in ogni realtà specifica (1000 nuovi decessi nella parte finale della pandemia avrebbero un significato certamente diverso da 1000 nuovi decessi nel pieno della stessa).

I test eseguiti

Infine, un dato di cui si è parlato poco all'inizio della pandemia, ma che è stato progressivamente incluso nella maggior parte dei dibattiti, è il numero di test eseguiti. Anche questo dato, per quanto apparentemente semplice, presenta debolezze, perché:

- non è sempre possibile capire di quali test si parla, perché spesso i risultati dei test eseguiti non sono disponibili il giorno stesso (e.g. i numeri pubblicati sono quelli il cui risultato è arrivato il giorno X? O sono quelli fatti il giorno X? O sono quelli il cui risultato è stato comunicato il giorno X?);
- non è chiaro se include tutti i test di controllo effettuati su pazienti che si ritengono guariti (e che vengono ripetuti su una stessa persona);
- per poter essere interpretato, dipende dai criteri con cui i test vengono eseguiti sulla popolazione, che hanno delle enormi conseguenze sull'interpretazione di tutti i dati legati all'epidemia, e spesso variano tra Stati e Regioni, ma anche – all'interno di una stessa realtà – nel tempo.

Altre misure e indici

Vi sono poi vari altri numeri che vengono citati più o meno spesso, e anche loro presentano le loro debolezze: il numero di ricoverati in terapia intensiva, ad esempio, è accurato solo fino a quando ci sono dei posti a disposizione nelle terapie intensive. Dal momento in cui le strutture sanitarie si saturano, invece, esclude inevitabilmente tutti quei pazienti che avrebbero bisogno di un ricovero ma per i quali non vi è più posto. E lo stesso è vero per quanto riguarda il numero di ricoverati in ospedale. Il numero dei pazienti guariti, invece, dipende direttamente dalla definizione di "guariti" in uso nelle diverse realtà (in Italia, ad esempio, questa include solo i pazienti precedentemente infetti – positivi al test – che poi risultano negativi a due test consecutivi) e, ovviamente, da tutti i limiti evidenziati per l'identificazione dei casi positivi.

A causa delle limitazioni sopra elencate, è facile capire come anche tutti gli indici derivati da questi numeri vadano sempre adeguatamente contestualizzati in base alla zona geografica e al contesto epidemico. Questo

è vero per il saldo giornaliero degli attualmente positivi (spesso incorrettamente usato al posto del numero di nuovi casi positivi), il numero di casi positivi per abitanti, il numero di decessi per abitanti, il tasso di diffusione del CoViD19, etc.

Conclusioni

Quando si parla di diminuzione dei casi da un giorno all'altro, quindi, un primo raccomandabile riflesso sarebbe quello di andare a vedere il numero di test svolti nelle due giornate in esame. Ad esempio, il 30 marzo il numero di nuovi casi positivi in Italia è diminuito del 57%, ma sia il 29 che il 30 marzo in Italia il numero di test fatti è diminuito del 31% rispetto al 26, 27 e 28 marzo.

Quando si vogliono confrontare Stati e Regioni, invece, un'informazione fondamentale è rappresentata dalla giusta comprensione dei criteri seguiti nella somministrazione dei test nelle diverse realtà. Ad esempio, in Germania l'età media delle persone sottoposte ai test è intorno ai 46 anni, mentre in Italia questa è intorno ai 63 anni. Per questa ragione, proporre dei confronti tra realtà diverse rischia di essere tremendamente fuorviante.

È quindi fondamentale che si continui a ribadire, in tutte le occasioni e sedi opportune, che analizzare questi dati per formulare previsioni o confronti tra Stati e Regioni, senza porsi il problema di come trattare le loro intrinseche debolezze, è sbagliato e pericoloso. Piuttosto che cercare di farlo da soli, dovremmo ascoltare ciò che dicono gli epidemiologi e gli statistici, e prestare attenzione a ciò che i loro modelli indicheranno giorno dopo giorno.

Invece, questi dati sono particolarmente preziosi in determinati contesti locali, laddove i fattori che limitano l'accuratezza delle figure elencate in precedenza variano poco o nulla. Anche in questo caso, però, bisognerebbe comunque ascoltare solo le proposte di epidemiologi e statistici, che conoscono bene i limiti di questi dati e hanno sviluppato modelli che possono tenerne conto.

Spero che questa nota abbia instillato nei lettori qualche dubbio in più, e un po' di sana diffidenza nei confronti di proclami altisonanti che vengono troppo spesso spacciati come verità indiscutibili e basate sui fatti. Sebbene possano effettivamente essere basati sui fatti, ciò non significa che la loro lettura dei fatti sia quella corretta.

Nota alla nota:

Questo breve pezzo non è l'unico ad essersi occupato di questo tema. Alcune buone letture le potete trovare alle pagine qui sotto:

Sky tg24 – Coronavirus, come leggere i dati e interpretare il bollettino della Protezione Civile

Sky tg24 – Il coronavirus e i dati dell'epidemia in Italia

Avvenire.it – Coronavirus. I numeri reali ci dicono che il rischio era avere oltre 1 milione di morti

epicentro.iss.it – Sorveglianza integrata COVID-19: nota metodologica

The Guardian – Coronavirus statistics: what can we trust and what should we ignore?

Neodemos.info – Tutti gli articoli sul COVID19