Covid: una caporetto statistica

Riccardo De Bonis
,
Giovanni D'Alessio
6 Maggio 2020

Premessa: tutti abbiamo sottovalutato il coronavirus

Nel fortunato *Homo Deus. Breve storia del futuro*, prosieguo di *Homo Sapiens. Da animali a dei*, Yuval Noah Harari aveva peccato di ottimismo, scrivendo: *le epidemie costituiscono oggi una minaccia assai meno seria per la salute umana di quanto lo siano state nei millenni precedenti*.

In difesa di Harari va detto che anche nella peggiore delle ipotesi l'epidemia del coronavirus non provocherà il numero di morti prodotti dai flagelli del passato, lontani e vicini nel tempo. La peste nera del 1300 causò, secondo Harari, un numero di morti stimato tra i 75 e i 200 milioni (altri studiosi forniscono stime più basse). Il vaiolo e le altre malattie portate dagli europei nel Nuovo Mondo causarono solo in Messico la scomparsa di 20 milioni di persone. La prima ondata dell'influenza spagnola sterminò tra i 50 e i 100 milioni di persone nel 1918.

Dalla sua apparizione, nei primi anni Ottanta del Novecento, l'AIDS ha causato la morte di più di 30 milioni di uomini e donne. Nel 2014 il virus Ebola ha determinato 11.000 vittime. Il coronavirus ha già superato per numero di morti le conseguenze di Ebola, ma resteremo lontani dalle conseguenze delle epidemie del passato. Se, invece, in tutto il mondo non avessimo preso le misure di contenimento, i morti avrebbero potuto raggiungere il numero di qualche milione.

Anche molti infettivologhi e virologhi intervistati in questi giorni da TV, radio e giornali, hanno ammesso che fino alla metà di febbraio avevano pensato che l'Italia se la sarebbe cavata. Tutti noi, in primo luogo gli estensori di questa nota, e poi scienziati, cittadini comuni ed esponenti politici, abbiamo sottovalutato il virus in gennaio e in febbraio (carnevali celebrati, eventi sportivi a porte aperte, etc. etc.). Errori più grandi di quelli italiani sono stati commessi in altri paesi (la celebrazione dell'8 marzo in Spagna, per fare un solo esempio), dove le misure di distanziamento sociale e di contenimento sono state introdotte con un ritardo

ancora maggiore.

Quando l’epidemia sarà finita si farà un’analisi degli errori commessi. Ad esempio, commissioni furono istituite per capire gli errori commessi nella gestione del virus Ebola. Nel 2014 un rapporto – è sempre Harari che lo racconta – criticò l’Organizzazione mondiale della sanità (OMS) per la sua reazione inadeguata; la responsabilità della pandemia fu attribuita alla corruzione e all’inefficienza della sezione africana dell’OMS. Altre critiche furono rivolte alla comunità internazionale per non aver risposto all’emergenza in tempo e con gli strumenti necessari. A pochi interessava occuparsi di un virus che colpiva l’Africa. Oggi sta accadendo la stessa cosa per la diffusione del coronavirus. Se è complicato raggiungere una cooperazione perfino in Europa, è purtroppo ancora più difficile aiutare l’Africa e le altre nazioni povere. Occorre fare di tutto per non ripetere errori ricorrenti. Il colera è scomparso nei paesi ricchi ma fa ancora 100.000 morti all’anno nel resto del mondo.

Parlando di coronavirus stiamo spesso usando statistiche sbagliate

L’obiettivo di questa nota è discutere come l’iniziale sottovalutazione del virus è stata poi accompagnata da un utilizzo disomogeneo ed errato di molte statistiche. È un errore infinitamente meno grave degli errori nella gestione della malattia, ma vale la pena di essere discusso, anche perché numeri sbagliati inducono messaggi errati tra le popolazioni, soprattutto nelle porzioni meno istruite. Il problema è acuto in Italia, perché i confronti internazionali mostrano che siamo quasi sempre agli ultimi posti per educazione matematica, educazione statistica, educazione finanziaria (mentre eccelliamo, purtroppo, nelle varie forme di educazione siberiana, e, per fortuna, in molte manifestazioni dell’educazione sentimentale).

La prima questione riguarda il numero dei contagiati. Essi sono stati determinati secondo una procedura che prevede l’accertamento della positività solo in presenza di alcune condizioni legate alla sintomatologia e alle frequentazioni di malati accertati. La scelta, indispensabile nelle prime settimane per la necessità di indirizzare le risorse verso le situazioni di maggiore rischio, non ha consentito una piena valutazione dell’effettivo numero di contagiati, spesso asintomatici. Ciò è avvenuto per vari motivi.

In primo luogo c’è stata una certa variabilità sul territorio nel mettere in pratica le indicazioni generali. In alcune regioni il numero dei tamponi è stato molto più elevato che in altre, riflettendosi in qualche modo nel numero dei contagiati rilevati. Il fatto che la disponibilità a svolgere tamponi sia cambiata nel corso del periodo di *lockdown* ha inoltre reso anche la dinamica di questa serie piuttosto incerta. Più in generale, però, sin dalle prime fasi è risultato chiaro che non venivano considerati quanti avevano contratto il virus in modo lieve o asintomatico, se non grazie ad alcune stime fornite da studiosi indipendenti che hanno cercato di colmare questa lacuna.

Alla fine di marzo l’Imperial College aveva stimato un numero, giudicato poco verosimile, di quasi 6 milioni di contagiati in Italia. L’Istituto Superiore di Sanità italiano aveva replicato fornendo una stima di 700.000 contagiati. È un numero vicino alla stima fornita dall’Istituto di Studi di Politica Internazionale (ISPI), circa 600.000 contagiati, con un campo di variazione molto vasto. Però intorno alla metà di aprile la stima fornita dall’Imperial College è stata giudicata non del tutto campata per aria. La Fondazione Hume ha ipotizzato un campo di variazione dei contagiati che va da 2 a 12 milioni. A fine aprile l’Istituto Superiore di Sanità ha fornito una stima di circa 3 milioni di contagiati asintomatici.

Va inoltre considerato il tema dello sfasamento temporale. Legare il conteggio dei contagiati alla manifestazione dei sintomi, come fatto in alcune regioni, vuol dire guardare costantemente una situazione non aggiornata, visto che i sintomi si manifestano dopo diversi giorni dal contagio. Nelle fasi iniziali del contagio, ogni 4-5 giorni si osservava un raddoppio delle persone coinvolte; diffondere per definizione dati non aggiornati, e quindi soggetti a sottostime consistenti, può avere avuto gravi ripercussioni sulla percezione del fenomeno e sulle misure intraprese. Come vedremo è stato un errore comune ad altri paesi.

È probabilmente da collegarsi alla sottostima dei contagiati, oltre al fatto che il nostro paese ha una grande quantità di anziani, il più l’elevato rapporto morti/affetti da coronavirus riscontrato in Italia rispetto ad altri paesi. Il numero che ci viene fornito, talvolta definito tasso di letalità, ha poco senso, perché noi non conosciamo il numero delle persone colpite da coronavirus. Come sopra riassunto, in Italia il denominatore del tasso di letalità è in realtà molto più alto dei numeri presentati. Ciò determinerebbe un tasso di letalità molto più basso.

Purtroppo, con il passare dei giorni, abbiamo capito che anche la statistica del numero dei morti non è esente da possibili e consistenti errori di misura. Al di là della linea di demarcazione, che in alcuni casi può essere difficile da tracciare, tra i morti “per” e “con” coronavirus (questione che veniva sovente ripetuta nelle prime fasi dell’epidemia), va segnalato che, al di fuori del perimetro degli ospedali, non sempre i morti da coronavirus sono stati collegati all’epidemia in corso, perché non si sono sempre eseguite le relative procedure di accertamento, in una situazione che rimane ancora adesso difficile in molte province. È un’osservazione più volte ripetuta dai sindaci lombardi e di altre regioni settentrionali.

Un metodo che ha permesso di evidenziare che i decessi ufficiali erano largamente sottostimati è stato quello di operare un confronto del numero dei morti nei mesi in cui ha avuto luogo l’epidemia con i dati degli anni passati, con la ragionevole approssimazione di addebitare al coronavirus la differenza sistematica di mortalità.

I dati forniti dall’Istat hanno riguardato un sottoinsieme dei comuni italiani, selezionati con modalità che purtroppo hanno reso difficile la stima complessiva a

livello dell’intero Paese (si vedano, ad esempio, gli interventi di Rettore e Tonini e di Borin et al. su lavoce.info). Sono però risultati chiari i reali ordini di grandezza dei morti da coronavirus, molto superiori a quelli forniti ufficialmente. Nei comuni più prossimi all’epicentro lombardo, il numero dei morti nel 2020 è risultato anche molte volte quello riscontrato storicamente negli anni passati.

In presenza di dati statistici così lacunosi, alcuni studiosi per monitorare l’avanzamento dell’epidemia hanno provato a utilizzare il numero dei malati in terapia intensiva, numero che però nelle fasi più acute della crisi può essere stato limitato dalle dotazioni del sistema sanitario (nell’emergenza, alcuni malati sono stati spostati in Germania).

Per sintetizzare, il numero dei contagiati che soprattutto nelle prime fasi dell’epidemia ha invaso le case degli italiani alle 18 del pomeriggio, è stato ampiamente sottostimato, così come il numero dei deceduti (ed entrambi i dati hanno riflesso differenze di comportamento tra le regioni). Il rapporto tra due quantità mal stimate, i morti e i colpiti dalla malattia, non può essere una buona statistica e infatti si confronta a fatica con i dati di altri paesi. Per usare le parole dell’ISPI, è solo un indicatore della letalità apparente, non di quella reale. Per avere il vero tasso di letalità, occorrerebbe avere stime non distorte del numeratore e del denominatore del rapporto. Ugualmente poco informativi sono risultati i dati sul numero dei ricoveri in ospedale, ampiamente soggetti a valutazioni estemporanee sull’opportunità di tenere o meno a casa i contagiati, e quelli sul numero dei guariti, influenzati anch’essi da modalità diverse adottate a livello regionale.

Insomma, una Caporetto statistica, che ha riguardato tutti i paesi del mondo e che trova parziale giustificazione nell’entità della crisi e nella sua improvvisa comparsa. Basti pensare alle statistiche sui morti che presentano chiare discontinuità in Francia ogni volta che arrivano i dati dalle residenze per anziani; o i valori elevati del Belgio che considerano morti da Covid-19 tutti quelli che hanno i sintomi, anche senza tamponi; o alla Germania, il paese con meno morti, che non sembra aver diffuso statistiche sul numero dei morti per tutte le cause (cfr. il Financial Times).

Il confronto internazionale dei dati forniti (o non-forniti, come in Germania) sembra mostrare che le statistiche sulla pandemia non sono armonizzate e non sono quindi confrontabili. Come ha sintetizzato l’epidemiologo Alessandro Vespignani “i dati di qualsiasi epidemia all’inizio sono sempre – mi [si] passi il

romanismo – un casino”.

Caporetto, la resistenza sul Piave e la battaglia di Vittorio Veneto

Negli ultimi 25 anni, la qualità delle statistiche economiche europee è migliorata. Abbiamo raggiunto grandi progressi nelle statistiche della contabilità nazionale, delle banche, della bilancia dei pagamenti, dei bilanci delle famiglie, della finanza pubblica. I progressi sono stati in buona parte legati all’integrazione europea e alla nascita dell’area dell’euro, che hanno costretto a una armonizzazione delle statistiche dei paesi. Dati non armonizzati sono inutili, perché inconfrontabili. L’armonizzazione e il miglioramento delle statistiche europee sono stati il frutto della collaborazione tra l’Eurostat (l’ufficio statistico della Commissione europea), gli istituti di statistica nazionali, la Banca centrale europea e le banche centrali nazionali.

Sono stati compiuti errori nel calcolo delle statistiche? Certo: come dice Popper è impossibile non fare errori. Il fallimento clamoroso delle statistiche è stata la falsificazione dei dati di finanza pubblica in Grecia, scoperta nel 2009. Sono seguiti rapporti, confronti tra i paesi, riforme, che hanno portato a nuovi miglioramenti delle statistiche – dagli errori si impara – in particolare in Grecia. Gli indicatori economici sono importanti, ma sicuramente lo sono meno dei dati sulla diffusione del virus. Le democrazie si fondano sulla disponibilità di buone statistiche, la cui produzione non deve essere messa in crisi neanche da eventi eccezionali.

In campo sanitario molto è già stato fatto e molto certamente verrà fatto in futuro nella direzione di una maggiore omogeneità internazionale nella raccolta dei dati. Esistono indicazioni europee in questo senso (qui un esempio), ma non allo stesso livello di precisione e di ampiezza esistenti per gli indicatori economici. Uno dei motivi è che le dimensioni dei fenomeni economico-finanziari sono spesso europei o globali, mentre nei fatti la dimensione sanitaria è principalmente locale. Ma la pandemia, con la rapida diffusione del virus da un paese a un altro, forse cambierà tutto.

Vi è oggi consapevolezza della necessità di un cambio di passo sulla raccolta di statistiche sulla diffusione del virus. In fondo, dopo Caporetto vi fu la vittoria di

Vittorio Veneto. In Italia già da alcune settimane si parla della necessità di svolgere indagini che producano statistiche più attendibili sul numero di quelli che hanno il virus (o che lo hanno avuto) e il numero di quelli che lo hanno evitato, dei guariti, dei morti e che chiarisca, per quanto possibile, aspetti rilevanti nel disegno delle politiche, come ad esempio la contagiosità dei guariti nel tempo e la loro eventuale immunità. È necessario svolgere indagini su campioni rappresentativi della popolazione, per sesso, età e distribuzione geografica, svolgendo eventualmente anche test clinici (avendo ben presenti le grandi difficoltà e incertezze del caso). Dalla fine di marzo diverse proposte sono andate in questa direzione (qui un es).

Il controllo di campioni rappresentativi della popolazione era impossibile nelle prime settimane della pandemia, dove giustamente tutti gli sforzi sono stati indirizzati a salvare vite umane. La buona notizia è che in aprile l'Istat ha annunciato un'indagine sierologica, d'intesa con l'Istituto Superiore di Sanità, partita il 4 maggio e diretta a misurare quante persone sono state contagiate dal virus, costruendo un campione di 150.000 individui dislocati in 200 comuni di tutte le regioni italiane. È un campione statisticamente rappresentativo della popolazione, per genere e per età, con un dettaglio particolare per le età superiori ai 50 anni, e per quattro differenti attività economiche. Il test sierologico ci consentirà di sapere, pur con un margine di errore, quanti italiani sono entrati in contatto con il virus, fornendo informazioni preziose su come disegnare l'allentamento delle misure restrittive.

Ma anche i dati amministrativi devono dare il loro contributo. Ad esempio sarebbe auspicabile migliorare la tempestività e l'estensione della raccolta dei dati sulla mortalità a livello territoriale, soprattutto in alcune regioni, rafforzando gli attuali sistemi di allerta sanitaria. È un investimento che oggi può tornare utile per il monitoraggio di nuove ondate epidemiche, ma che domani potrà segnalare situazioni critiche legate a fattori ambientali che necessitano di interventi mirati.

Le stesse informazioni raccolte nella fase di monitoraggio dei contatti, il c.d. *tracing*, dovrebbero essere utilizzate – nel rispetto della privacy – anche per migliorare la nostra comprensione su come l'epidemia si diffonde nella popolazione, aiutandoci in questa fase di convivenza con il virus.

Come ha ricordato l’Associazione Italiana di Epidemiologia, per limitare il rischio di nuove ondate epidemiche sono necessarie azioni, come “il rafforzamento delle attività di identificazione dei casi, il loro isolamento extradomiciliare, l’identificazione tempestiva dei contatti, l’estensione delle capacità di accertamento virologico dell’infezione a partire delle categorie ad alto rischio e l’identificazione dei focolai di contagio. … queste attività necessitano di sistemi di raccolta dei dati in modo standardizzato, che consentano una loro tempestiva registrazione ed elaborazione, in grado di guidare la valutazione del rischio e le relative contromisure. È fondamentale che tutte le aree del Paese adottino strumenti efficaci e condivisi.”

Insomma “testare”, “tracciare” e “trattare”. Statistiche migliori sul virus contribuiranno a realizzare la profezia di Harari di un mondo in grado di prevenire le epidemie.

coronavirus-2-1.jpg

Se continuiamo a tenere vivo questo spazio è grazie a te. Anche un solo euro per noi significa molto. Torna presto a leggerci e SOSTIENI DOPPIOZERO