



COSA “POSSONO” DIRE I DATI SULLA MORTALITÀ DEL CAMPIONE ISTAT?

**ALESSANDRO BORIN, DOMENICO DEPALO, GIUSEPPE
ILARDI, ALFONSO ROSOLIA E LUCA ROSSI¹**

L'Istat ha recentemente reso disponibili dati sul numero giornaliero di decessi in 1.450 comuni italiani che hanno registrato un forte aumento della mortalità nei primi mesi del 2020. È possibile che proprio questa modalità di selezione dei comuni per cui diffondere i dati comporti una sovrastima della mortalità in eccesso. Tuttavia, il lavoro mostra che questa potenziale distorsione è meno rilevante al crescere della quota di popolazione selezionata in una data area. Ne consegue che nelle aree più colpite, come nel caso delle province lombarde per le quali i dati si riferiscono a un ampio numero di comuni, i dati Istat rappresentano una misura affidabile – e in alcuni casi forse prudenziale – dell'effettiva mortalità.

Per fornire tempestivamente informazioni aggiornate sull'andamento della mortalità in Italia, l'Istat ha pubblicato i dati sul numero di decessi giornalieri per i primi quattro mesi dell'anno per il periodo 2015-2020. Fino al 2019 le informazioni sono disponibili per tutti i comuni italiani; per il 2020 i dati, riferiti al periodo dall'1 gennaio al 28 marzo sono relativi a 1.450 tra i circa 6.000 comuni che partecipano nell'Anagrafe Nazionale della Popolazione Residente (ANPR); questi comuni sono inclusi nel campione diffuso dall'Istat per il 2020 se hanno registrato (a) almeno 10 decessi tra l'1 gennaio e il 28 marzo 2020 e (b) un aumento della mortalità totale nelle prime quattro settimane di marzo di almeno il 20% rispetto alla media osservata per lo stesso periodo nel quinquennio 2015-2019.

Come sottolineato da Enrico Rettore e Sara Tonini (2020) in due recenti articoli pubblicati su *LaVoce.info*, l'applicazione di criteri di selezione del campione direttamente collegati alla variabile di studio può creare delle distorsioni nelle stime di interesse. Per questa ragione, come sottolineato anche nella documentazione fornita dall'Istat, i risultati delle analisi basate su questi dati vanno considerati con grande cautela (Colombo e Impicciatore 2020; Istituto Cattaneo. ISTAT 2020). Ciononostante, questi dati possono offrire importanti indicazioni, in alcuni casi addirittura prudenziali, sull'andamento della mortalità, in particolare nelle aree geografiche più colpite dall'epidemia.

Nel loro articolo, Enrico Rettore e Sara Tonini mostrano come applicando la regola di selezione dell'Istat a un anno del periodo 2015-19 non interessato dal Covid-19 si giungerebbe per costruzione

¹ Banca d'Italia. Le opinioni espresse sono personali e non riflettono necessariamente la posizione della Banca d'Italia.

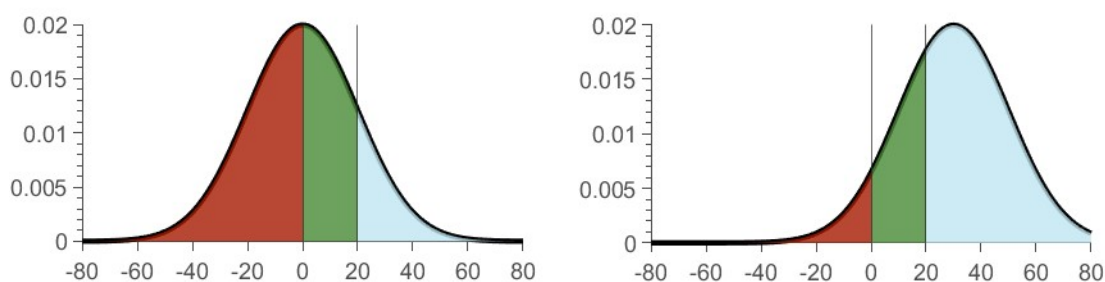
a selezionare un campione caratterizzato da una mortalità superiore alla media, a fronte della *minore* mortalità nei rimanenti comuni. L'esercizio è molto interessante e dimostra perché, in condizioni normali, le stime basate su questo campione possano essere distorte per eccesso. Tuttavia, il 2020 non è un anno normale. In alcune aree del Paese lo shock indotto dal Covid-19 ha comportato un drammatico aumento nei tassi di mortalità che, tra le altre cose, modifica sostanzialmente la natura della selezione non casuale del campione e le distorsioni che ne derivano². Questo avviene essenzialmente per due ragioni.

1) Rispetto a un anno normale, se lo shock alla mortalità fosse così ampio da spostare tutti comuni oltre la soglia, il criterio di selezione non sarebbe stringente e il campione selezionato corrisponderebbe all'universo di interesse. In questo caso, le stime basate su questi dati non sarebbero soggette ad alcuna distorsione.

2) Coerentemente con la natura spaziale del contagio, è probabile che nelle aree più colpite anche i comuni non selezionati dal criterio adottato abbiano registrato una maggiore mortalità, ma non tale da rientrare nel campione; questa mortalità in eccesso pertanto sfuggirebbe alla misurazione.

La figura 1 illustra questi due canali. Il riquadro di sinistra riporta un'ipotetica distribuzione degli scostamenti annuali della mortalità comunale da una media storica di riferimento, per convenienza centrata sullo zero e la cui dispersione riflette fattori e shock idiosincratichi. Il riquadro di destra rappresenta la distribuzione degli scostamenti tra gli stessi comuni in un anno in cui tutti sono colpiti da uno shock comune che aumenta il tasso di mortalità. La barra verticale indica uno scostamento del 20%, coerente con la regola adottata dall'Istat. È evidente che la massa al di sopra della soglia dopo lo shock (area azzurra) è molto superiore a quella colta in un anno normale, e lo è tanto più quanto più forte è lo shock. Allo stesso modo, dopo lo shock la regola di selezione trascura un ampio insieme di comuni in cui l'aumento di mortalità, pur evidente, non è abbastanza forte da indurne l'inclusione nel campione (area verde).

Figura 1: Esempio di una traslazione verso destra della distribuzione della mortalità in eccesso

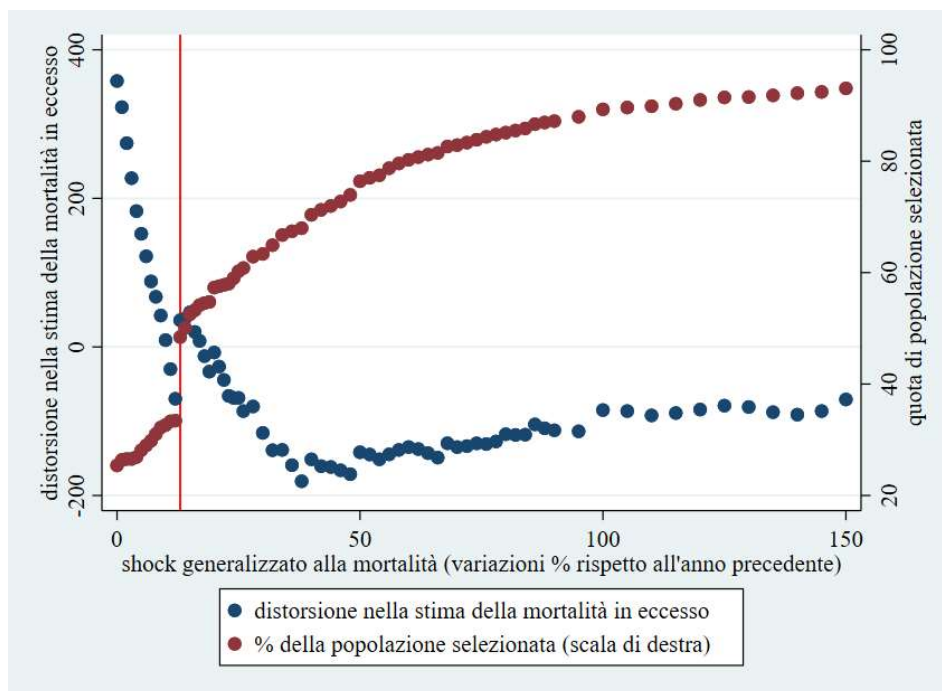


Per quantificare la rilevanza di questi due canali abbiamo considerato i dati disponibili per i comuni della Lombardia per il 2019, e abbiamo simulato una serie di variazioni generalizzate della mortalità di diversa entità. Gli shock sono stati definiti in termini di aumento percentuale rispetto al 2018, da 0 a 150%. Si tratta di un approccio prudentiale, in quanto lo shock non viene applicato ai comuni più piccoli che presentavano mortalità pari a zero nel marzo dell'anno precedente. I risultati delle simulazioni sono riportati nella Figura 2. In linea con quanto mostrato da Enrico Rettore e Sara Tonini, si può notare che, in assenza di shock, il criterio Istat seleziona circa un quinto della

² Evidenze in questo senso emergono dal "[Rapporto settimanale](#)" del Sistema di Sorveglianza della Mortalità Giornaliera del Ministro della Salute basata su un campione di 19 grandi comuni seguiti nel tempo. Secondo questi dati, partendo da livelli storicamente molto ridotti in gennaio e febbraio, a marzo la mortalità è passata rapidamente a valori estremamente elevati.

popolazione lombarda e determina una sovrastima di circa 400 decessi per il mese di marzo 2019. La distorsione si riduce al rafforzarsi dello shock, e coerentemente cresce la quota di popolazione residente nei comuni selezionati. La sovrastima si annulla in corrispondenza di uno shock generalizzato a tutti comuni della Lombardia pari al 12% della mortalità, che induce una selezione di circa il 40% della popolazione nel campione. Shock più forti comportano addirittura una sottostima della mortalità in eccesso, dovuta all'esclusione dei comuni che pur registrando una mortalità maggiore non rientrano nella soglia (aree verde e rossa della figura 1). Shock di intensità ancora maggiore sono associati a una progressiva riduzione della sottostima, riconducibile al fatto che con ampi shock la maggior parte dei comuni rientra nel campione fino a che questo coincide con l'universo di riferimento.

Figura 2: Simulazione dell'effetto di uno shock alla mortalità ai comuni della Lombardia nel 2019

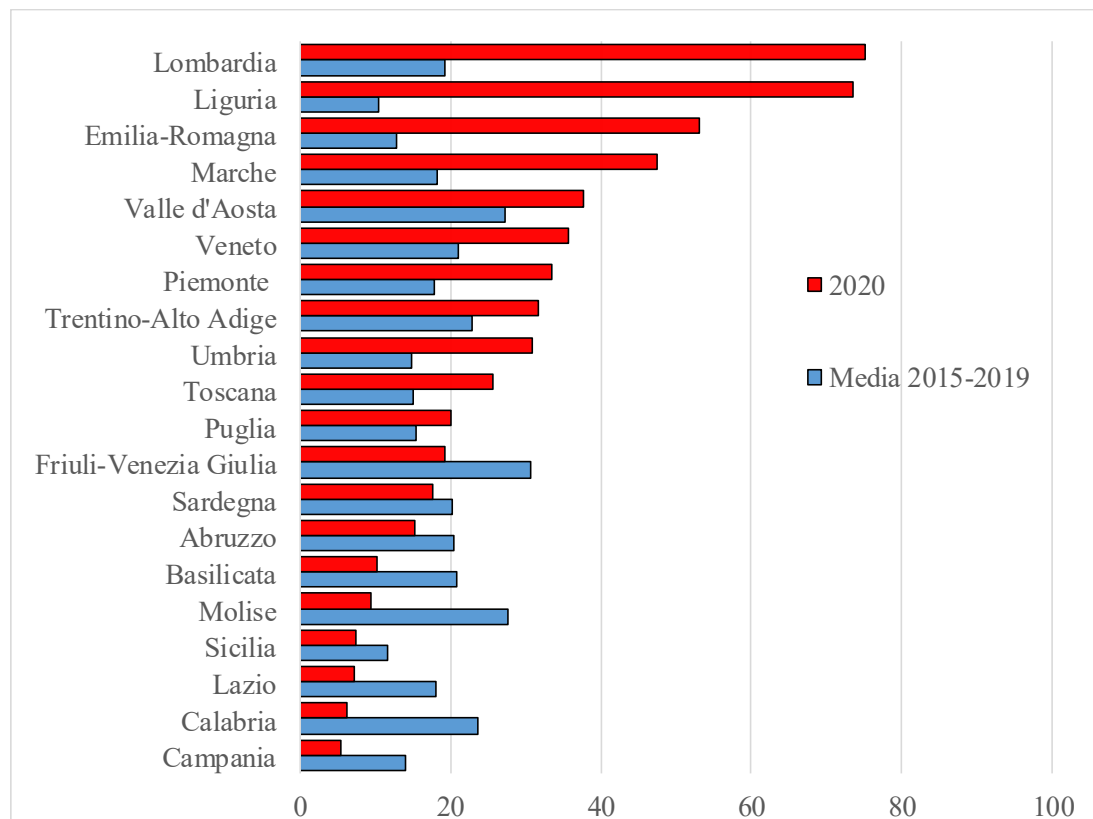


L'esercizio precedente suggerisce quindi che la quota di popolazione che rientra nel campione selezionato dall'Istat sia un buon indicatore dell'affidabilità delle stime e del segno dell'eventuale distorsione. La figura 3 riporta tali quote per le regioni italiane nel quinquennio 2015-2019 e nel 2020 per il campione di comuni ANPR. A fronte della sostanziale omogeneità tra le regioni tra il 2015 e il 2019, con quote di popolazione incluse nel campione selezionato tra il 10 e 25%, nel 2020 si osserva una notevole eterogeneità. In Calabria, Campania, Lazio e Sicilia la quota di popolazione selezionata è inferiore al 10%, mentre in Lombardia e Liguria supera il 70%.

Come sostenuto da Rettore e Tonini, queste evidenze suggeriscono che utilizzare il campione Istat per le regioni in cui la diffusione del contagio è stata relativamente contenuta (Sud, Isole e Lazio), cioè in situazioni pressoché normali, induca una sostanziale sovrastima della mortalità effettiva. Per contro, per molti regioni del Nord il campione fornirebbe una stima affidabile e addirittura plausibilmente cautelativa dell'effettivo eccesso di mortalità, dal momento che sarebbero ragionevolmente esclusi comuni con mortalità comunque maggiore della media storica. Quest'ultimo punto è più efficacemente analizzato sfruttando l'eterogeneità territoriale del contagio, che ci permette di fornire qualche valutazione in più anche per le regioni in cui la quota di popolazione selezionata nel 2020 risulta di poco superiore alla media storica. Ad esempio, per la maggior parte delle province della Toscana la quota inclusa nel campione non differisce significativamente da quella

degli anni precedenti, ma per tre province (Massa-Carrara, Pisa e Pistoia) l'incremento è rilevante e quindi le stime di eccesso di mortalità su queste aree potrebbero essere affidabili (Figura 4). Al contrario, nel caso dell'Emilia-Romagna molte province risultano ampiamente rappresentate nel campione Istat del 2020, tuttavia per quelle di Bologna e Ferrara la quota selezionata non è sostanzialmente diversa dal passato, per cui non è opportuno effettuare valutazioni sulla variazione della mortalità legata al Covid-19.

Figura 3: Quota di popolazione di ciascuna regione rappresentata dai comuni selezionati dal criterio ISTAT

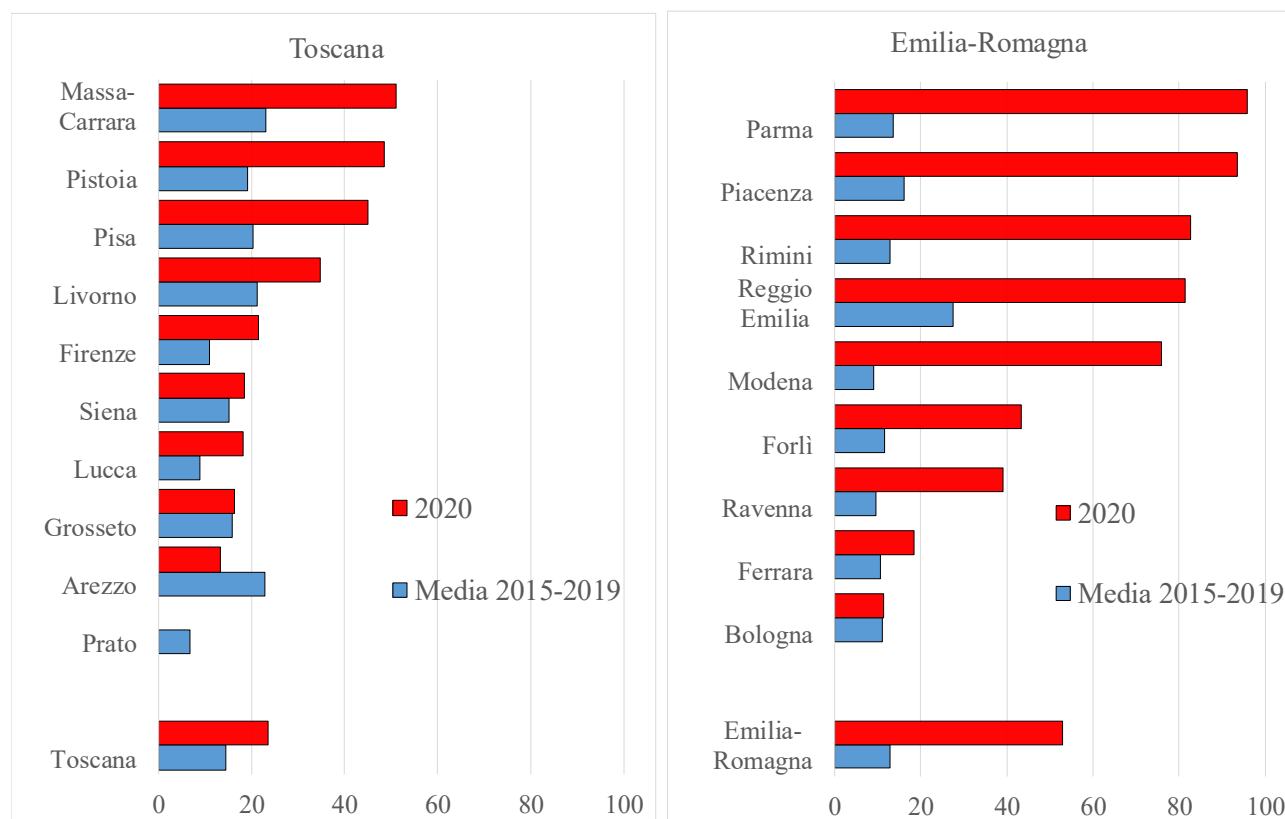


Queste considerazioni indicano che il campione di comuni selezionato dall'Istat può fornire elementi utili e importanti per misurare in modo affidabile la mortalità in eccesso riconducibile, direttamente o indirettamente, al Covid-19 soprattutto nelle aree maggiormente interessate dal contagio e per cui è più urgente elaborare una strategia sanitaria di risposta che tenga pienamente conto degli effetti complessivi dello shock.

Prima di concludere, è opportuno aggiungere alcune considerazioni generali riguardo alla possibilità di confrontare le stime sull'eccesso di mortalità con i dati sui decessi per Covid-19 certificati sulla base di test di laboratorio. Innanzitutto, sebbene l'eccesso di mortalità sia con ogni probabilità collegabile all'epidemia, non è possibile tramite questo esercizio quantificare in che misura questo sia imputabile direttamente all'epidemia e in che misura sia attribuibile ad altre cause connesse (es. allo stress del sistema sanitario in alcune regioni). Vi potrebbero essere anche fattori di segno opposto, come il livello di mortalità stagionale più basso della norma associato all'inverno particolarmente mite, la riduzione delle morti bianche e di quelle per incidenti stradali a causa del *lockdown*. Infine, un'ultima considerazione riguarda il periodo di osservazione: questo non coglie né i decessi della diffusione del virus a marzo, né la sua natura endemica, né la possibile anticipazione di decessi che sarebbero comunque avvenuti nel giro di poco tempo, considerata la maggiore letalità del virus per i

sogetti più fragili. Pertanto, una valutazione precisa dell'effetto del Covid-19 sulla mortalità dovrà attendere la diffusione dei dati sul numero e sulle caratteristiche dei decessi per tutti i comuni italiani su un orizzonte temporale più ampio.

Figura 4: Quota di popolazione di ciascuna provincia della Toscana e dell'Emilia-Romagna rappresentata dai comuni selezionati dal criterio ISTAT



Nel frattempo, le informazioni che l'Istat diffonde con eccezionale tempestività rappresentano una solida base da cui partire per valutare l'ampiezza e la diffusione dello shock e su cui impostare analisi e risposte di policy.

Bibliografia

Colombo, A.D., e R. Impicciatore, "[La crescita della mortalità ai tempi del Covid-19](#)", Istituto Cattaneo, 1 aprile 2020.

ISTAT, "[I decessi del 2020. Dati anticipatori sulla base di un sottoinsieme di comuni del sistema ANPR](#)", 9 aprile 2020.

Rettore, E., e S. Tonini, "[Decessi da Covid, facciamo chiarezza sui dati Istat](#)", *lavoce.info*, 2 aprile 2020.

Rettore, E., e S. Tonini, "[Morti da coronavirus: calcoli sul campione inadatto](#)", *lavoce.info*, 6 aprile 2020.

Sistema di sorveglianza della mortalità giornaliera, "[Rapporto settimanale. Settimana 25-31 Marzo](#)".