

# DAL COLERA DEL 1854 AL COVID-19

*Abstract.* John Snow's method of limiting contagion and death in the 1854 cholera epidemic is one of the foundations of modern epidemiology: a fine example of how data, statistical information and knowledge can make an essential contribution. Yet today, in the course of the COVID-19 pandemic, what Snow did cannot be done. I refer to the Italian case (although some points can be easily generalised) to identify some obstacles and propose possible solutions: we do not know where the infected live and where they have become infected; we do not know their demographic and socio-economic characteristics; we do not have sufficient information on mobility; we do not make full use of the conceptual tools we have; the principle of privacy overrides the principle of public health. In the conclusions, I offer an example of what could be done if these obstacles were removed. In particular, I refer to a recent paper to illustrate the contribution that statistics and mathematical models could make to inform public opinion and to provide authorities with suggestions for more effective and equitable pandemic policies.

*Keywords:* COVID-19, statistical information, models, privacy.

## 1. Uno strano monumento

Nel cuore di Londra, a Soho, c'è uno strano monumento: una fontanella, dove attingeva chi non aveva l'acqua corrente in casa.

La storia di quella fontanella inizia nel 1854, il 31 agosto. Gli epidemiologi parlano di una pandemia di colera tra il 1846 e il 1860, e a Londra l'infezione serpeggia già da tempo. Ma quel giorno scoppia a Soho un'epidemia importante: nei primi tre giorni muoiono 127 persone. Anche se i tre quarti degli abitanti si allontanano dal quartiere, il 15 settembre i morti sono saliti a 500 e alla fine dell'epidemia a 616.

Soho è un quartiere povero e degradato. Stalle, mattatoi, povere abitazioni, bordelli e camere a ore si trovano fianco a fianco. Le strade

---

\* Libero ricercatore

piene di sterco di cavallo e di rifiuti. L'ambiente ideale per il diffondersi del colera.

La maggioranza dei medici ritiene infatti che l'infezione sia diffusa dai miasmi (come per la malaria); una minoranza l'attribuisce ai germi (ma la teoria di Pasteur si afferma solo nel decennio successivo). Tra i fautori della teoria dei germi c'è John Snow, che si mette al lavoro sull'epidemia londinese. Sospettando da tempo che l'agente infettivo sia legato all'acqua inquinata dalle feci umane, mette in piedi un esperimento a doppio cieco, esaminando la suscettibilità al contagio di famiglie rifornite da diversi acquedotti (tutta l'acqua di Londra proviene dal Tamigi, ma è filtrata e trattata con maggior o minore efficacia dai diversi fornitori). Soprattutto, ha l'idea di rappresentare sulla mappa di Londra i bacini d'utenza di tutte le fontane pubbliche, definendo le singole 'celle' di strade e isolati sulla base della loro vicinanza alle fonti d'approvvigionamento: un'applicazione di quelli che sarebbero divenuti noti come diagrammi di Voronoi. Poi segna sulla mappa con un punto ogni caso di colera diagnosticato. È subito evidente che il maggior numero di casi (con 61 morti) è in prossimità della fontanella di Broad Street. Il sospetto che sia quello il focolaio dell'infezione trova conferma in due circostanze: è alimentata da un pozzo, e non dall'acquedotto; e il pozzo è scavato a meno di un metro di distanza da una fossa settica. Qualche eccezione serve da test dell'ipotesi: alcune famiglie della zona preferiscono attingere a una fontanella più lontana, perché ritengono l'acqua migliore, e non si ammalano; tre bambini morti di colera non abitano in zona, ma vanno a scuola lì e hanno bevuto alla fontanella; nessuno degli operai della fabbrica di birra di Broad Street si ammala, perché nel processo di produzione l'acqua si deve bollire. Il 7 settembre si rimuove la maniglia della pompa e in pochi giorni l'epidemia si placa<sup>1</sup>.

## **2. Quello che ha fatto John Snow oggi non si può fare**

Il metodo seguito da Snow per limitare contagi e morti nel colera del 1854 è una delle basi dell'epidemiologia moderna: un bell'esempio di come dati, informazione statistica e conoscenza possano dare un contributo essenziale. Eppure oggi, nel corso della pandemia COVID-19, quello che ha fatto John Snow non si può fare. Non lo può fare – certamente – né la comunità scientifica, né il singolo ricercatore. Non

---

<sup>1</sup> Lo stesso John Snow ha pubblicato i suoi risultati (Snow, 1855).

lo possono fare nemmeno – temo – le autorità sanitarie e politiche cui è demandato il governo delle misure di contrasto alla pandemia.

Faccio riferimento al caso italiano (anche se alcuni punti si possono generalizzare) per individuare alcuni ostacoli e proporre possibili soluzioni:

- non conosciamo i luoghi del contagio (dove dimorano i contagiati e dove si sono infettati);
- non conosciamo le caratteristiche demografiche dei contagiati;
- non ne conosciamo le caratteristiche socio-economiche;
- non abbiamo informazioni sufficienti sulla mobilità;
- non utilizziamo appieno gli strumenti concettuali che abbiamo;
- il principio della *privacy* sovrasta quello della salute pubblica.

Vediamo ora i punti uno per uno.

### *2.1. Non conosciamo i luoghi del contagio*

Per essere più precisi, non conosciamo né dove risiedono i contagiati (ciò che invece Snow non ha avuto difficoltà a conoscere in breve tempo, sia pure su scala molto più piccola), né dove si sono infettati.

Quanto al primo punto, la *Scheda di dimissione ospedaliera* – principale strumento di raccolta delle informazioni relative a tutti i ricoveri nelle strutture ospedaliere pubbliche e private sul territorio nazionale<sup>2</sup> – registra, tra l’altro, l’indirizzo di residenza del paziente. Anche senza considerare i problemi di accuratezza e completezza che affliggono le rilevazioni a carattere amministrativo (il cui primo intento è quello della gestione di un servizio, e non quello della rilevazione di dati statistici), le schede sono trasmesse dopo le dimissioni per guarigione o decesso del paziente, e non sono perciò sufficientemente tempestive per il monitoraggio di un’epidemia. Inoltre, a mia conoscenza non esiste un analogo flusso codificato di informazioni per gli interventi diagnostici come i tamponi e gli esami sierologici, soprattutto se erogati da strutture private. Né mi risulta che questi dati – se esistono – siano messi a disposizione della comunità scientifica.

---

<sup>2</sup> Per maggiori informazioni si veda: [www.salute.gov.it](http://www.salute.gov.it).

Quanto al secondo punto, in un’audizione informale<sup>3</sup> presso la XII Commissione permanente Igiene e sanità del Senato della repubblica, il presidente dell’Accademia nazionale dei lincei, prof. Giorgio Parisi, è intervenuto proprio sui temi del ricorso ai test e al tracciamento per il contenimento della pandemia. Tra le molte cose interessanti che ha detto, ha presentato i dati raccolti dal Robert Koch Institut sui luoghi del contagio in Germania (RKI, 2020). Con riferimento alle 20 tipologie in cui lo studio suddivide le situazioni in cui l’infezione è stata trasmessa, emergono con chiarezza l’ambito familiare, quello medico-ospedaliero, il luogo di lavoro e la ristorazione (peraltro con un numero di contagi molto diverso per ogni tipo di focolaio). Ma l’evidenza più clamorosa è che in tre casi su quattro il luogo del contagio è ignoto: quando è avvenuto in luoghi in cui il contatto è casuale (mezzi pubblici, negozi, palestre), non ripetuto e sistematico (abitazioni, uffici e ospedali), rintracciare le catene di trasmissione è molto più difficile.

## 2.2. Non conosciamo le caratteristiche demografiche dei contagiati

La conoscenza delle caratteristiche demografiche dei contagiati incontra un secondo ostacolo – accanto a quello della non esaustività e della scarsa tempestività dei dati delle *Schede di dimissione ospedaliera* – nella struttura delle informazioni anagrafiche. In Italia le anagrafi sono organizzate su base comunale e soltanto nel 2005, con l’entrata in vigore del *Codice dell’amministrazione digitale*, è stata istituita un’anagrafe nazionale della popolazione residente (ANPR) presso il Ministero dell’interno<sup>4</sup>. Come spesso accade, per complicazioni amministrative e l’inefficacia degli incentivi e delle sanzioni, l’attuazione è marciata a rilento, anche se il decreto istitutivo prevedeva che le operazioni si completassero entro il 31 dicembre 2014. Invece, alla fine del 2016 un solo Comune (quello di Bagnacavallo, RA) era “subentrato” in ANPR. Soltanto nel 2019 e 2020 c’è stata un’accelerazione (in parte stimolata dalla pandemia) che ha portato i Comuni subentrati a 6.858 con una popolazione di poco più di 50 milioni. Sono in corso le attività preliminari per altri 433 Comuni (tra cui Roma) con circa 6,5 milioni di residenti<sup>5</sup>.

I vantaggi di un’anagrafe nazionale completa e funzionante per il monitoraggio e il tracciamento dei contagiati è agevole da intuire: di

---

<sup>3</sup> 10 novembre 2020, ore 14:30. Il video è disponibile *online*: [webtv.senato](https://www.webtv.senato.it/).

<sup>4</sup> Codice dell’amministrazione digitale (D.Lgs. n. 82/2005), art. 62.

<sup>5</sup> [stato-migrazione.anpr.it/](https://stato-migrazione.anpr.it/), consultato il 1° dicembre 2020.

ogni persona sottoposta ad accertamento diagnostico presso qualunque struttura (purché ne sia stata registrata l'identità tramite la *Tessera sanitaria*) conosceremmo – oltre a data ed esito dell'accertamento diagnostico e dell'eventuale malattia – tutti i dati disponibili sui normali certificati anagrafici: nome, cognome, indirizzo, età, genere... Inoltre, l'informazione sarebbe molto tempestiva, dal momento che l'archivio è informatizzato (affidato alla gestione della SOGEI<sup>6</sup>) e “mette a disposizione dei comuni un sistema di controllo, gestione e interscambio, puntuale e massivo, di dati, servizi e transazioni”<sup>7</sup>.

Per inciso, si noti che dal completamento dell'ANPR trarrebbe giovamento anche la statistica ufficiale. Poiché le anagrafi sono a base comunale, l'ISTAT opera tradizionalmente per le sue rilevazioni un campionamento a due stadi: in un primo stadio estrae un campione di comuni (solo in parte casuale, perché i comuni di maggiore dimensione demografica sono sempre presenti nel campione)<sup>8</sup>; nel secondo stadio, all'interno di ciascun comune estratto (o selezionato) si estrae a sua volta un campione casuale secondo un ulteriore piano di campionamento. L'operatività di ANPR allargherebbe le scelte del disegno campionario, rendendo disponibile una lista di individui e famiglie su base nazionale (come già avviene per le rilevazioni sulle imprese) e quindi possibile l'opzione di un campionamento casuale semplice.

### 2.3. Non conosciamo le loro caratteristiche socio-economiche

Oltre alle caratteristiche demografiche di ogni contagiato, sarebbe utile conoscerne anche le caratteristiche socio-economiche: quali sono le condizioni della famiglia cui appartiene (reddito, *status*, abitudini di consumo), se e presso quale impresa lavora (con quale contratto e mansione, in quale settore, a quale indirizzo), e così via.

---

<sup>6</sup> “SOGEI – Società Generale d'Informatica S.p.A. – è la società di *information technology* al 100% del Ministero dell'economia e delle finanze e opera sulla base del modello organizzativo dell'*in-house providing*” ([www.sogei.it](http://www.sogei.it)).

<sup>7</sup> D.Lgs. n. 82/2005, art. 62, c. 3.

<sup>8</sup> Ad esempio, nella *Rilevazione sulle forze di lavoro* “i due stadi del campionamento sono rispettivamente i comuni e le famiglie. I comuni vengono stratificati rispetto alla dimensione demografica su base provinciale.” Nel primo stadio “i comuni sono classificati in base alla dimensione demografica e vengono suddivisi in: 346 autorappresentativi (comuni in cui la dimensione demografica supera una prefissata soglia) e non-autorappresentativi” (per un totale di 1.246 comuni). Il campione è indicativamente di 76.000 famiglie. Cfr. [siqua.istat.it](http://siqua.istat.it).

Affinché questo sia possibile, è necessario che l'ISTAT completi la realizzazione di un'infrastruttura statistica: il “sistema dei registri”, che consentirà la creazione di “nuove basi dati statistiche, svincolate dal disegno d'indagine, micro-fondate su intere popolazioni di riferimento e poste a sistema mediante integrazione concettuale, fisica e statistica delle unità che le compongono” (Monducci, 2019)<sup>9</sup>.

La chiave del sistema dei registri è l'integrazione delle informazioni relative alle singole unità statistiche: a ogni persona (o famiglia, o impresa, o luogo) sono ricondotte le informazioni a essa riferite (rilevate direttamente o tramite procedure di stima), quale che ne sia la fonte (rilevazione statistica o dato amministrativo). Inoltre, i registri sono connessi tra loro logicamente (e informaticamente), in modo da collegare – ad esempio – un individuo al nucleo familiare dove risiede e all'impresa dove lavora, oltre alla geo-referenziazione dell'abitazione e del luogo di lavoro (Wallgren e Wallgren, 2014).

Le statistiche basate sui registri hanno una tradizione secolare: la stima pionieristica della popolazione svedese basata sui registri parrocchiali compiuta dal matematico Pehr Elvius porta nel 1748 alla legge che introduce il registro nazionale e nel 1756 all'istituzione del primo istituto di statistica del mondo (Whitby, 2020). Un'organizzazione degli istituti di statistica basata sui registri e sulla loro integrazione è raccomandata dagli organismi statistici delle Nazioni unite da oltre trent'anni (Sundgren, 1999). Tanto dovrebbe bastare per segnalare il ritardo dell'ISTAT nella modernizzazione del suo modo di operare, anche se sarebbe ingeneroso non ricordare che altri istituti di statistica dei paesi sviluppati, anche assai blasonati, scontano ritardi simili.

#### *2.4. Non abbiamo informazioni sufficienti sulla mobilità*

I dispositivi che portiamo sempre con noi – dal cellulare allo *smartwatch* – sono dotati di uno o più strumenti di geo-referenziazione<sup>10</sup>. Anche se a volte non ne siamo pienamente consapevoli<sup>11</sup>, in genere abbiamo concesso al produttore del nostro

---

<sup>9</sup> Anche il Censimento permanente della popolazione è un passo importante in questa direzione.

<sup>10</sup> Il GPS (Global Positioning System), gestito dal governo statunitense, è solo il più noto dei sistemi di posizionamento globale.

<sup>11</sup> Raramente leggiamo fino in fondo i contratti di licenza che ci permettono l'uso (spesso gratuito) delle applicazioni che usiamo quotidianamente.

apparecchio e ad alcuni fornitori di servizi di utilizzare, in forma anonima, la nostra posizione e i nostri movimenti.

Sono informazioni che sarebbero molto utili per tracciare la pandemia.

Facciamo l'esempio di Google. Se e quando un utente attiva la 'cronologia delle posizioni' (la funzione deve essere approvata esplicitamente), i suoi dati – resi anonimi e aggregati – consentono di creare rapporti sugli spostamenti (Google, 2020a). I rapporti illustrano la variazione delle visite e della durata della permanenza presso luoghi diversi (negozi e tempo libero, alimentari e farmacie, parchi, stazioni del trasporto pubblico, luoghi di lavoro, zone residenziali) rispetto al valore mediano nello stesso giorno della settimana nel periodo di riferimento (3 gennaio-6 febbraio 2020). Per l'Italia, le informazioni sono rese disponibili a tutti e gratuitamente, con aggiornamenti frequenti e tempestivi. Per il Lazio, risulta che al 1° dicembre 2020 gli spostamenti verso gli interscambi del trasporto pubblico sono diminuiti del 47%, verso i parchi del 42%, verso i luoghi di lavoro del 37%, verso i luoghi del commercio e del tempo libero (ristoranti, bar, centri commerciali, parchi a tema, musei, biblioteche e cinema) del 26%, verso i negozi di alimentari e le farmacie del 5%, mentre gli spostamenti all'interno delle zone residenziali sono aumentati del 13% (Google, 2020b).

Sono informazioni utili per valutare l'impatto della pandemia e delle restrizioni introdotte sulle nostre abitudini, ma troppo poco granulari per orientare le strategie di salute pubblica e per prendere decisioni critiche nella lotta a COVID-19. Non so se gli organi di governo abbiano chiesto e ottenuto da Google informazioni di maggiore dettaglio. Se così fosse, saremmo di fronte a un'attenuazione – dettata dall'emergenza sanitaria – dei principi del rispetto della *privacy* dei cittadini. Ma le scelte operate per l'applicazione *Immuni* – che non utilizzano le informazioni di geo-localizzazione – sono un forte indizio nella direzione contraria.

Negli USA, con una normativa completamente diversa da quella in vigore in Italia e nell'Unione europea, una società privata come SafeGraph<sup>12</sup> dispone di tre *dataset* primari:

- *Core places*: informazioni di base – come il nome della località, l'indirizzo, la categoria e l'associazione del marchio per i punti di interesse (POI) dove le persone spendono tempo o denaro – per circa 6,1 milioni di POI;

---

<sup>12</sup> [safegraph.com](https://www.safegraph.com)

- *Geometry*: impronte dei POI con metadati di gerarchia spaziale che rappresentano quando i poligoni ‘figli’ sono contenuti dai ‘genitori’ o quando due ‘inquilini’ condividono lo stesso poligono: disponibile per circa 6,1 milioni di POI;
- *Patterns*: sono disponibili per circa 4,2 milioni di POI e localizzano il traffico e le aggregazioni demografiche che rispondono alle domande: quanto spesso le persone visitano, quanto a lungo si fermano, da dove vengono, dove vanno e altro ancora.

Come vedremo, l’utilizzo di queste informazioni insieme a quelle delle residenze e dei singoli spostamenti tra residenze e POI consente la creazione di modelli previsionali della diffusione e dell’andamento della pandemia molto utili per le politiche di contenimento.

### *2.5. Non utilizziamo appieno gli strumenti concettuali che abbiamo*

Dal 1981 l’ISTAT produce – con cadenza decennale in occasione dei censimenti della popolazione – una zonizzazione funzionale del territorio italiano basato sulla mobilità quotidiana, i Sistemi locali: hanno una dimensione sovra-comunale e nella loro definizione più recente sono 610; tra questi, 503 sono persistenti (erano presenti anche nella zonizzazione definita sulla base del Censimento 2001) e robusti (la loro individuazione non dipende dall’algoritmo utilizzato, ma solo dai dati che ne descrivono le caratteristiche relazionali in termini di flussi pendolari). Rappresentano pertanto “un nocciolo invariante di sistemi locali che costituisce un elemento permanente dell’auto-organizzazione dei territori e ne definisce l’ossatura” (ISTAT, 2015, p. 109).

Ai nostri fini, la loro caratteristica essenziale è quella dell’auto-contenimento dei flussi pendolari giornalieri: all’interno dei sistemi locali i movimenti quotidiani non incontrano ostacoli e sono radicati nelle abitudini delle persone, mentre i loro confini testimoniano dell’esistenza di barriere alla mobilità (ISTAT, 2015, p. 151).

Da questa considerazione discende l’utilità del riferimento ai Sistemi locali per valutare i processi di diffusione della pandemia. I confini amministrativi – comunali, provinciali e regionali – non tengono conto delle necessità e delle abitudini di comportamento delle persone nei loro spostamenti quotidiani a corto raggio.

Come ha autorevolmente affermato il prof. Luigi Biggeri, ex presidente dell’ISTAT, i Sistemi locali “individuati dall’ISTAT sono ‘aree territoriali identificate da un insieme di comuni legati tra loro da flussi degli spostamenti quotidiani per motivi di lavoro e di studio’. È



chiaro che se si verificano contagiati, siano essi asintomatici o sintomatici, questi si spostano soprattutto all'interno di quelle aree. [...] Perché a questo livello sono disponibili anche altri dati – livello di occupazione, intensità della mobilità eccetera – che aiutano a capire se [un nuovo] focolaio può diventare pericoloso o meno. Se scatta un campanello d'allarme, è possibile attuare un tracciamento completo” (Biggeri, 2020).<sup>13</sup>

Inspiegabilmente l'ISTAT – che pure ha una forte tradizione nell'uso e nella valorizzazione dei sistemi locali, e una *leadership* in sede internazionale – non li ha per nulla utilizzati nella sezione del sito dedicata all'emergenza sanitaria.<sup>14</sup>

## 2.6. Il principio della privacy sovrasta quello della salute pubblica

Il testo unico *Codice in materia di protezione dei dati personali*<sup>15</sup> raccoglie e sistematizza la normativa vigente in materia accumulatasi dal 1996<sup>16</sup>. Inoltre, il 4 maggio 2016 è stato pubblicato sulla Gazzetta ufficiale dell'Unione europea il nuovo *Regolamento generale sulla protezione dei dati* (GDPR, *General Data Protection Regulation*)<sup>17</sup>, cui la normativa nazionale si è adeguata nel 2018<sup>18</sup>. Sull'applicazione della normativa vigila l'*Autorità garante per la protezione dei dati personali* (informalmente: “Garante della *privacy*”), istituita dalla legge del 1996 e confermata anche dalla normativa seguente.

Non entro nel merito delle norme: sono complesse e non dispongo delle competenze necessarie, anche se la comprensibilità delle leggi dovrebbe essere un elemento essenziale della democrazia. Noto però una sorta di slittamento progressivo del *focus*, evidente già nei primi tre articoli del *Codice*: l'accento si sposta dalla *tutela dei dati personali* (da verificare *ex post*, in sede di giudizio, verificando la fattispecie specifica, bilanciando gli interessi delle parti) alla *liceità del loro trattamento* (stabilita *ex ante*, in sede di configurazione del sistema informativo, con il controllo preventivo del Garante, un'autorità

---

<sup>13</sup> Altri due ex presidenti dell'ISTAT, Giorgio Alleva e Alberto Zuliani, sono intervenuti più volte con proposte di rilevazioni campionarie utili a soddisfare il fabbisogno di dati e informazioni quantitative sulla pandemia in corso, a partire dalla fine di marzo del 2020 (Alleva et al., 2020).

<sup>14</sup> Si veda in proposito: [www.istat.it/it/archivio/239854](http://www.istat.it/it/archivio/239854).

<sup>15</sup> D.Lgs. n. 196/2003.

<sup>16</sup> In particolare, la legge n. 675/1996.

<sup>17</sup> Regolamento UE 2016/679.

<sup>18</sup> D.Lgs. n. 101/2018.

amministrativa). Segnalo, in proposito, che anche prima della pandemia del COVID-19 molte voci<sup>19</sup> hanno lamentato che i pareri preventivi e le autorizzazioni del Garante rappresentano un ostacolo allo svolgimento della ricerca scientifica, soprattutto in materia statistica. In particolare, i sistemi informativi basati sui registri – per loro natura – non appaiono compatibili con il requisito di “*ridurre al minimo l'utilizzazione di dati personali e di dati identificativi*”.

La situazione sanitaria d'emergenza complica ulteriormente il quadro: senza entrare nel merito degli strumenti giuridici utilizzati, sono state introdotte misure straordinarie con restrizioni a diritti costituzionalmente garantiti (ad esempio, domicilio, circolazione e soggiorno, istruzione, lavoro, iniziativa economica privata ...). Non così per i dati personali, e proprio in virtù del controllo preventivo del Garante, intervenuto in più occasioni<sup>20</sup> a ritenere necessarie: nel caso di uso di dati aggregati e anonimi, la verifica puntuale *ex ante* delle caratteristiche specifiche del trattamento; nel caso di uso di dati individuali, l'approvazione di una norma apposita, con efficacia limitata temporalmente, rispettosa dei principi di proporzionalità, necessità e ragionevolezza e con adeguate misure di garanzia per la tutela dei dati personali.

In questo quadro, mi pare ragionevole la proposta formulata da Riccardo Innocenti, ex presidente dell'USCI: “Il Sistema Statistico Nazionale [venga] investito dei compiti di analisi statistica dell'evoluzione pandemica in corso, indicando gli estremi del segreto statistico, con un provvedimento *ad hoc* che sospenda il potere interdittivo preliminare del *Garante della privacy*, al quale viene lasciata la possibilità di verificare eventuali violazioni della riservatezza individuale verificatesi in concreto, e non solo ipotizzabili in astratto, comunque procedibili solo su esposto dell'interessato (per evitare liti temerarie)”<sup>21</sup>.

In questo modo, si potrebbe restituire alle persone la libertà di valutare – sulla base delle loro preferenze – il bilanciamento tra costi e benefici della comunicazione di alcuni dei propri dati personali<sup>22</sup>, invece di affidarsi al parere preventivo del Garante. In ogni caso – per

---

<sup>19</sup> Tra cui l'ISTAT, la Commissione per la garanzia della qualità dell'informazione statistica (COGIS), la Società italiana di statistica (SIS), l'Unione statistica dei Comuni italiani (USCI).

<sup>20</sup> Gli interventi sono pubblicati sul sito [www.garanteprivacy.it](http://www.garanteprivacy.it).

<sup>21</sup> [www.facebook.com/story.php?story\\_fbid=10224868964579243&id=1468374656](https://www.facebook.com/story.php?story_fbid=10224868964579243&id=1468374656)

<sup>22</sup> Molti hanno osservato – con riferimento al *cashback* di Stato – che chi aderisce all'iniziativa valuta il costo della cessione dei propri dati personali inferiore al *bonus* massimo di 150 euro.

effetto del GDPR, fondato sul principio della *privacy by default*<sup>23</sup> – la scelta avverrebbe in regime di *opting-in* (come quello in essere per la ‘cronologia delle posizioni’ di Google). L’emergenza sanitaria, peraltro, potrebbe motivare anche la rimozione di questo vincolo, consentendo un regime di *opting-out*.<sup>24</sup> Personalmente, sono favorevole a questa seconda ipotesi.

### **3. Che cosa si potrebbe fare se questi ostacoli fossero rimossi: un esempio**

Per illustrare il contributo che la statistica e i modelli matematici potrebbero dare per informare l’opinione pubblica e per fornire alle autorità suggerimenti per politiche di contrasto alla pandemia più efficaci ed eque faccio riferimento a un articolo pubblicato *online* su *Nature* il 10 novembre 2020: è un esempio rilevante di come si possano combinare dati della statistica ufficiale e dati dei telefoni mobili (Chang et al., 2020).

Gli autori partono dall’ipotesi che la pandemia COVID-19 abbia cambiato le abitudini di mobilità e che dunque siano necessari modelli che catturino gli effetti di questi cambiamenti sulla diffusione del virus. Hanno pertanto costruito un “classico” modello SEIR,<sup>25</sup> integrando reti di mobilità dinamiche a grana fine per simulare la diffusione del virus in dieci aree statistiche metropolitane statunitensi. Utilizzando i movimenti di 98 milioni di persone (suddivisi per ora e derivati dai dati dei telefoni cellulari), il modello traccia gli spostamenti dai luoghi di residenza (57 mila gruppi di sezioni di censimento, forniti dal *Bureau*

---

<sup>23</sup> “Il titolare del trattamento mette in atto misure tecniche e organizzative adeguate per garantire che siano trattati, per impostazione predefinita, solo i dati personali necessari per ogni specifica finalità del trattamento.” (art. 25, c. 2, Regolamento UE 2016/679).

<sup>24</sup> Per la differenza tra i regimi di *opting-in* e *opting-out* il riferimento classico è a Coase (1960), secondo il quale entrambi i regimi conducono allo stesso risultato. Kahneman et al. (1991) sostengono invece che la distorsione dello *status quo* e il modo in cui la scelta è presentata inducono a scegliere l’opzione predefinita (*default*), anche se la scelta diversa è relativamente facile. In linea con questa tesi, l’impostazione o la modifica delle scelte predefinite è stata proposta (Thaler e Sunstein, 2008) come modo efficace per influenzare il comportamento, ad esempio per decidere se diventare un donatore di organi.

<sup>25</sup> Nel modello SEIR (susceptible-exposed-infectious-recovered, cioè suscettibile-esposto-contagioso-guarito) si ipotizza un periodo di incubazione durante il quale gli individui sono stati infettati ma non sono ancora contagiosi. Durante questo periodo l’individuo si trova nel compartimento E (esposti). Il compartimento R comprende chi ha smesso di essere contagioso, quale che sia l’esito dell’infezione (guarigione o morte).

of Census) ai “punti di interesse” (dati acquisiti dai *dataset* di SafeGraph per 553 mila POI), creando in tal modo 5,4 miliardi di segmenti riferibili ai singoli spostamenti.

I risultati sono estremamente rilevanti: un modello relativamente semplice – una volta alimentato da una mole di dati vasta e granulare – si adatta molto bene alla traiettoria reale dei casi, nonostante i cambiamenti nel comportamento della popolazione nel tempo. Sulla base del modello, si rileva come una piccola minoranza di POI ‘superdiffusori’ sia responsabile della maggior parte delle infezioni. Di conseguenza, interventi mirati volti a limitare l’affollamento e l’occupazione massima in ogni POI risultano più efficaci di misure intese a ridurre uniformemente la mobilità.

Inoltre, il modello mette in luce come gli appartenenti a gruppi svantaggiati sotto il profilo razziale o per *status* socioeconomico siano a rischio di tassi di infezione più elevati, esclusivamente per effetto delle differenze di mobilità: queste persone hanno ridotto la loro mobilità in misura inferiore rispetto alla media, e i POI che visitano, relativamente più affollati, sono a rischio più elevato.

Individuando chi è infetto e mettendolo in relazione ai luoghi frequentati, il modello rende possibili analisi dettagliate che possono fornire risposte politiche più efficaci ed eque: ad esempio, gli autori sono in grado di calcolare gli effetti della riapertura di diverse tipologie di POI<sup>26</sup> rispetto al mantenimento della chiusura, con valori che spaziano dai 10 mila casi aggiuntivi per centomila residenti per i ristoranti in senso stretto, agli oltre mille per palestre, bar, hotel, tavole calde e centri religiosi, a valori inferiori a cento per la maggior parte dei negozi, inclusi supermercati e grandi magazzini.

## Riferimenti bibliografici

ALLEVA G., ARBIA G., FALORSI P.D., PELLEGRINI G., ZULIANI, A. (2020), “Proposta di una indagine a campione per una stima affidabile dei parametri fondamentali della epidemia da SARS-COV-2”, 26 marzo 2020. Disponibile *online*: [www.eticapa.it/eticapa/proposta-a-cura-di-statistici-di-alto-livello-su-unindagine-campionaria-accurata-sul-covid-19/?cn-reloaded=1#more-11040](http://www.eticapa.it/eticapa/proposta-a-cura-di-statistici-di-alto-livello-su-unindagine-campionaria-accurata-sul-covid-19/?cn-reloaded=1#more-11040)

---

<sup>26</sup> Purtroppo tra i POI non sono incluse le scuole e il modello utilizzato non prende in considerazione l’uso del trasporto pubblico, se non implicitamente.

- BIGGERI L. (2020), “Caos numeri COVID/ Biggeri: mancano criteri su tutto, perché non si vuole usare l’ISTAT?”, Intervista rilasciata a [ilsussidiario.net](http://ilsussidiario.net) il 17 novembre 2020. Disponibile online: [www.ilsussidiario.net](http://www.ilsussidiario.net).
- CHANG, S. et al. (2020). “Mobility network models of COVID-19 explain inequities and inform reopening”. *Nature* (2020). <https://doi.org/10.1038/s41586-020-2923-3>.
- COASE R. (1960), “The Problem of Social Cost”, *Journal of Law and Economics*, Vol. 3 (Oct. 1960): 1-44.
- GOOGLE (2020a), *COVID-19 - Rapporti sugli spostamenti della comunità*. Disponibile online: [www.google.com/covid19/mobility/index.html?hl=it](http://www.google.com/covid19/mobility/index.html?hl=it).
- GOOGLE (2020b), *COVID-19 - Rapporti sugli spostamenti della comunità. Lazio 1 dicembre 2020*. Disponibile online: [2020-12-01 IT Lazio Mobility Report it.pdf](https://www.google.com/covid19/mobility/index.html?hl=it).
- ISTAT (2015), *La nuova geografia dei sistemi locali*, ISTAT, Roma. Disponibile online: [La-nuova-geografia-dei-sistemi-locali.pdf](http://www.istat.it/it/files/2015/05/La_nuova_geografia_dei_sistemi_locali.pdf).
- KAHNEMAN, D., KNETSCH, J. L., THALER, R. H. (1991). “Anomalies. The Endowment Effect, Loss Aversion, and Status Quo Bias”. *Journal of Economic Perspectives*, Vol. 5, No. 1 (Winter 1991): 193-206.
- MONDUCCI R. (2019), “Il patrimonio informativo del sistema integrato dei registri statistici”, *L’intelligenza artificiale al servizio della democrazia*, Senato della Repubblica, 15 ottobre 2019. Disponibile online: [www2.slideshare.net](http://www2.slideshare.net).
- RKI (2020), *Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19)*, 20 ottobre 2020. Disponibile online: [www.rki.de](http://www.rki.de).
- SNOW J. (1855), *On the Mode of Communication of Cholera*, John Churchill, Londra. Il testo è disponibile online.
- SUNDGREN B. (1999). *Information systems architecture for national and international statistical offices. Guidelines and recommendations*, Conference of European Statisticians Statistical Standards and Studies No. 51, United Nations, Geneva.
- THALER, R. H., SUNSTEIN, C. R. (2008), *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Yale University Press, New Haven (CT).
- WALLGREN A. e WALLGREN B. (2014), *Register-based Statistics*, John Wiley & Sons, Hoboken (NJ).
- WHITBY A. (2020), *The Sum of the People*, Basic Books, New York (NY).