

# Monitorare e comprendere la pandemia nel nostro Paese: un bilancio da un punto di vista statistico e alcune proposte

Giornate della ricerca del Dipartimento MEMOTEF

Roma, 31 maggio 2022

Tavola rotonda con il gruppo di ricerca attivo da marzo 2020

Giorgio Alleva, Giuseppe Arbia, Piero Falorsi, Vincenzo Nardelli, Alberto Zuliani,  
integrato successivamente con Elena Ambrosetti e Alessandra De Rose

# Premessa

- Fin dall'inizio della pandemia con un gruppo di ricercatori - statistici e statistici economici - abbiamo condiviso dapprima la sorpresa, poi la delusione e l'amarezza per la carente costruzione e gestione del quadro informativo necessario per assumere le decisioni del governo, ai vari livelli, e per offrire ai cittadini e agli operatori economici e sociali una comunicazione corretta e utile sulle decisioni prese e, più in generale, sull'evoluzione della situazione.
- Non siamo rimasti passivi; abbiamo illustrato sui media e attraverso webinar il nostro disagio e le nostre esigenze informative, come statistici e come cittadini; abbiamo elaborato proposte che abbiamo condiviso con la comunità scientifica e con diverse istituzioni di ricerca; con scarso successo.
- Vogliamo condividere il bilancio, a nostro avviso quello di un'occasione mancata, per la statistica e la statistica ufficiale; raccogliere osservazioni sulle nostre proposte e rispondere alle vostre domande.

# Sommario

- Il rito quotidiano della diffusione dei dati
- Una variabilità non sempre decifrabile
- I problemi principali e la mancanza di un quadro di coerenza dell'informazione
- Gli aspetti oggetto di osservazione: considerazioni critiche
- I dati necessari per capire, intervenire, suscitare comportamenti
- Un sistema di monitoraggio multi-fonte
- Bilancio di un'occasione mancata
- Un esempio di buona pratica statistica: l'ONS nel Regno Unito
- Le nostre proposte
- Conclusioni

# Il rito quotidiano della diffusione dei dati

Ancora oggi, a distanza di oltre due anni dall'insorgenza della pandemia, vengono resi disponibili giornalmente sempre gli stessi dati che riguardano:

- test processati,
- contagiati e tasso di positività al test,
- stock dei ricoverati con il dettaglio per quelli in terapia intensiva e, per questa, gli ingressi nella giornata,
- morti,
- qualche indice derivato (contagi per 100.000 residenti),
- indice di trasmissione dei contagi  $R_t$  (con cadenza settimanale).

In generale i dati sono dettagliati per regione; in qualche caso, vengono forniti per provincia e grandi città.

L'ISS riporta età e genere. Settimanalmente, aggiorna un bollettino rimasto sostanzialmente immutato nel corso del tempo.

# Una variabilità non sempre decifrabile

Molti fattori possono determinare la **variabilità osservata** nel tempo e nello spazio:

- Specificità della **sindrome e mutazioni** del virus.
- Caratteristiche della **popolazione**, specialmente composizione per **età**.
- Presenza di **comorbidità**.
- Stato dei **sistemi sanitari**.
- Tipologia e intensità delle **misure di contenimento** introdotte.
- Organizzazione dei **sistemi di raccolta** delle informazioni.
- Eventuale disomogeneità del **quadro definitorio**.
- Da ultimo, l'andamento della **campagna vaccinale** e l'adesione delle persone al ciclo proposto.

# I problemi principali/1

I problemi principali sul piano dell'informazione, in particolare statistica:

- **manca** di alcuni dati fondamentali,
- **quadro definitivo non nitido**, né per la fase di raccolta dei dati, né per la loro comprensione (i metadati pubblicati sono insufficienti),
- organizzazione della raccolta inadeguata con ritardi nella disponibilità,
- **assenza di collegamento fra i dati** disponibili (difficile comprendere le relazioni),
- **scarsa iniziativa per approfondire** problemi e cogliere opportunità,
- **limitata produzione di informazione adeguata** a: sostenere le decisioni dei **governi** ai diversi livelli; convincere i **cittadini** ad assumere comportamenti adeguati e le **imprese** ad attivare misure di contrasto nei luoghi di lavoro,
- infine, **scarsa "sensibilità statistica"** nella raccolta, nell'elaborazione e nella comunicazione dei dati.

## I problemi principali/2: un approfondimento

- **Assenza di collegamento fra i dati**, cosicché non è possibile produrre stime delle probabilità di passaggio fra gli stati delle persone che si imbattono nella patologia: da asintomatico a eventualmente sintomatico, al trattamento domiciliare, al ricovero in reparti Covid, in terapia 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.
- **Assenza di un'informazione di carattere longitudinale** per consentire, fra l'altro, l'osservazione sul long-covid e sugli effetti della pandemia sulla popolazione non contagiata.
- **Mancata disponibilità** dei dati **in formati aperti e/o interoperabili** che ha impedito in generale alla comunità scientifica di contribuire alla comprensione della pandemia.
- **Mancato utilizzo dei big data.**

## I problemi principali/3: mancanza di un quadro di coerenza

- Sulla spinta dell'emergenza di marzo-aprile 2020, i dati sono stati **raccolti senza un disegno orientato a comprendere** nitidamente l'evoluzione dell'epidemia.
- Si sarebbe dovuto capire che i dati raccolti **non sarebbero stati sufficienti** a fornire un quadro informativo perspicuo.
- Ci si sarebbe dovuto chiedere quali **dati** fossero **necessari**.
- Si sarebbe dovuto comporli in un **quadro di coerenza generale** opportunamente progettato.
- Invece, **la base informativa è rimasta sostanzialmente immutata** e ha mostrato la sua **fragilità** anche nelle ondate successive.
- La **finalizzazione** dell'informazione è stata **congiunturale e non strategica**.

## Gli aspetti oggetto di osservazione: i test

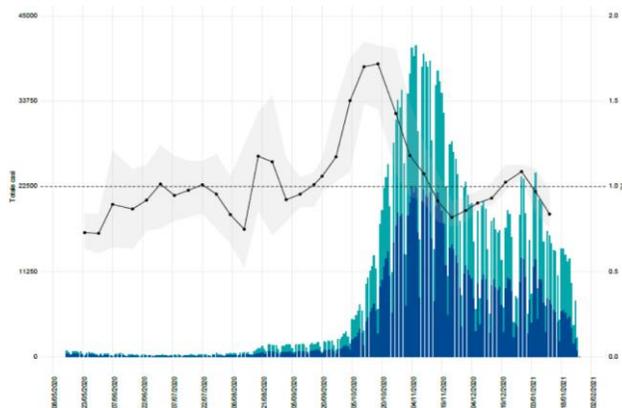
- Fin **dall'inizio** si è trattato di **test molecolari** attraverso tampone oro-naso-faringeo; da **metà gennaio 2021** si sono aggiunti i **test rapidi**, da qualche mese anche i **self test**. L'inclusione dei test rapidi e dei self test altera la serie storica e crea **confusione** nella lettura delle dinamiche temporali e spaziali.
- Il **numero di tamponi processati per 100.000 abitanti** è risultato molto **variabile nel tempo e fra territori**.
- A marzo-aprile 2020 i test erano riservati essenzialmente ai sintomatici; attualmente si distribuiscono fra **sintomatici, contatti, screening**. Il loro **mix** è **variabile** nel tempo e fra regioni.

## Gli aspetti oggetto di osservazione: i contagi

- Il **numero dei contagi** accertati è **strettamente collegato al numero di test** processati (non soltanto alla diffusione dell'infezione).
- I **test** effettuati e processati **sui contatti e per screening** sono stati via via più numerosi nel corso del tempo; **hanno fatto emergere gli asintomatici e i pauci-sintomatici** condizionando la lettura dell'evoluzione della pandemia.
- La categoria dei “**guariti clinicamente**” che viene considerata separatamente è riconducibili presumibilmente a **re-testing**.
- I casi di **re-testing** e anche quelli emersi da **screening andrebbero scorporati** ai fini del calcolo del tasso di positività, sia dal numeratore (test positivi) sia dal denominatore (test eseguiti).

# Gli aspetti oggetto di osservazione: l'indice di trasmissione del contagio $R_t$

- **Presentato** dai media **come elemento fondamentale di preoccupazione o di conforto** e soprattutto come elemento di prova dell'efficacia degli interventi operati nelle settimane precedenti o come indizio per quelli da assumere successivamente.
- Calcolato **sui soli sintomatici dei quali si conosca la data di comparsa dei sintomi**.
- Ambiguamente accolto come stima puntuale e **trattato come deterministico**, mentre sconta un errore statistico abbastanza ampio.
- Nel corso del tempo sono stati forniti un **intervallo di credibilità al 95%** e poi un **range** pure al 95% la cui **ampiezza è risultata molto variabile**, indizio preoccupante per l'**affidabilità della stima**.



Stima dell'indice  $R_t$  medio a 14 giorni e numero di casi per data di prelievo/diagnosi distinti per presenza o assenza della data di inizio dei sintomi

## Gli aspetti oggetto di osservazione: i ricoveri

- I dati sui ricoveri **sono i meno ambigui**.
- Rappresentano il **riferimento principale** per valutare la pressione sul sistema sanitario e quindi calibrare le restrizioni da adottare a livello regionale per contrastare il diffondersi dell'epidemia.
- Durante la campagna vaccinale sarebbe stato importante dare risalto appropriato a stock, flussi e tassi di ricovero **secondo la fase del ciclo vaccinale** in corso al momento del ricovero (1°, 2° e buster), distintamente per reparti Covid e terapia intensiva.

# Gli aspetti oggetto di osservazione: i decessi

- Per i **morti** registrati 'da Covid-19' si dispone di **informazioni** di sintesi **abbastanza dettagliate**: luogo, genere ed età.
- Il **numero di morti per Covid-19** è stato **sottostimato** nella prima fase della pandemia e forse **sovrastimato** successivamente.
- Il **bilancio finale** dei morti per la pandemia dovrà **tenere conto** in qualche misura anche dell'**eccesso di mortalità** generale rispetto agli anni precedenti.
- Si possono calcolare **tassi di letalità** (morti per 100 contagiati) e **tassi di mortalità** (morti per 100.000 abitanti).
- Tuttavia, per il primo tasso, si dispone dei **solli contagiati accertati**, con una presenza (assoluta e relativa) di asintomatici e *pauci*-sintomatici variabile nel tempo e territorialmente e ciò ne limita l'interpretazione nel tempo e nello spazio.
- La **letalità** è **correlata positivamente** con l'**età dei contagiati e soprattutto con la presenza di altre patologie**. Dati e studi sulla presenza di co-morbilità nei casi di decessi registrati come Covid sono quindi importanti.

# I dati necessari per capire, intervenire, suscitare comportamenti/1

Nella fase di insorgenza della pandemia le **domande principali** che i cittadini si ponevano e alle quali occorreva dare risposta erano:

1. **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, eventualmente per terapie intensive;
- se entrati in ospedale, di uscirne guariti;

2. in **quali luoghi e contesti c'è rischio maggiore di contagio** (si sarebbe potuto sfruttare i big data);

3. infine, **sulla base di quali informazioni vengono prese le decisioni** circa le misure restrittive da adottare.

# I dati necessari per capire, intervenire, suscitare comportamenti/2

## Successivamente sono sorte nuove domande:

1. quali sono gli **effetti della campagna vaccinale** sulla probabilità di contagio, di tipologia di ricovero, di guarigione e di decesso. Gli effetti della campagna vaccinale sui ricoveri e i decessi sono ben documentati dall'ISS, non altrettanto sui contagi.
2. qual è l'**impatto di medio e lungo periodo del contagio** sulla salute e la vita delle persone che si sono ammalate.
3. qual è l'**impatto di medio e lungo periodo della pandemia** sulla vita delle persone che non sono state contagiate.
4. quali **effetti** si sono prodotti **per popolazioni selezionate**: studenti dei vari ordini di scuola; studenti universitari; persone alla ricerca di occupazione; anziani autosufficienti eccetera

Dati inutili: Immuni/1



Dati aggiornati al 21 maggio 2022

**Totale casi da inizio pandemia: 17.229.263 (Protezione civile)**

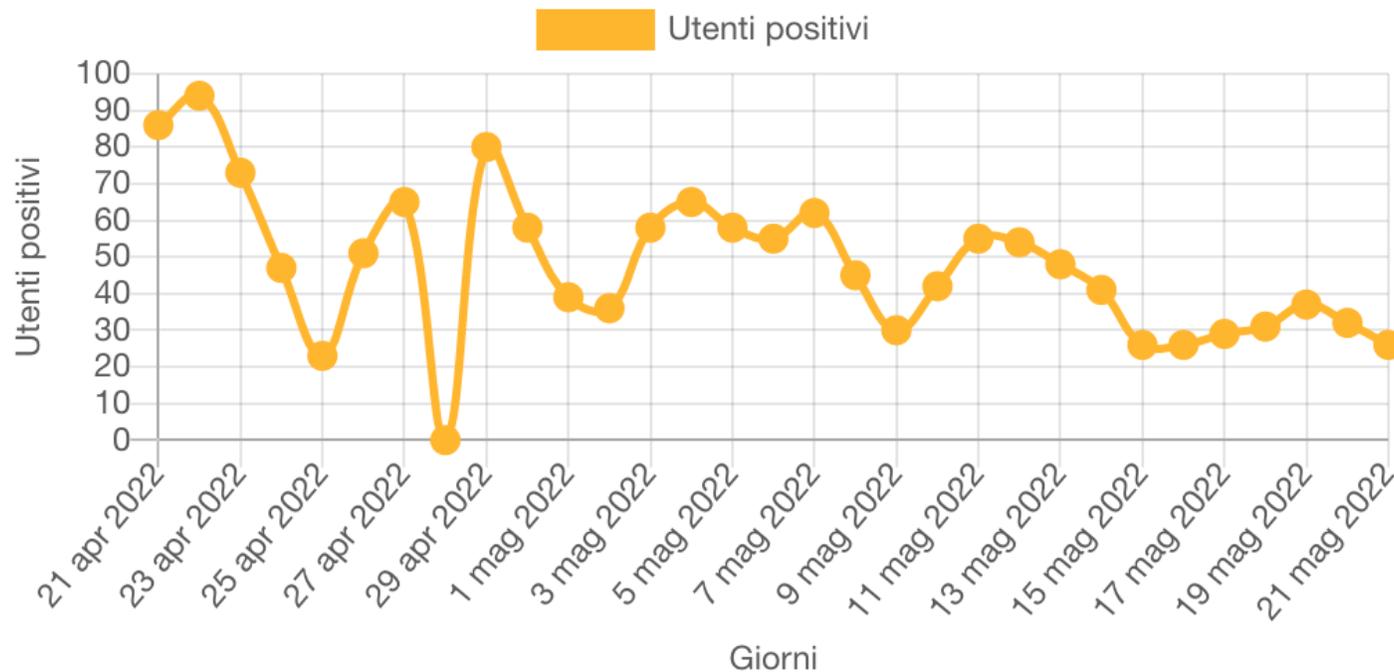
Download (1): Utenti che hanno scaricato l'applicazione (non sono compresi gli aggiornamenti e le reinstallazioni).

Utenti positivi (2): Utenti positivi che hanno caricato le loro chiavi.

Notifiche inviate (3): Notifiche di possibile esposizione al rischio generate dall'applicazione. La rilevazione è parziale poiché vengono rilevate tutte le notifiche per i dispositivi iOS e solamente un terzo di quelle inviate da Android.

## Dati inutili: Immuni/2

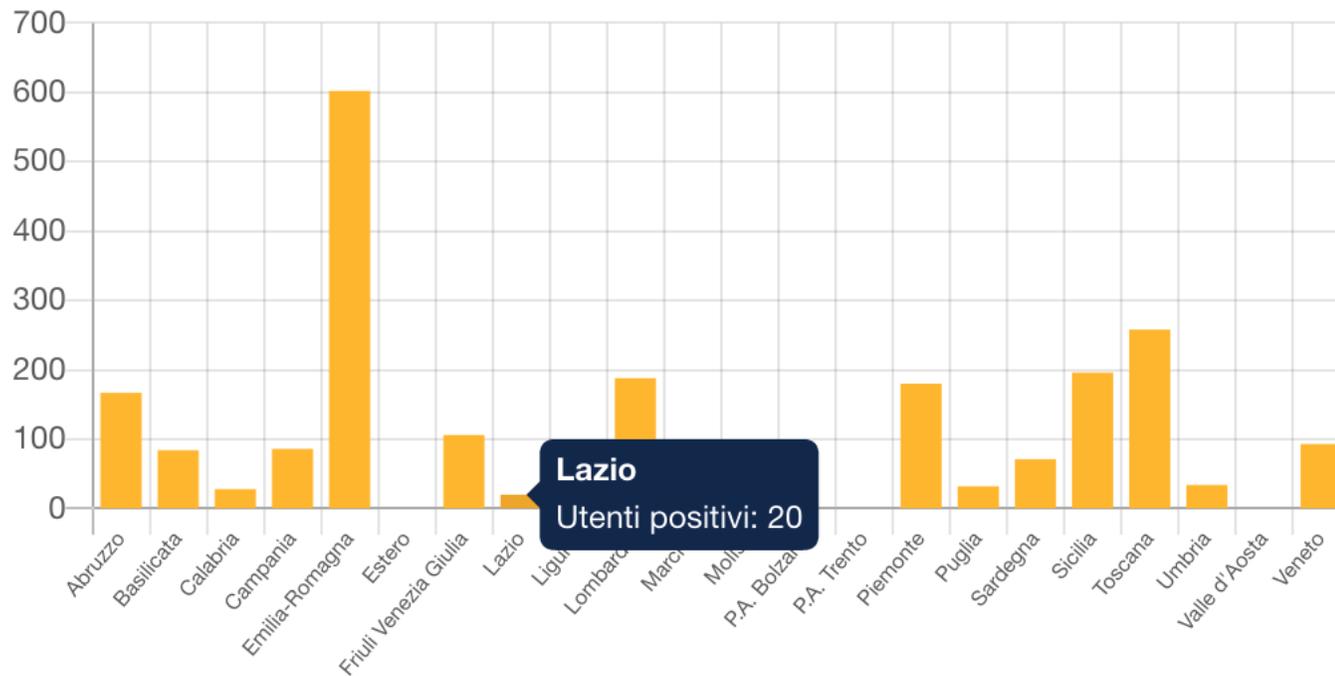
### Andamento degli utenti positivi



- Ad esempio, il 22 aprile sono stati registrati 94 utenti positivi che corrispondono allo **0,01% del totale dei positivi** rilevati in quel giorno dalla Protezione civile (73.212).
- Serie storica **non correlata** con i conteggi ufficiali: nei weekend non è visibile la riduzione del conteggio dei positivi.

# Dati inutili: Immuni/3

## Utenti positivi per regione su base mensile



Dati riferiti al mese di Aprile 2022.

### BUONI DATI NELLA FORMA

Dati diffusi con buona **granularità** che includono anche i casi importati **dall'estero**, distribuiti quotidianamente sia in **open data** (formato csv) sia in una **dashboard** accessibile e ben organizzata.

### MA NON NELLA SOSTANZA

Purtroppo **non informativi** sulla condizione attuale della pandemia in Italia.

Dati aggiornati al 21 Maggio 2022

# Un sistema di monitoraggio multi-fonte/1

## 1. Integrare informazioni sanitarie, raccolte in modo rigoroso, con altre campionarie.

- Campione ripetuto nel corso del tempo per **stimare la prevalenza** e seguirne l'evoluzione.
- Campione panel per stimare i **cambiamenti di stato** dei contagiati.
- Campione **panel di scuole e alunni** per stimare il tasso di contagio nel sistema educativo e seguirne l'evoluzione.
- Campione **panel di medici di base** per stimare il numero di pazienti con quadri sintomatici riconducibili a Covid-19, il loro tasso di positività al test e rilevare dati riguardo al decorso successivo.
- Monitoraggio costante della situazione nelle **residenze sanitarie assistite**.

## 2. Individuare, attraverso sistemi di tracciamento, **luoghi e contesti nei quali la trasmissione del SARS-CoV-2 è più frequente**.

## 3. Utilizzare le **tecnologie dell'informazione** e i **big data** per stimare gli effetti delle restrizioni adottate.

## 4. Assicurare la copertura informativa del **piano vaccinale**, non limitata al solo numero di vaccinati ma orientata alla lettura dell'evoluzione dei contagi e delle conseguenze sul sistema sanitario.

## Un sistema di monitoraggio multi-fonte/2

- Valutazione delle **conseguenze di Covid-19 sulle condizioni di salute delle persone** dopo la guarigione (dati sanitari e campionari).
- L'**Invalsi** ha svolto meritoriamente una valutazione dell'effetto differenziale della didattica a distanza rispetto a quella in presenza sul rendimento scolastico degli alunni.
- Le **università**, in modo coordinato, dovrebbero misurare gli effetti sulle carriere studentesche e sul rendimento agli esami della didattica a distanza e in modalità mista.
- L'**Istat dovrebbe assumere un ruolo significativo** che si è espresso finora soltanto in due occasioni: l'indagine sierologica di maggio-luglio 2020 (con i forti limiti che ha avuto) e la valutazione della sovra-mortalità rispetto agli anni precedenti (con qualche *defaillance*). L'aggiunta di alcune domande nei questionari di rilevazione periodica è apparsa occasionale. Gioverebbe una **organizzazione sistematica di raccolta, elaborazione e diffusione di dati** relativamente a presenza e intensità dello *smart working*; adozione delle norme igieniche essenziali e rispetto delle misure di contenimento via via introdotte da parte delle persone, delle imprese e delle istituzioni. Un impegno specifico dovrebbe essere dedicato al long-covid e agli effetti della pandemia sulla popolazione generale mediante **indagini longitudinali**.
- Il sistema dovrebbe essere in grado di **attrezzarsi rapidamente** per produrre stime affidabili **su questioni rilevanti che emergessero** via via.

# Bilancio di un'occasione mancata/1

- Il «rito quotidiano» doveva essere un appuntamento di grande **chiarezza e autorevolezza**, fondato sulla solidità dei processi di raccolta e integrazione dei dati, ben documentato, frutto di una **collaborazione tra istituzioni e competenze** (forse sarebbe potuto essere a cadenza non giornaliera).
- Occorreva mettere a punto un **sistema di monitoraggio dinamico**, consapevoli dei limiti delle informazioni provenienti dalle sole strutture ospedaliere, della necessità di miglioramento nei processi di raccolta dei dati, della responsabilità di intervenire con rigore e autorevolezza nel dibattito sulle misure da adottare e, in generale, su come fronteggiare l'emergenza (forse un portavoce unico avrebbe giovato).
- L'impreparazione e **gli errori emersi** riguardo alla misura della dimensione e degli effetti della pandemia durante la prima ondata **non hanno indotto a migliorare** successivamente.
- Non si è investito in un **sistema informativo** derivante dai dati delle autorità sanitarie **integrato** con quelli raccolti attraverso indagini statistiche realizzate mediante campioni casuali).

## Bilancio di un'occasione mancata/2

- La **fiducia nelle evidenze numeriche**, nelle previsioni da modello ne escono rafforzate? I dati e le informazioni statistiche si sono tradotti nella conoscenza necessaria per valutare e decidere interventi e comportamenti? La risposta è negativa.
- La **trasparenza** non sempre è stata assunta come valore distintivo da parte delle istituzioni responsabili delle decisioni per il contrasto dell'epidemia.
- L'insufficienza e talvolta l'ambiguità dei dati hanno alimentato la **distanza delle posizioni** fra i tecnici e nel confronto politico e istituzionale.
- La **confusione** fra i cittadini è stata grande e talvolta è sembrato anche nei luoghi delle decisioni.
- Non è stata **contrastata la comunicazione senza informazione**. Si sarebbe dovuto fare chiarezza sulla distinzione fra evidenza scientifica e opinione (anche autorevole), fra competenza, genericità e sprovvedutezza.
- Alle **esortazioni** (ad esempio quelle contenute e ripetute pedissequamente nei bollettini settimanali dell'ISS) **non sono seguite le azioni** evocate.

## Bilancio di un'occasione mancata/3

- Non sono stati resi disponibili alla comunità scientifica i **dati individuali anonimizzati** (le resistenze nascondono un problema di qualità?).
- Sono state mantenute sostanzialmente immutate le **gelosie istituzionali e disciplinari**.
- L'accademia si è esercitata sui **modelli** di tendenza e previsione della diffusione del contagio ma molto poco sulla **qualità dei dati** sottostanti.
- La **comunità statistica** non è riuscita ad avere un ruolo significativo nei luoghi delle decisioni.

# Un esempio di buona pratica: l'Office for National Statistics (ONS) del Regno Unito

- **Coronavirus (COVID-19) Infection Survey, UK**
- **Sample survey from 18 April 2020, results every two weeks.**
- Estimates for England, Wales, Northern Ireland and Scotland, and for the main cities.
- This survey is being delivered in partnership with University of Oxford, University of Manchester, UK Health Security Agency and Wellcome Trust.
- This study is jointly led by the Office for National Statistics (ONS) and the Department for Health and Social Care (DHSC) working with the University of Oxford and Lighthouse laboratory to collect and test samples.

Un sistema informativo dinamico che ha sfruttato il mandato istituzionale, la competenza dell'Istituto nazionale di statistica e la storica indagine longitudinale [Understanding Society](#).

## Explore by topic

Comparisons

Infections

Hospitals

Deaths

Vaccines

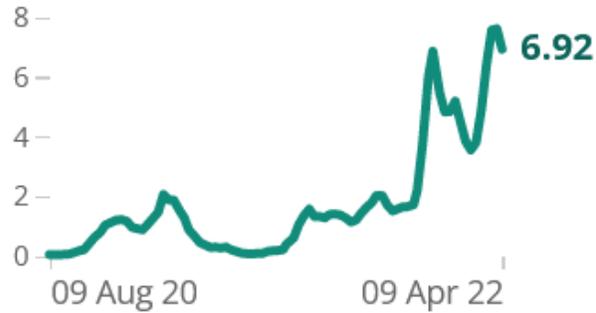
Antibodies

Well-being

Lifestyle

## Infections decreased

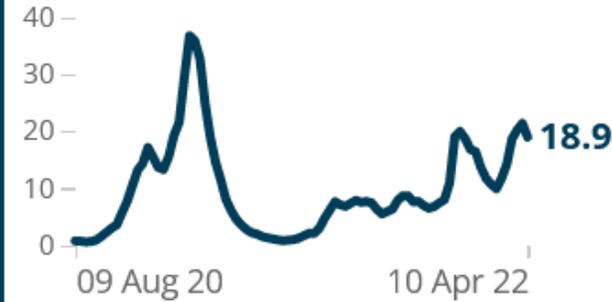
Percentage testing positive for COVID-19, England



Office for National Statistics

## Hospital admissions decreased

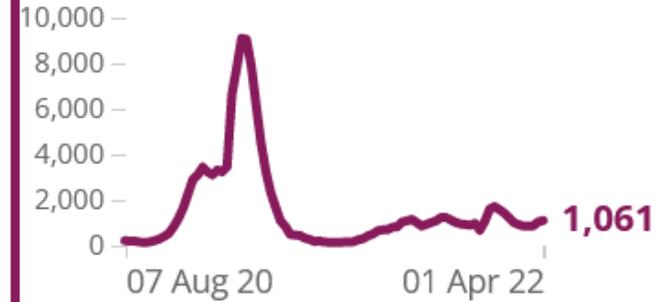
Hospital admissions involving COVID-19 per 100,000 people, England



UK Health Security Agency

## Deaths continued to increase

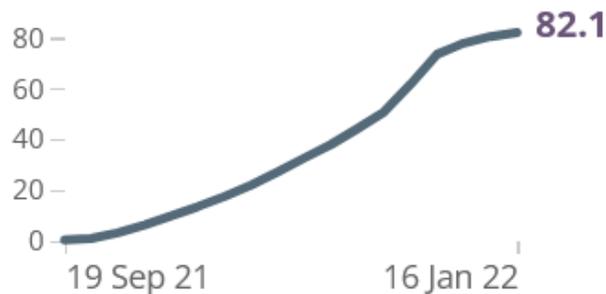
Deaths registered by week involving COVID-19, UK



Office for National Statistics

## Around 8 in 10 adults have had 3 vaccine doses

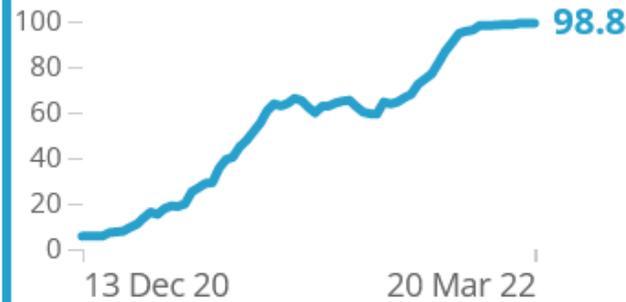
Percentage who have received three doses of a COVID-19 vaccine, England



Office for National Statistics

## Over 9 in 10 adults had antibodies

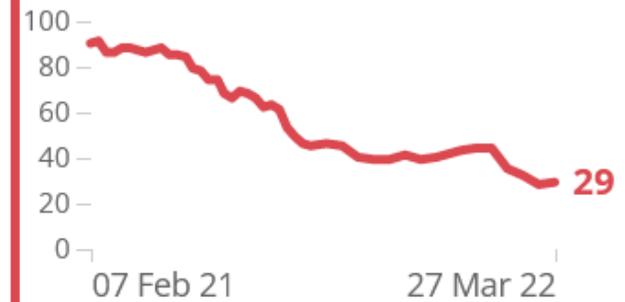
Percentage testing positive for antibodies to COVID-19, England



Office for National Statistics

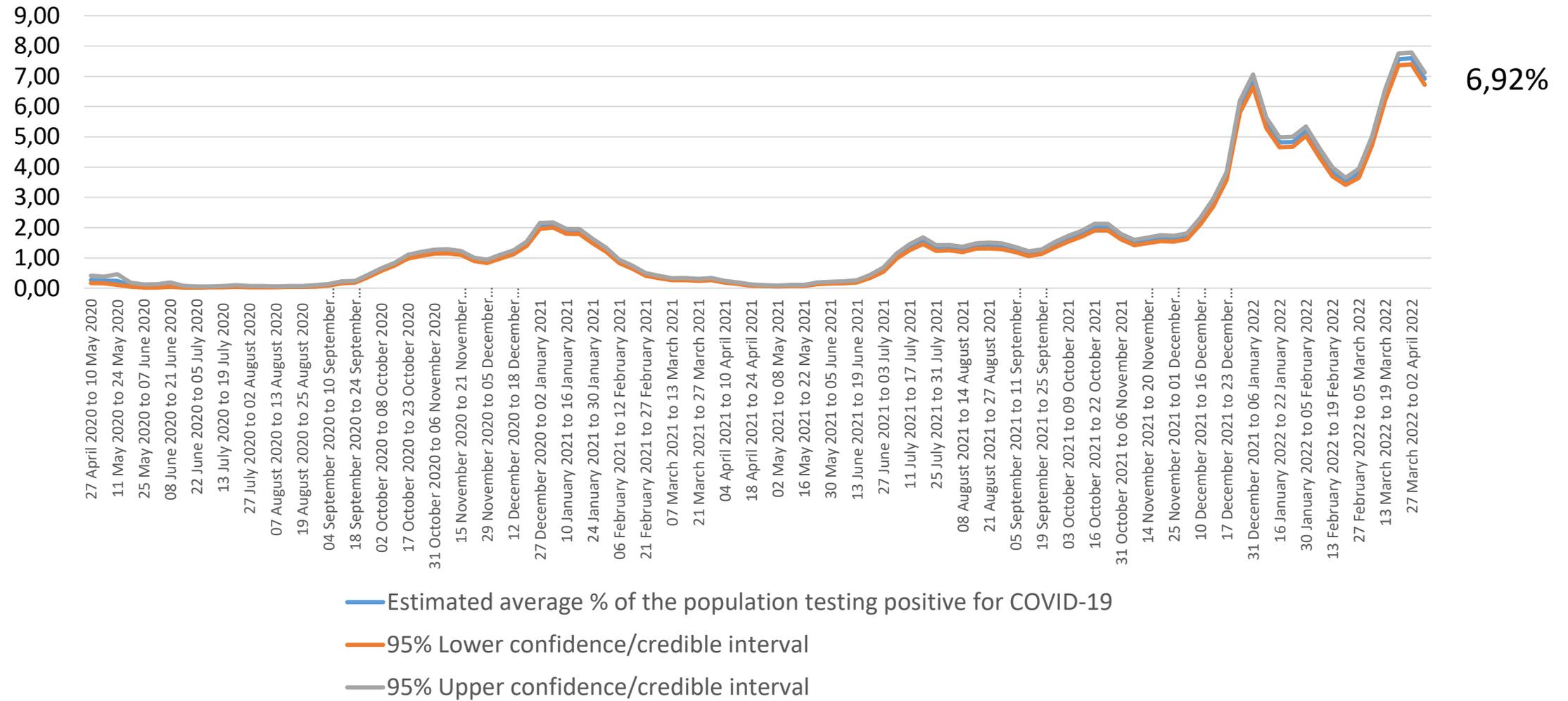
## Social distancing remained low

Percentage always/often maintaining social distancing



Office for National Statistics

## Official reported estimates of the percentage of the population testing positive for COVID-19, UK countries



# Coronavirus (COVID-19) latest insights

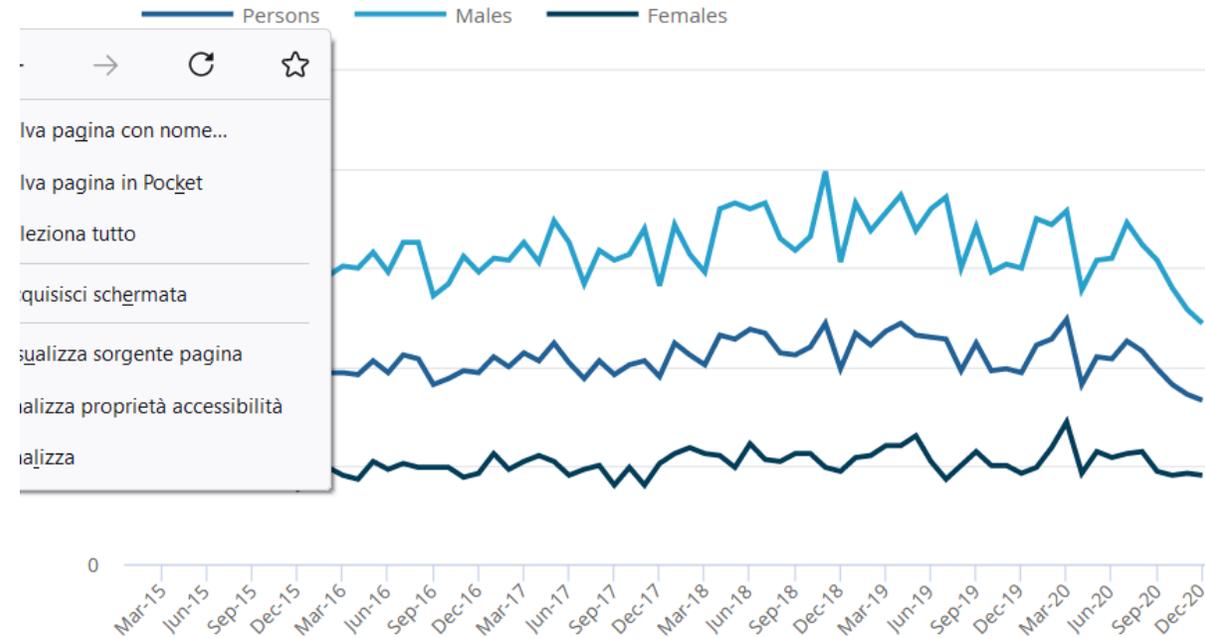
## 14 April 2022

### Latest COVID-19 headlines

- [Infections remain high in Wales but fall in other UK countries](#)
- [No evidence suicides increased during the COVID-19 pandemic](#)
- [Retail and recreation visits remain below pre-COVID-19 levels](#)

### Suicide rates in 2020 remained fairly consistent with previous years

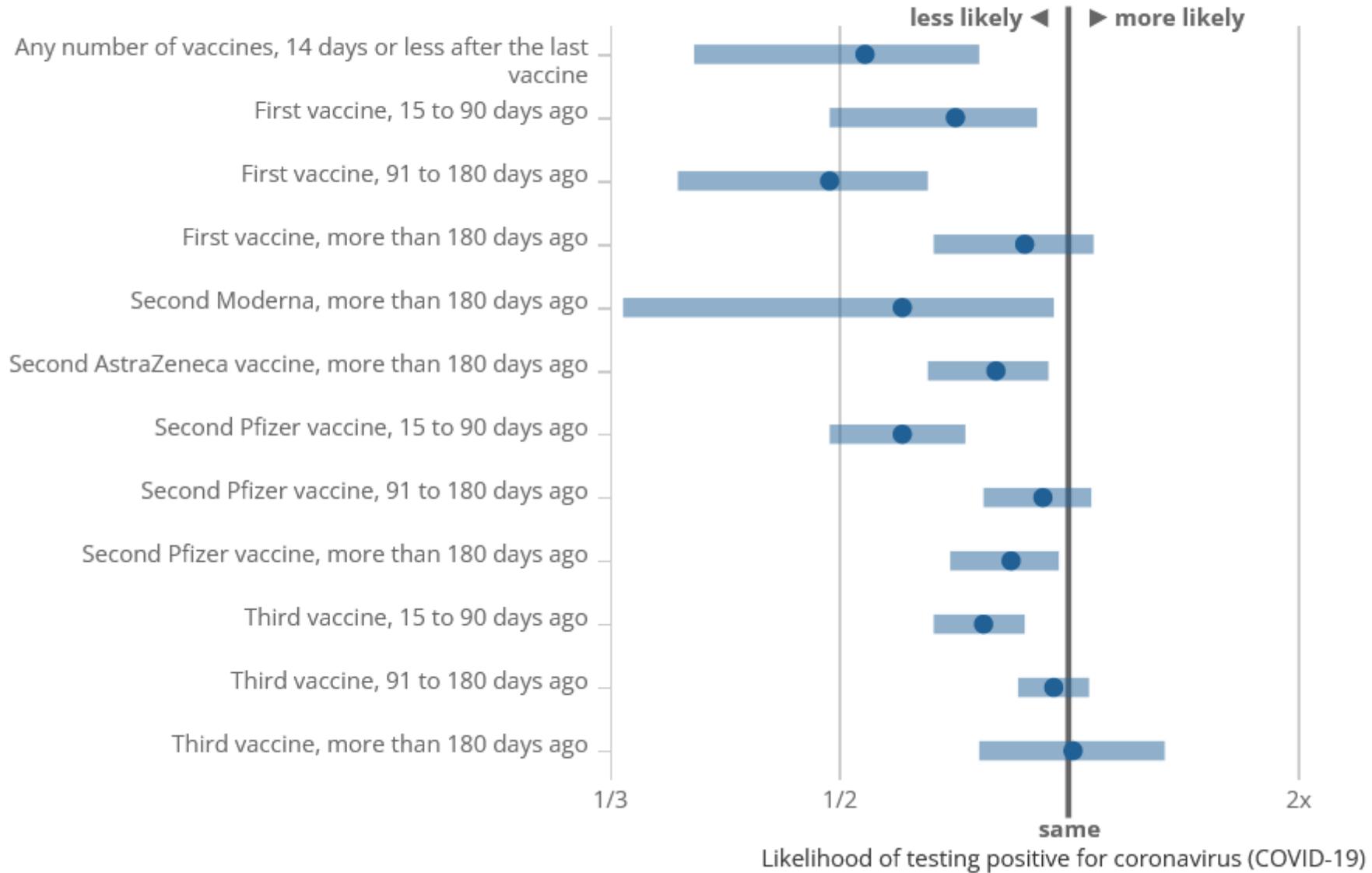
Monthly age-standardised suicide rates by sex, England and Wales, deaths occurring between January 2015 and December 2020



Source: Office for National Statistics – Deaths from suicide that occurred in England and Wales

● Estimates with 95% confidence intervals

**Vaccination status (Reference group: Not vaccinated)**



# Impact of coronavirus (COVID-19) on lifestyle

---

- [Travel](#)
- [Impact on society](#)
- [Household spending and disposable income](#)
- [Preventative measures](#)
- [Homeworking](#)
- [Impact on education](#)
- [Impact on disabled people](#)
- [Further information](#)

# Le nostre proposte

## Marzo 2020

**1.** Indagine campionaria ripetuta nel tempo per stimare **il tasso di contagio** riferito alla popolazione (passaggio da campione 'di convenienza' a campione probabilistico).

## Marzo 2022

**2.** Impianto di una indagine campionaria longitudinale sugli **effetti della pandemia** sulla vita delle persone, nel medio e lungo periodo.

## In corso

**3.** Studio delle caratteristiche di efficienza e fattibilità di un indagine standard a due stadi per la **stima degli infetti tenendo conto della autocorrelazione spaziale delle infezioni**.

Le idee progettuali sono state elaborate da un **gruppo di ricerca** che ha portato avanti, dall'inizio del 2020, studi relativi all'impianto di sistemi integrati di raccolta di dati sulla diffusione della pandemia Covid-19. Ne fanno parte: Giorgio Alleva, Giuseppe Arbia, Piero Falorsi, Vincenzo Nardelli, Alberto Zuliani. Per la seconda proposta il gruppo è stato integrato da Alessandra De Rose ed Elena Ambrosetti.

## Proposta di indagine campionaria/1 (Alleva *et al.*, 2022, JOS)

- Realizzazione di un'indagine campionaria ripetuta nel tempo per fare inferenza sul tasso di contagio riferito alla popolazione, proposta avanzata a marzo 2020, in piena prima ondata dell'epidemia, quando si conosceva ancora molto poco di Covid-19.
- Utilizzo di campioni probabilistici di contagiati e di loro contatti (Gruppo A) e di residenti (Gruppo B), quest'ultimo rappresentativo della popolazione, eventualmente a partire da una certa età, seguito nel tempo.
- Effettuazione del test diagnostico (non per i contagiati) per valutare la prevalenza e raccolta di informazioni utili a interpretare le dinamiche di contagiosità.
- Oltre a stimare la prevalenza puntuale, era quindi previsto di seguire longitudinalmente parte del campione relativo al gruppo B per valutare i passaggi di stato.
- Le stime ottenibili soddisfano le caratteristiche di non distorsione e presentano un'efficienza di gran lunga superiore a quella ottenibile mediante un campionamento casuale semplice.

## Proposta di indagine campionaria/2

- **Gruppo A** è il sottogruppo costituito dagli individui per i quali è stato **accertato uno stato di infezione** (potrebbero essere ricoverati in ospedale o in quarantena obbligatoria) e da tutte le persone che hanno avuto contatti con loro negli X giorni precedenti l'accertamento del contagio (conosciuti attraverso i dati raccolti dalle autorità sanitarie).

Concentrare gli investimenti in risorse sull'osservazione dei contatti di questo gruppo massimizza il numero di persone infette osservate nel campione..

- **Gruppo B** costituito da persone sane, persone infette ma non consapevoli di esserlo, persone la cui infezione è ancora in una fase di incubazione e tutte quelle che hanno avuto contatti con loro.

È necessario osservare il Gruppo B per produrre stime affidabili per l'intera popolazione relativamente al tasso di persone infette e al tasso di mortalità.

- Inoltre si può ottenere un'informazione sui **passaggi di stato**.

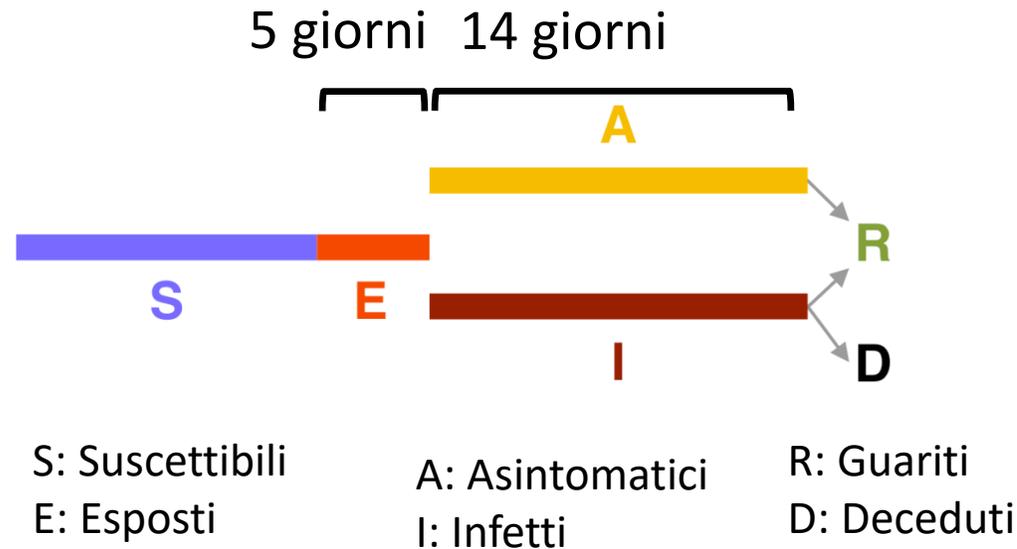
## Proposta di indagine campionaria/3

- 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.
- **Allora non se ne fece nulla.**
- Una giustificazione potrebbe essere rinvenuta nella situazione di emergenza sanitaria che si stava vivendo, che suggeriva di destinare i tamponi ai soli sintomatici.
- Nel corso della seconda ondata si processavano circa 200.000 tamponi ogni giorno **e destinarne il 5% una volta ogni due settimane** alla stima della prevalenza sarebbe stato del tutto fattibile. A maggior ragione nelle ondate successive.
- Complessivamente si è trattato di un'**occasione perduta** da parte della statistica e della statistica ufficiale per contribuire alla definizione di un quadro evolutivo affidabile della pandemia.
- Di fronte alla grande incertezza che ha generato il Covid-19 la **statistica**, che di incertezza si nutre avrebbe dovuto svolgere un **ruolo più importante**. Un'occasione perduta per la statistica ufficiale considerato il mandato e l'indipendenza dell'Istituto, le competenze e professionalità che possiede, la fiducia di cui gode da parte dei cittadini.

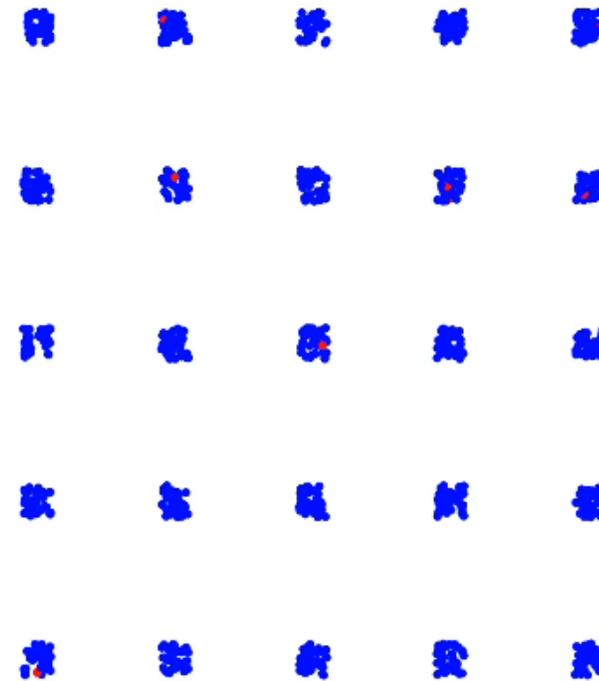
# Simulazione ①

Considerata l'impossibilità di ottenere **dati individuali anonimizzati** per finalità scientifiche, i metodi di campionamento sono testati su un **dataset generato** in cui vengono tracciati 20.000 individui.

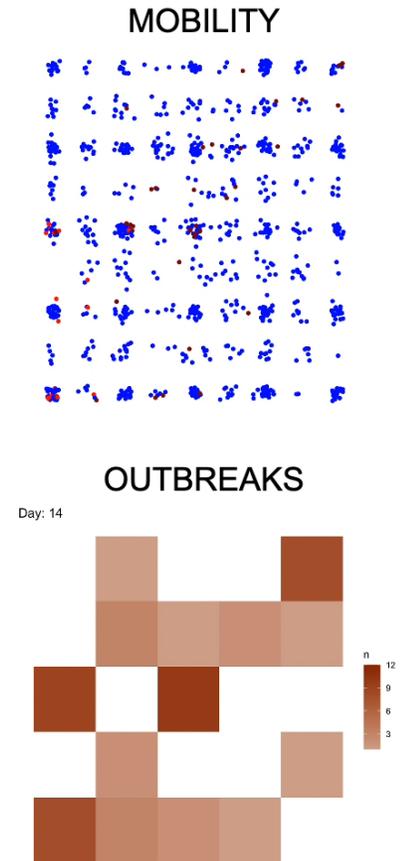
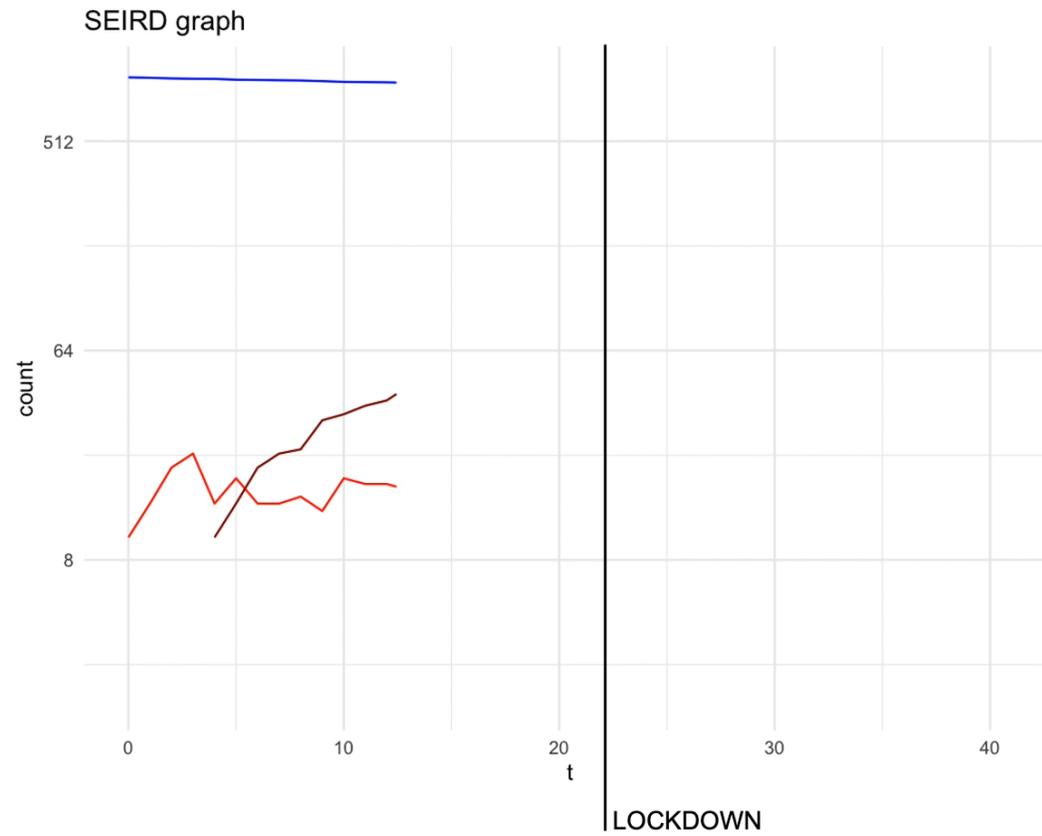
## Simulazione epidemia



## Simulazione mobilità ed interazioni sociali



# Simulazione ①



## Risultati della simulazione: Distorsione ed Efficienza ①

<b>Giorno</b>	<b>Percentuale individui infetti nella popolazione</b>	<b>Distorsione relativa</b>	<b>Numerosità campione tradizionale equivalente a 1000 testati con tracciamento</b>
<b>15</b>	0.58%	0.0019	72
<b>25</b>	3.94%	0.0000	161
<b>35</b>	6.54%	0.0045	190

## Optimal two stages spatial sampling design

Come sfruttare l'informazione spaziale della popolazione per migliorare il disegno di campionamento standard a due stadi, finalizzato alla stima dei contagi?

Coniugando **efficienza e fattibilità**, è proposto un **disegno campionario basato su tecniche di bilanciamento** mostrando analiticamente la performance rispetto ad altri piani di campionamento e studiando il comportamento attraverso simulazioni Monte Carlo.

## Efficienza e fattibilità: disegno campionario standard di campionamento standard a due stadi

<b>Giorno</b>	<b>Autocorrelazione spaziale</b>	<b>Componente geografica (I livello)</b>	<b>Componente individuale (II livello)</b>	<b>Componente non riducibile</b>	<b>Componente geografica DOPO applicazione del metodo</b>
<b>7</b>	0,59	<b>83,00%</b>	11,33%	5,60%	<b>10,72%</b>
<b>15</b>	0,63	<b>84,50%</b>	10,31%	5,10%	<b>8,56%</b>
<b>43</b>	0,96	<b>98,50%</b>	1,01%	0,50%	<b>2,91%</b>

Optimal two stages spatial sampling design for the estimation of the critical parameters of the SARS-CoV-2 epidemic: Efficiency vs. feasibility: submitted to The American Statistician

# Proposta di uno studio longitudinale sugli effetti del Covid-19 (in corso)

- A distanza di due anni dall'insorgenza della pandemia, **manca in Italia uno studio longitudinale** che consenta di valutare in modo sistematico gli effetti di Covid-19 sulla vita delle persone, elemento cruciale per comprenderne l'impatto sulla società e sostenere le scelte ai diversi livelli di governo e ambiti di responsabilità.
- Le **esperienze in corso in altri Paesi** hanno consentito e consentono di dare risposta a numerose questioni che altrimenti rimarrebbero sospese.
- È stata delineata una proposta di indagine longitudinale riferita a Covid-19, tratteggiando lo **schema di disegno campionario** e fornendo una **stima delle risorse finanziarie e organizzative necessarie**.
- È fondamentale prevedere la regia da parte delle **autorità sanitarie nazionali** e il coinvolgimento attivo delle **Regioni** e dell'**Istat**.

# Studi longitudinali/1

- L'indagine più significativa a **livello europeo** è la **SHARE** (*Survey of Health, Ageing and Retirement in Europe*), un'infrastruttura di ricerca per lo studio degli effetti delle politiche sanitarie, sociali, economiche e ambientali sul percorso di vita dei cittadini europei.
- È realizzata dal 2004 in 28 paesi europei e Israele ed è coordinata presso il *Max Planck Institute for Social Law and Social Policy*. Sono state realizzate otto waves d'indagine tra il 2004 e il 2020.
- Il campione italiano è di circa 2000 individui di 50 anni e più.
  
- A **livello nazionale** in miglior esempio di indagine longitudinale sul Covid-19 è quello del **Regno Unito**, nell'ambito della storica indagine longitudinale [Understanding Society](#) condotta dall'*Office for National Statistics* (ONS). Sono state realizzate numerose ondate di indagini specifiche su Covid-19, a partire dal mese di aprile 2020, per valutare l'impatto della pandemia sul benessere degli individui e delle famiglie.
- Pure nel Regno Unito sono state attuate modifiche e aggiornamenti ai questionari di altre indagini panel esistenti per monitorare gli effetti della pandemia sulla popolazione.

## Gli studi longitudinali/2

- L'indagine panel francese **ELIPSS** (Étude Longitudinale par Internet Pour les Sciences Sociales) è progettata e gestita da Sciences Po e CNRS. Rilevazione mensile partita nel 2012; si è giovata di un finanziamento dell'Agence Nationale de la Recherche (ANR). **I partecipanti sono dotati di un tablet e di un abbonamento a internet mobile** rispondendo da remoto, via web, in occasione delle diverse waves.
- E' realizzata ad hoc **per la comunità scientifica** in modo da colmare la carenza di mezzi per la realizzazione di indagini longitudinali sulle tematiche proprie alle scienze umane e sociali.
- Altri **paesi europei** hanno proposto **questionari ad hoc** nell'ambito di indagini longitudinali esistenti per studiare gli effetti di Covid-19 sulla salute fisica e mentale dei cittadini e valutarne l'impatto sociale ed economico.
- In **Italia** si segnala l'utilizzo dell'indagine longitudinale **Italian Lives** (ITA.LI) realizzata dal Dipartimento di Sociologia e ricerca sociale dell'Università di Milano-Bicocca nel corso del 2020. Un **campione di 950 intervistati** ha risposto, tra aprile e agosto 2020, a un breve questionario inteso a individuare i cambiamenti intervenuti per i principali aspetti di qualità di vita a causa di Covid-19.

# La proposta: qualche dettaglio/1

- La proposta è di suddividere la popolazione in tre sottopopolazioni:

A: **ricoverati finora** in strutture ospedaliere per Covid-19 e guariti o ancora in terapia ospedaliera;

B: **contagiati conosciuti finora** non ospedalizzati e guariti o ancora in terapia domiciliare;

C: **restante popolazione** (compresi i contagiati non conosciuti perché asintomatici o *pauci*-sintomatici, non diagnosticati o volutamente occultati).

- Potrebbero verificarsi transiti e sovrapposizioni fra le tre sottopopolazioni dopo l'inizio della rilevazione.
- Per la sottopopolazione A la soglia minima di età sarebbe indicata tenendo conto del consenso dei genitori dei rispondenti. L'idea sarebbe di estenderla alle persone con 10 anni o più attingendo le risposte dai soggetti adulti della famiglia, tipicamente la coppia genitoriale.
- Nella proposta è suggerita una possibile estensione dello studio ai nuclei famigliari.

## La proposta: qualche dettaglio/2

- Per la **sottopopolazione A** degli ospedalizzati sarebbe assai conveniente poter effettuare il campionamento a partire da **elenchi in possesso delle autorità sanitarie**. Collettività di questo primo gruppo sono in qualche caso già oggetto di analisi longitudinali locali da parte delle strutture ospedaliere e delle autorità sanitarie. Sarebbe interessante procedere a una ricognizione dell'esistente e all'integrazione delle informazioni eventualmente disponibili.
- Anche per la **sottopopolazione B** dei contagiati sarebbe importante poter accedere a elenchi disponibili (presumibilmente **elenchi della medicina territoriale**).
- La costruzione del panel relativo alla terza **sottopopolazione C** non pone problemi particolari.
- La periodicità delle ondate andrà definita tenendo conto dell'evoluzione della pandemia.
- Sono state formulate ipotesi operative per poter procedere alla stima dei costi.

# La proposta: la dimensione campionaria

- Si ritiene idonea una numerosità di 1.000 unità per ciascuna delle tre sottopopolazioni per ogni realtà geografica di interesse (ad esempio una regione, una città metropolitana).
- Sarebbero prodotte stime accettabili (con  $CV < 15$ ) nel caso di un'incidenza del carattere investigato nella popolazione superiore al 5%.

*Errori di campionamento (CV) secondo l'incidenza del carattere sulla popolazione e la numerosità campionaria*

Valore della proporzione espressa in termini percentuali	Numerosità campionaria			
	1.000	5.000	10.000	20.000
30%	4,83	2,16	1,53	1,08
20%	6,32	2,83	2,00	1,41
10%	9,49	4,24	3,00	2,12
5%	13,78	6,16	4,36	3,08
1%	31,46	14,07	9,95	7,04

- Il suggerimento per un'indagine nazionale che punti a produrre stime regionali è di operare su campioni di circa 1.000 unità per ognuna delle tre sottopopolazioni di interesse per ogni regione e di puntare, quindi, a un campione di 20.000 unità per ciascuna sottopopolazione, e dunque di complessive 60.000 unità.
- L'indagine dovrebbe essere promossa dal Ministero della salute, dall'Istituto superiore di sanità e dall'Istat, con l'auspicabile collaborazione delle regioni e dei comuni, e potrebbe giovare di un finanziamento del PNRR.

## La proposta: le variazioni nel corso del tempo da osservare attraverso i campioni relativi alle tre sottopopolazioni

- salute oggettiva e salute percepita;
- prevenzione e cura;
- impatto di Covid-19 sulla salute mentale;
- comportamenti assunti rispetto alla pandemia;
- comportamenti a rischio;
- stili di consumo;
- relazioni all'interno o all'esterno della famiglia;
- aiuti forniti e ricevuti;
- utilizzo e soddisfazione dei servizi pubblici e privati, con specifica attenzione a quelli sanitari;
- situazione di benessere oggettivo e percepito;
- condizione occupazionale;
- situazione abitativa;
- altre.

# Comunicazione dei dati

- Comunicazione giornaliera:
  - **statica**, mai critica, **talvolta fuorviante**; l'instabilità delle definizioni per i dati oggetto di rilevazione hanno reso impossibile effettuare confronti corretti nel tempo e nello spazio
  - **assenza** sostanziale **di controlli di qualità** e nessun investimento per migliorare.
- Comunicazione settimanale:
  - **statica**; **talvolta tendenziosa** (enfasi sul rischio relativo, anziché assoluto; anziani/ giovani; novax/vax);
  - ripetizione monotona (senza convinzione) di raccomandazioni ai cittadini;
  - **assenza** sostanziale **di controlli di qualità** e nessun investimento apparente per migliorare.
- **I 21 (in realtà 16) indicatori** per la classificazione per livello di rischio delle regioni; **fragilità dell'impianto** e quindi contenzioso fra Stato e Regioni.
- **I mass-media non hanno fatto un buon servizio**, salvo eccezioni (il problema non riguarda soltanto la pandemia Covid 19).

# Conclusioni/1

- Di fronte a un virus sconosciuto **l'incertezza è ineludibile ma può essere ridotta** e si può alimentare la consapevolezza sociale ponendosi le **domande giuste** e cercando di dare loro **risposte adeguate**.
- La **produzione di dati è centrale**. È necessario che essi siano raccolti ed elaborati all'interno di un disegno organico nel quale siano dominanti i valori dell'**interdisciplinarietà** e della **cooperazione interistituzionale**.
- Occorre promuovere l'integrazione fruttuosa fra i dati della **sorveglianza sanitaria** e quelli derivanti da **rilevazioni**, totali o campionarie.
- **Statistica e statistica ufficiale dovrebbero essere coinvolte maggiormente**. È importante che l'**Istat** diventi **protagonista**. L'esempio dell'ONS e della sua partnership con le autorità sanitarie e le università mostra la strada da intraprendere quanto prima.
- Uno specifico finanziamento all'interno del *Recovery Plan* potrebbe essere riservato all'allestimento di un **quadro informativo dinamico su Covid-19 e sui suoi effetti** su diversi aspetti della vita delle persone contagiate e non contagiate. Occorre attrezzarsi per il futuro.
- L'impianto di una **indagine longitudinale significativa** sarebbe l'investimento coerente.

# Cronologia delle grida di dolore e delle proposte

[La lotta al virus ha bisogno dei dati](#), Alleva G. e Zuliani A. (2020) Lettera al *Corriere della sera*, 24 marzo 2020.

[Numeri affidabili per uscire dalle nebbie dell'epidemia](#), Alleva G. e Zuliani A. (2020), *lavoce.info*, 30 marzo 2020.

[Audizione presso la Commissione Sanità del Senato della Repubblica](#), Alleva G. (2020).

[Coronavirus: chiarezza sui dati](#), Alleva G. e Zuliani A. (2020), *Bancaria*, 2020/3: 2-26, ISSN: 0005-4623.

[La pandemia dei dati: intervista a Giorgio Alleva e Alberto Zuliani](#). Patone M. (2020), *Scienza in rete*, 10 dicembre 2020.

[Covid 19: i dati comunicati giornalmente sono utili per capire?](#) Alleva G., Zuliani A. (2021) *Etica e Economia*, Menabò n.145/2021, 1 marzo 2021.

[Orientarsi nella pandemia Covid-19: i dati necessari](#), Alleva G. e Zuliani A. (2021), *Bancaria*, 2021/3: 20-35, ISSN: 0005-4623.

[Orientarsi nella pandemia Covid-19: i dati necessari](#), Alleva G. e Zuliani A. (2021), Conferenza invitata, *Invalsi*, 10 gennaio 2021.

[Orientarsi nella pandemia Covid-19: il ruolo della statistica e della statistica ufficiale](#), Alleva G., Arbia G., Zuliani A. (2021), Conferenza invitata, Accademia nazionale dei Lincei, 21 febbraio 2021.

[I ventuno indicatori e i colori delle regioni](#), Alleva G. e Zuliani A. (2021), *Etica e Economia*, Menabò n. 147/2021, 31 marzo 2021.

[A spatial sampling design for efficient estimation of the parameters of the SARS-CoV-2 epidemic](#), Alleva, G., Arbia, G., Falorsi, P., Nardelli, V., Zuliani, A. (2021), SEA 2021 Conference, Tokyo, 28 May 2021.

[A sample approach to the estimation of the critical parameters of the SARS-CoV-2 epidemics: an operational design](#), Alleva G., Arbia G., Falorsi P.D., Nardelli V., Zuliani A. (2022), accepted by *Journal of Official Statistics*.

[Spatial sampling design to improve the efficiency of the estimation of the critical parameters of the SARS-CoV-2 epidemic](#), Alleva G., Arbia G., Falorsi P.D., Nardelli V., Zuliani A. (2021), submitted to *Journal of Official Statistics*.

[Analyzing vulnerability in the context of the COVID-19 pandemic: An intersectional approach](#), Alleva G., Ambrosetti E., De Rose A., Miccoli S., Strangio D., Zannella M (2021), LVII Riunione Scientifica della SIEDS, 27 Maggio 2021.

[Epidemia di Covid-19: informazione, misure, comunicazione, comportamenti](#), Alleva G. e Zuliani A. (2022), in *Rapporto sullo stato sociale 2022*, a cura di Pizzuti F. R. *et al.*

