

# **Orientarsi nella pandemia Covid-19: i dati necessari**

Giorgio Alleva e Alberto Zuliani

Sapienza, Università di Roma

*Invalsi, 20 gennaio 2021*

# Impreparati alla seconda ondata

- A ottobre ci siamo ritrovati di fronte ai numeri di contagi e decessi di marzo, questa volta spalmati sull'intero territorio nazionale: **un'emergenza per la seconda volta drammatica.**
- Tutti abbiamo riflettuto sui ritardi, le inadeguatezze, la disorganizzazione, i rinvii, gli scaricabarile, le furbizie, ....
- **I dati sono centrali**; la confusione di quelli disponibili è grande e alimenta la distanza delle posizioni negli organi tecnici e nel confronto politico e istituzionale.
- Abbiamo testimoniato la nostra **sofferenza come statistici e come cittadini.**
- ... e forse (?) **ci sarà una terza ondata**, se non è già iniziata.

# Una variabilità elevata non facilmente decifrabile

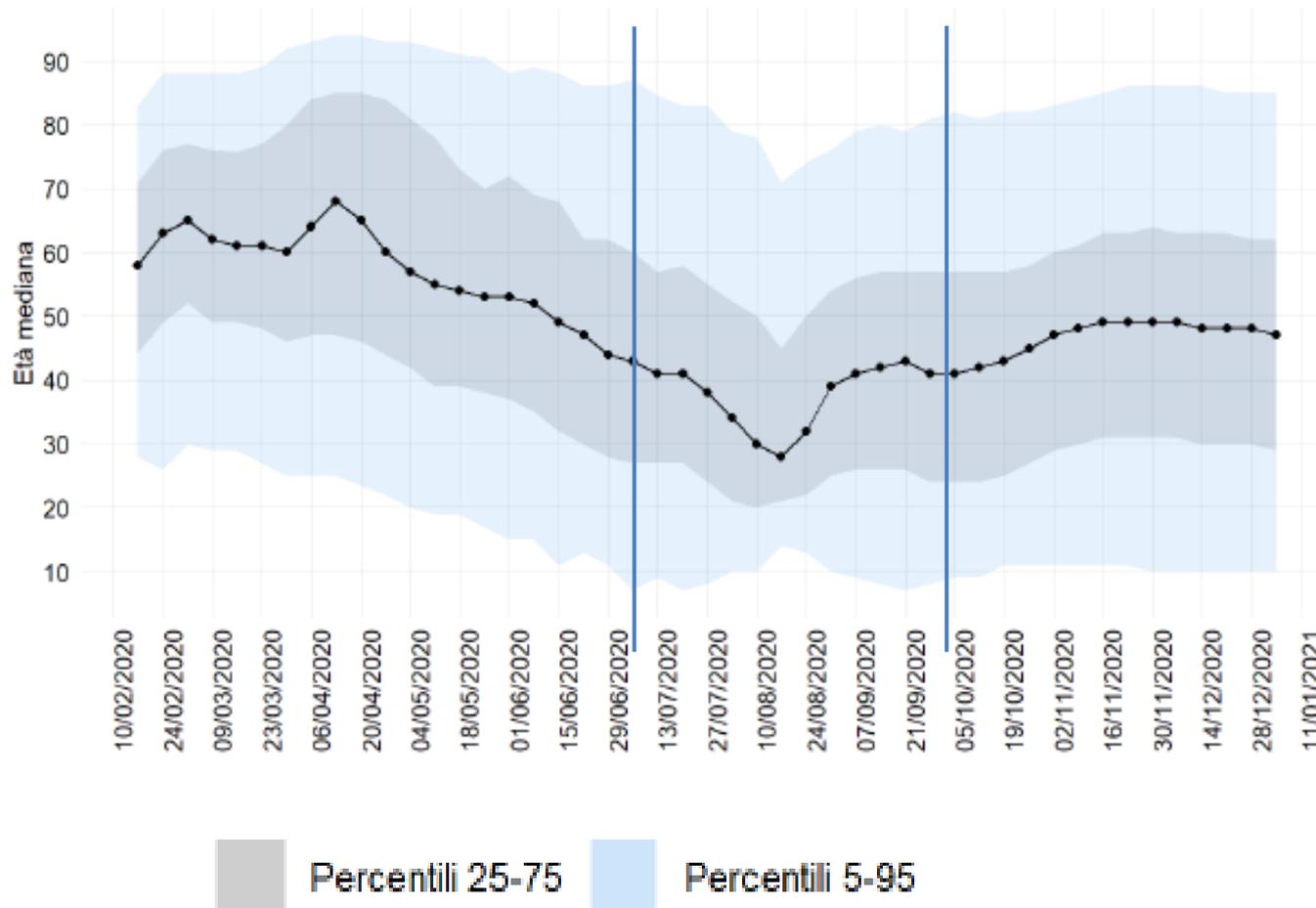
- La variabilità è molto elevata: i confronti nel tempo, fra paesi e da noi fra territori sono difficoltosi;  
**non emergono regolarità chiare.**
- I motivi
  - Specificità della sindrome e mutazioni del virus
  - Composizione per età della popolazione
  - Stato dei sistemi sanitari
  - Tipologia e intensità delle misure di contenimento
  - Stato dei sistemi di raccolta dei dati
  - Disomogeneità delle definizioni adottate
  - ...

# Qualche dato internazionale

- **Morti per milione di abitanti** (dato aggiornato al 19 gennaio 2021): Italia, 1.375; UK, 1.350; USA, 1.213; Spagna, 1.158; Francia, 1.095; Germania, 584; Giappone, 37; Corea del Sud, 25; Cina, 3.
- **Abbiamo imparato poco** dai paesi dell'estremo oriente quanto a sistemi di tracciamento e misure di contenimento.
- Lo **stato dei sistemi sanitari** (per la Cina si hanno poche informazioni) potrebbe spiegare una parte delle differenze.  
Prima della crisi per Covid-19 il numero di posti letto per 1.000 abitanti nella Corea del Sud (12,3) e in Giappone (12,8) superava di 1,6 volte quello della Germania, di 2,2 volte quello della Francia, 4,1 volte l'Italia, 4,4 volte la Spagna, oltre 5 volte Regno unito e USA.

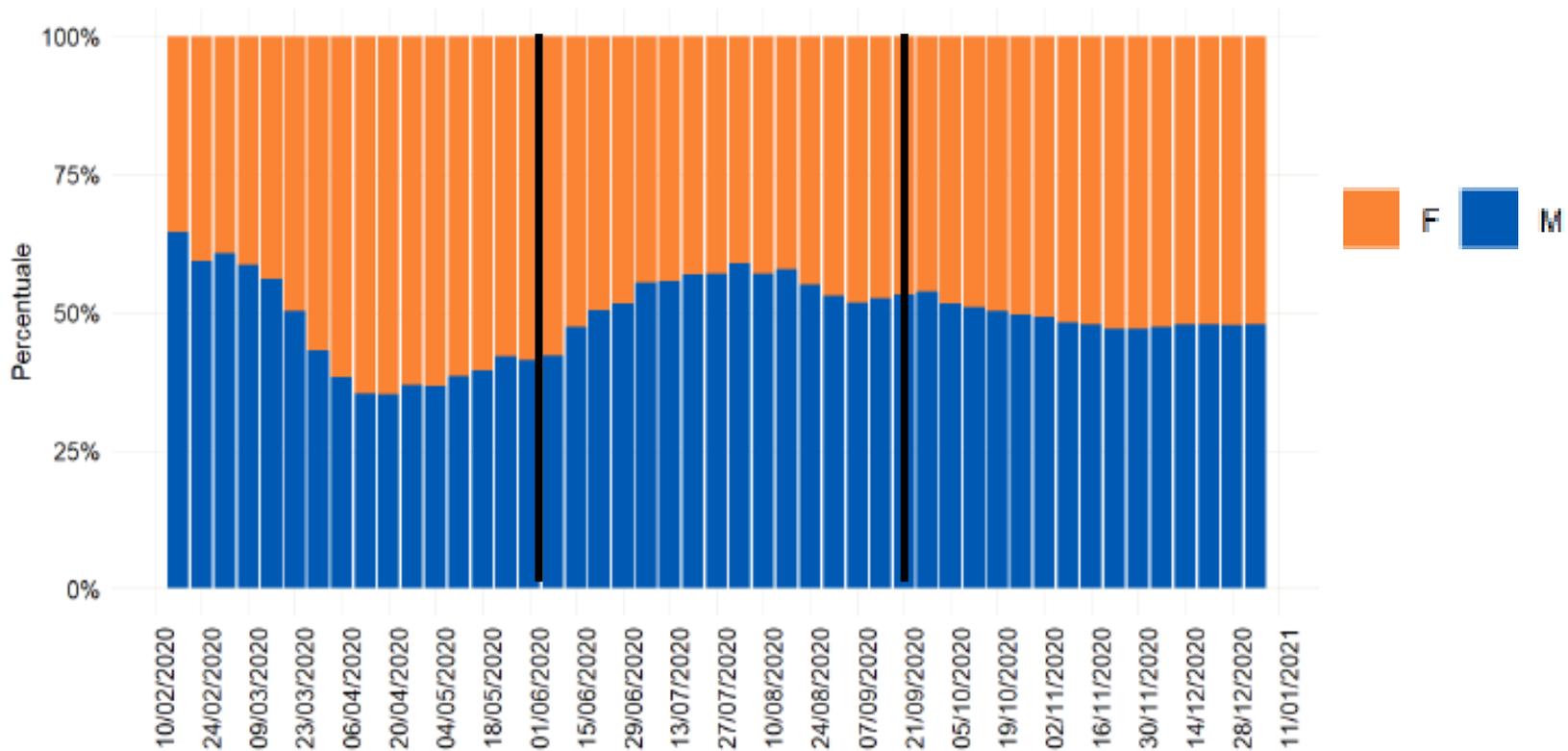
# Regolarità nel tempo non nitide

**Età mediana** dei casi diagnosticati per settimana di diagnosi.



# Regolarità nel tempo non nitide (segue)

Percentuale di casi diagnosticati per  **sesso**  e settimana di diagnosi



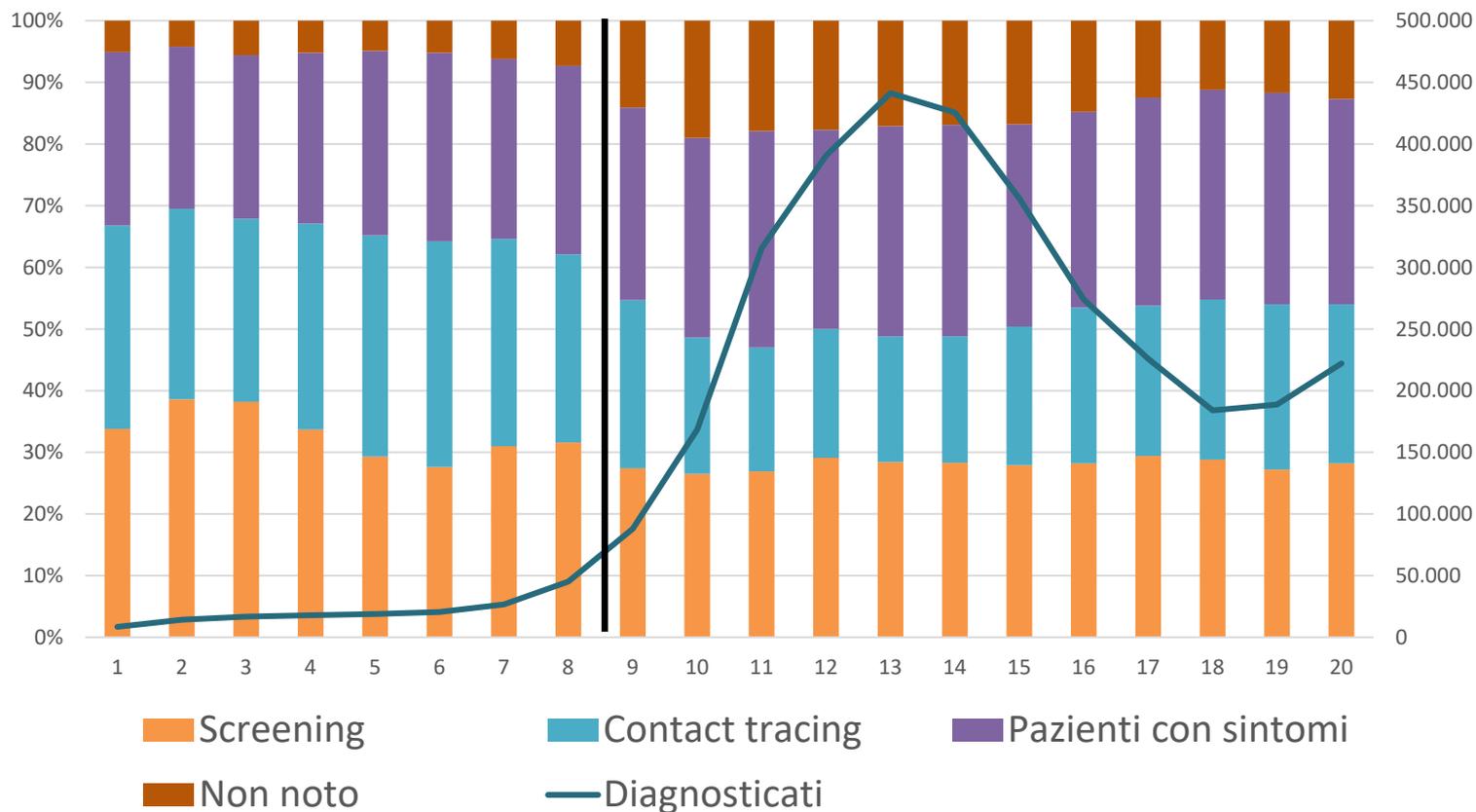
# Molte domande ancora senza risposta

- **Non** è citando insieme, giorno per giorno, il numero di tamponi processati, di casi positivi, di ricoverati e di decessi che **possiamo capire cosa stia accadendo**.
- Ognuno di noi si chiede: 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, se necessario presso i reparti di terapia sub-intensiva e intensiva; se entrati in ospedale, di uscirne guariti; in quali luoghi e contesti c'è rischio maggiore di contagio; infine, sulla base di quali informazioni vengono prese le decisioni circa le misure restrittive via via adottate?

# I contagi

- Il numero dei contagiati è strettamente collegato al numero di tamponi processati. Il numero di tamponi per 100.000 abitanti è risultato variabile nel corso del tempo ed è stato ed è tuttora variabile fra le regioni. A marzo-aprile scorso, i tamponi erano riservati essenzialmente ai sintomatici; attualmente, si distribuiscono fra sintomatici, contatti, *screening*. Il mix è cambiato nel corso del tempo ed è presumibilmente diverso regione da regione, cosicché i confronti risultano problematici.
- Non si conosce il numero di tamponi effettuati e processati per i tre motivi precedenti e verosimilmente i tassi di positività specifici sono differenti (quello per i sintomatici è prevedibilmente maggiore).
- Sul tasso di positività può influire la numerosità del *re-testing* che andrebbe scorporata.
- Da pochi giorni sono compresi nel denominatore anche i test rapidi e nel numeratore i contagiati che sono stati diagnosticati tramite essi. La serie dei tassi di positività avrà una discontinuità.

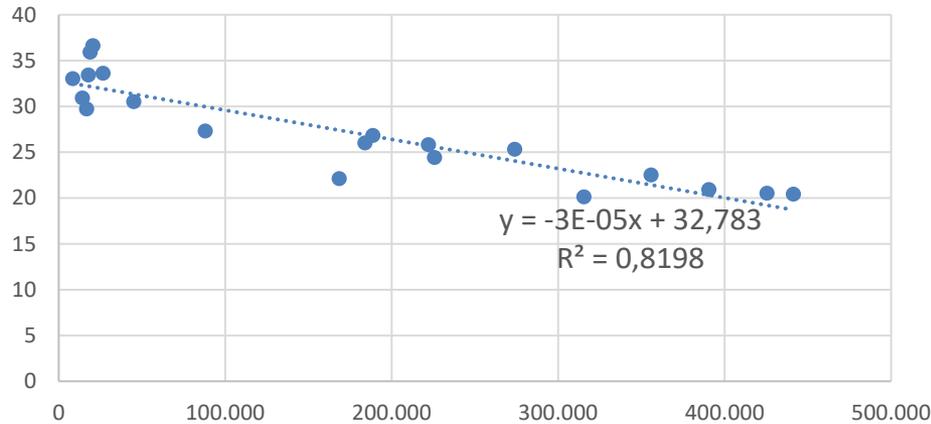
## Composizione percentuale dei tamponi secondo il motivo per il quale sono effettuati e per settimana (dal 25/8/2020 al 13/1/2021)



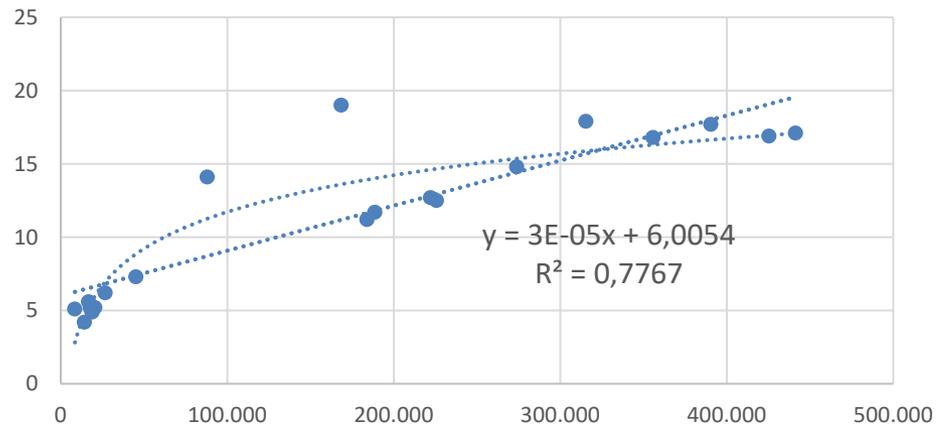
Nostra elaborazione su dati ISS

# Contagiati per motivo dell'accertamento (segue)

Percentuale di tamponi per *contact tracing* secondo il numero di casi diagnosticati

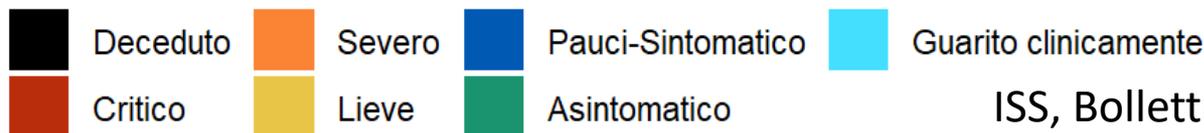
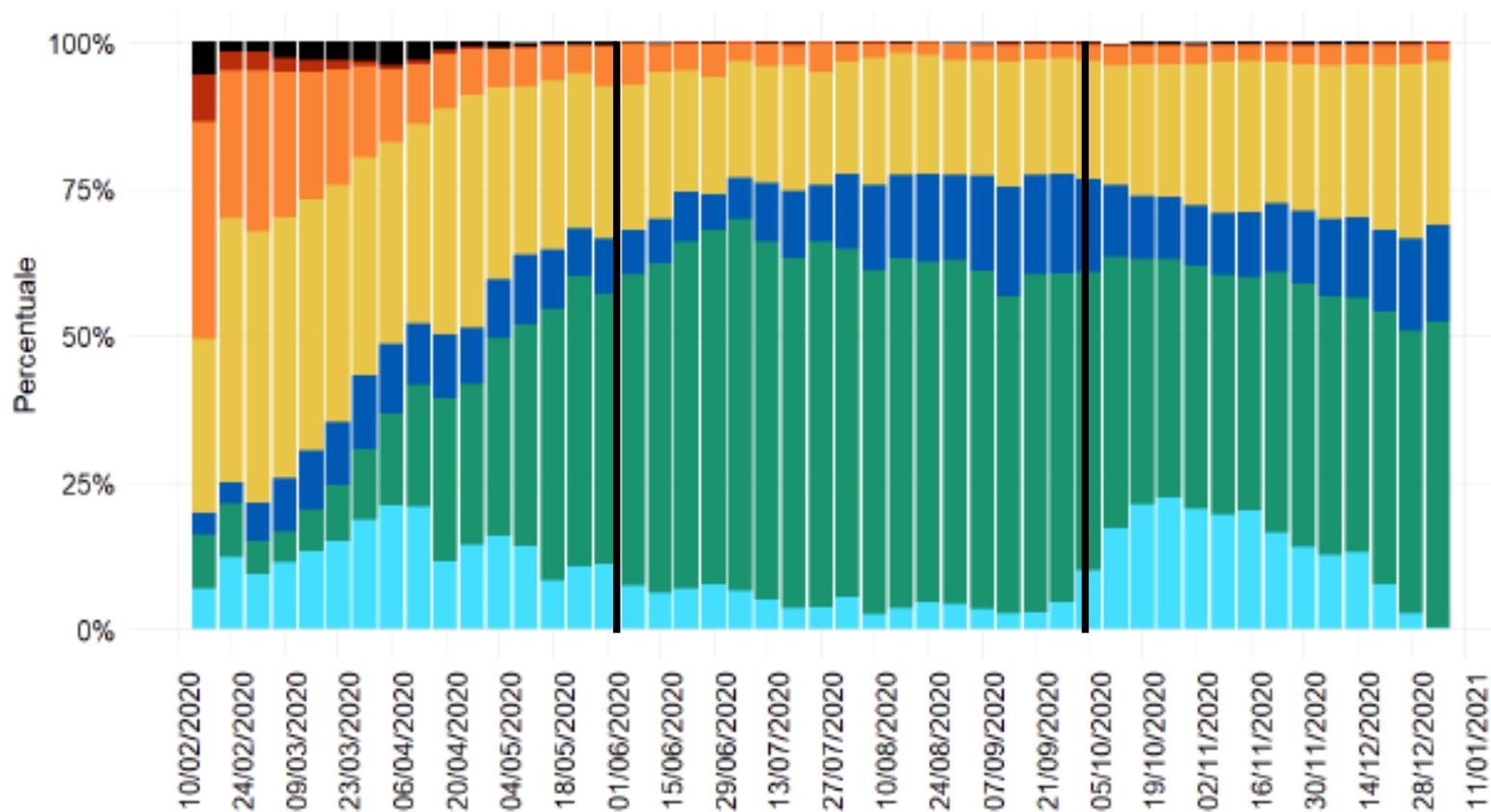


Percentuale di tamponi per motivo non noto secondo il numero di casi diagnosticati



# Contagiati per quadro clinico

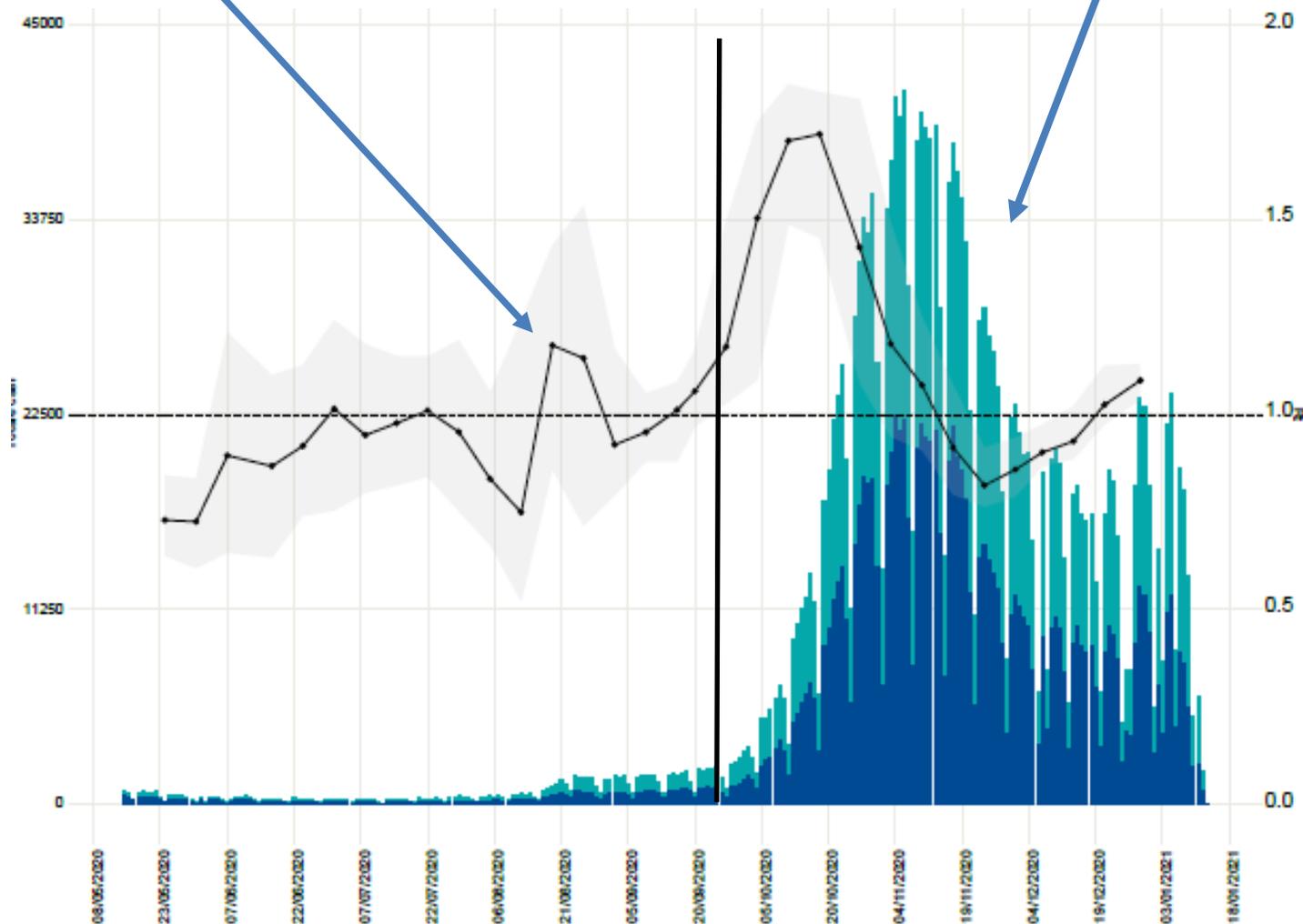
Percentuale di casi diagnosticati per stato clinico al momento della diagnosi nel tempo



# L'indice Rt

- È calcolato attualmente sui soli sintomatici, considerati maggiormente stabili.
- Viene comunicato quale stima puntuale e **trattato**, usualmente, **come se fosse deterministico**; mentre sconta un errore statistico abbastanza ampio.
- Fino al 25 novembre 2020 è stato fornito un *intervallo di credibilità al 95%*; dal 2 gennaio 2021 viene fornito un non meglio specificato *range* molto più ristretto.
- L'**intervallo di confidenza/credibilità** è risultato **estremamente variabile**, anche in settimane contigue: nel mese di settembre 2020, per il dato nazionale, ha raggiunto un'ampiezza massima di 0,82 punti; a ottobre di 0,66 punti; a novembre di 0,73 punti; a dicembre di 0,23 punti e a gennaio 2021, di 0,15 punti. Gli intervalli riferiti alle regioni sono in generale molto più ampi.
- Nell'ultimo bollettino pubblicato dall'Istituto superiore di sanità, l'indice relativo ai soli casi sintomatici, calcolato **per il periodo 23 dicembre 2020 – 5 gennaio 2021 è risultato pari a 1,09**, con un *range* 1,04-1,13, cioè 0,09 punti.

# Rt medio nel tempo e numero di casi (8/5/2020-13/01/2021)



Possiede data inizio sintomi ■ No ■ Si

per data prelievo/diagnosi

# Ricoveri

- I dati sui **ricoveri** sono **referiti agli stock** dei quali viene fornita la variazione giornaliera in più o in meno, un'informazione che non dà conto dei flussi. Sarebbe necessario che gli stock fossero aggiornati giorno per giorno con le **uscite** - per guarigione, trasferimento o morte - e gli **ingressi**, dettagliatamente per reparti Covid, terapia sub-intensiva e terapia intensiva, in modo da poter valutare appropriatamente la pressione sui reparti.
- Sarebbe possibile ottenere informazioni su: tasso di saturazione giornaliero delle strutture, durata media della degenza, tasso di guarigione e tasso di letalità dei ricoverati, informazioni forse disponibili ma non diffuse.

# Morti

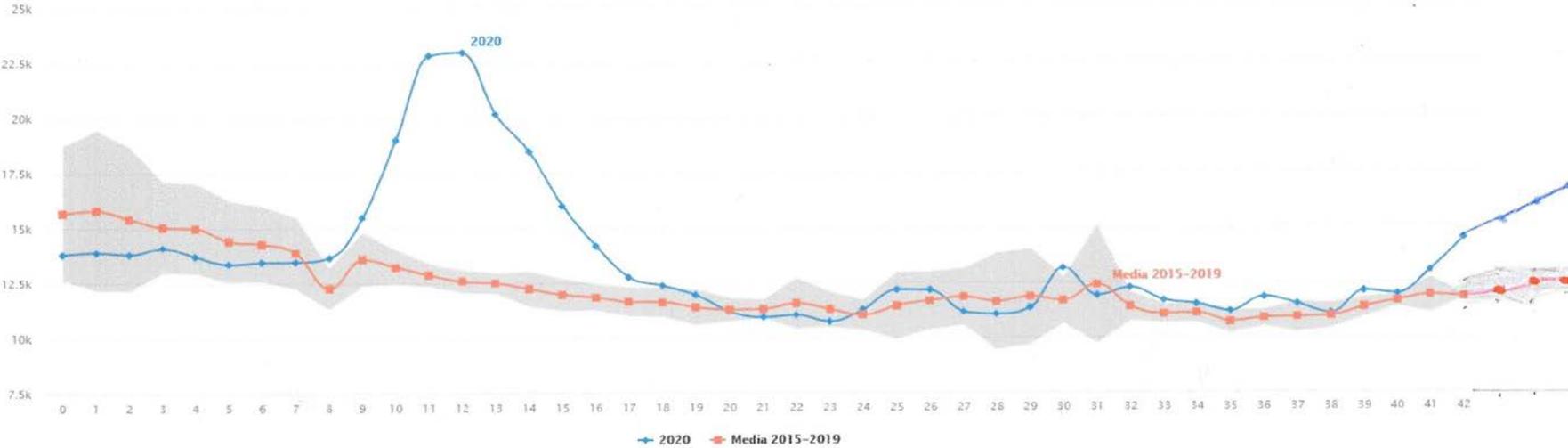
- **Confronti** temporali e spaziali del tasso di letalità per 100 contagiati e di mortalità per 100.000 residenti sono **problematici**.
- Il primo tasso ha seri problemi al denominatore dal quale mancano molti asintomatici e *pauci*-sintomatici rilevati soltanto parzialmente.
- Esso andrebbe fornito **distintamente** per i contagiati in terapia **domiciliare**, con dettaglio per le residenze per anziani, e **ospedaliera**.
- Si potrebbe **capire meglio per quali motivi i morti siano così numerosi**: carenze della medicina territoriale, anche riferita agli ospiti delle residenze per anziani; insufficienza di posti letto, specialmente di terapia sub-intensiva e intensiva; carenza di personale medico e infermieristico, oltre all'ormai evidente rischio per età.
- Sul secondo tasso, oltre alla sottostima del numero di morti, **influisce la diversa composizione per età della popolazione nei territori** e si può ricorrere a tassi standardizzati.

# Tasso di letalità al 13 gennaio 2021 (decessi per 100 contagiati)

Classi di età/genere	Tasso di letalità (%)		
	uomini	donne	totale
0-29	< 2/10.000	< 1/10.000	1/10.000
30-59	0,5	0,2	0,3
60-69	4,1	1,6	2,9
70-79	13,2	6,5	10,0
80-89	26,5	14,9	19,6
90 e più	36,8	21,3	24,9
<b>Totale</b>	<b>4,0</b>	<b>2,9</b>	<b>3,4</b>

ISS, Bollettino 13 gennaio 2021

# Eccesso di mortalità nel tempo (fino al 30 novembre)



## Eccesso di mortalità: prima e seconda ondata (per età)

Descrizione	Marzo-maggio				Ottobre-novembre			
	Decessi 2020	Eccesso su media 2015-19	Stime		Decessi 2020	Eccesso su media 2015-19	Stime	
			Per Covid- 19 (%)	Differenza (%)			Per Covid-19 (%)	Differenza (%)
<b>Totale</b>	211.047	50.254	68%	32%	134.767	31.715	67%	33%
<b>&gt; 80 anni</b>	138.307	36.424	52%	48%	88.425	23.428	55%	45%

Nostra elaborazione su dati ISS e Istat.

- Nella seconda ondata la dimensione dell'eccesso di mortalità è un poco inferiore, ma la **quota non registrata come 'da Covid' (1/3)** è la stessa.
- Nella prima ondata l'eccesso non registrato può essere soprattutto un 'Covid sommerso'; nella seconda, una mortalità «collegata» all'epidemia.
- **Negli over 80 la quota è intorno al 50%.**
- Ci sono evidenze analoghe in altri paesi (Bucci E., L'orrida conta dei morti).

# Prima e seconda ondata a confronto

- Modello moltiplicativo:

$$\frac{M}{P} = \frac{M}{C} \times \frac{C}{T} \times \frac{T}{P}$$

- Linearizzazione logaritmica

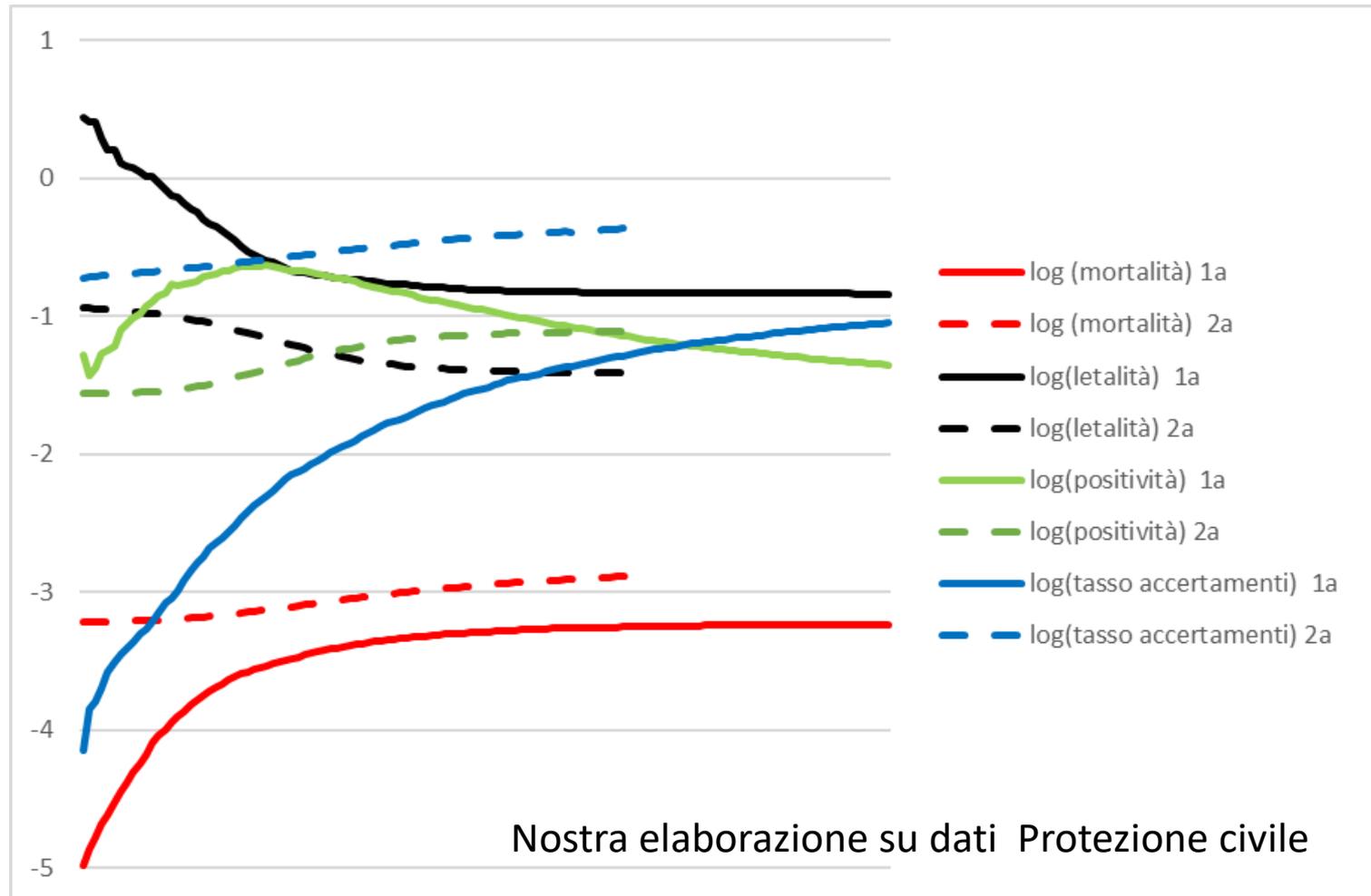
$$\log \frac{M}{P} = \log \frac{M}{C} + \log \frac{C}{T} + \log \frac{T}{P}$$

*(con M riferito al 15° giorno successivo all'accertamento dei contagi)*

- Apprezzamento dei **ritmi di evoluzione nel tempo** di: tasso di mortalità (M/P), tasso di letalità (M/C), tasso di positività (C/T), tasso di accertamento dei contagi (T/P).
- Confronto fra le due ondate.

# Prima e seconda ondata a confronto (*segue*)

Serie giornaliera dei logaritmi dei tassi nella prima e nella seconda ondata  
(24.02 - 30.06 e 01.10 - 02.01)



# Cosa funziona

- L'uso di mascherine, igiene personale, distanziamento sociale e *lockdown* hanno dimostrato di essere **buone pratiche** per contrastare il contagio (Brauner J. M. *et al.*, 2020).
- L'anticipo dei tempi del *lockdown* (a suo tempo previsto in modo generalizzato nel Paese) **ha fermato la prima ondata nelle regioni centrali e nel Mezzogiorno.**
- Il **comportamento delle persone è stato fondamentale** nel corso della prima ondata.
- Il **numero di posti in terapia** intensiva e sub-intensiva è stato **ampliato.**
- I **protocolli terapeutici** si sono **affinati.**
- Vedremo come funzioneranno i **vaccini**; stiamo per assistere a una **sperimentazione planetaria.**

# Cosa non ha funzionato

- Tracciamento; **il flop di *Immuni***.
- **Medicina territoriale**: né argine al ricorso ai ricoveri ospedalieri, né protocolli uniformi di trattamento domiciliare, né coinvolgimento nei programmi nazionali; grande variabilità.
- **Rapporti istituzionali** fra Stato, Regioni ed Enti locali.
- In generale, **organizzazione di sistema** debole.

# I dati necessari per capire

- Non accontentarsi dei dati disponibili.
- Costruire e alimentare un **sistema di monitoraggio multi-fonte**.
- Integrare **dati sanitari e dati campionari**.
- Raccogliere i **dati secondo un disegno** che li renda realmente informativi.
- Predisporre **disegni sperimentali** (ad esempio, per capire come abbiano funzionato le misure di contenimento del contagio. Le scuole sarebbero state un buon campo di sperimentazione).
- Possibile (necessaria) **presenza dell'Istat**.

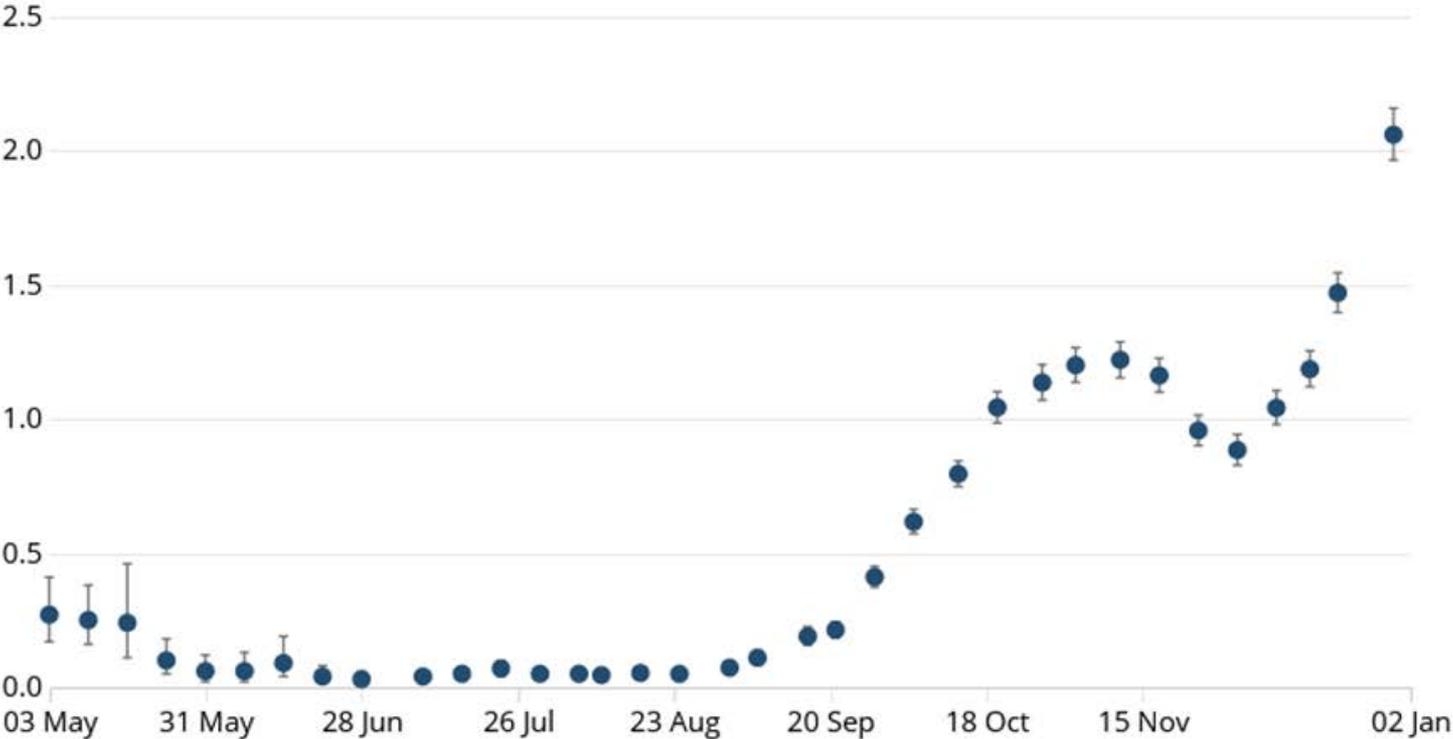
# Sistema informativo Covid-19

## Una proposta

- Integrazione fra sorveglianza sanitaria e monitoraggio statistico.
- Campione ripetuto nel corso del tempo per stimare la prevalenza e seguirne l'evoluzione.
- Panel per verificare i cambiamenti di stato delle persone in rapporto alla pandemia.
- Campione (o rilievo totale) di scuole e alunni per seguire l'evoluzione dei contagi nel sistema educativo.
- Campione (o rilievo totale) di medici di base (I stadio) mediante il quale stimare il numero di pazienti (II stadio) con quadri sintomatici riconducibili a Covid-19 e seguirne gli esiti.
- Copertura informativa del piano vaccinale.

*I dati riferiti alle persone dovrebbero essere resi disponibili per classi di età*

Percentage testing positive for COVID-19



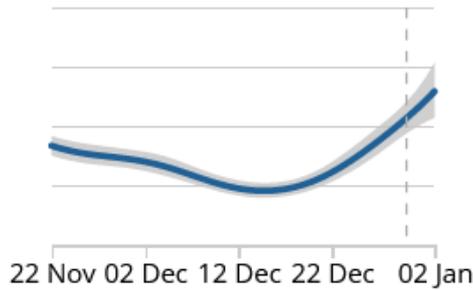
Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

■ 95% credible interval — Modelled percent testing positive for COVID-19 by region

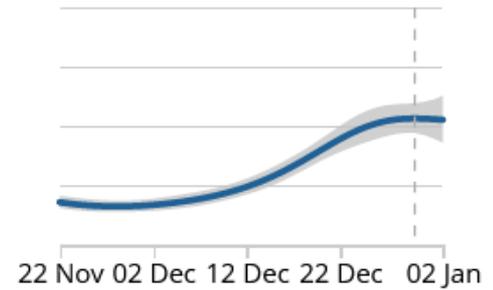
### London



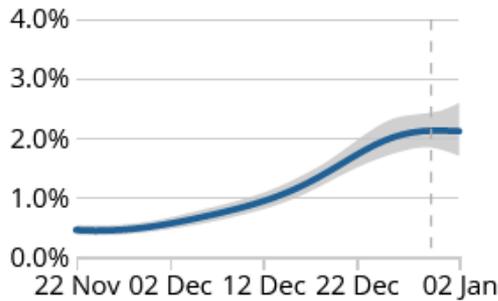
### North West



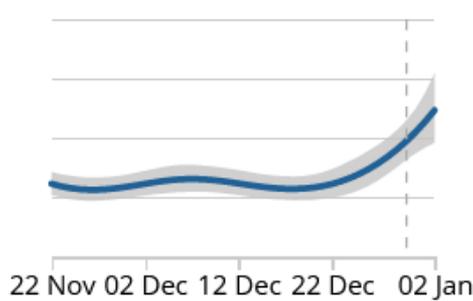
### South East



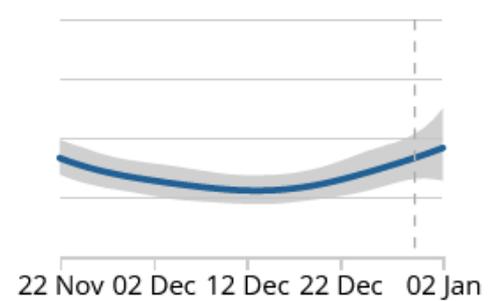
### East of England



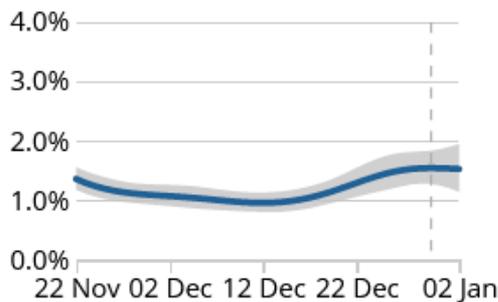
### East Midlands



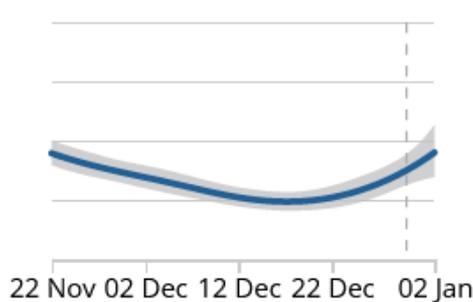
### North East



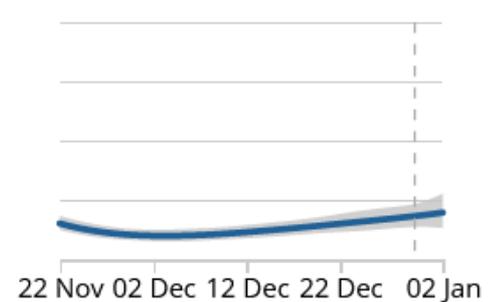
### West Midlands



### Yorkshire and The Humber



### South West



# Speciale scuola

- «Il 30 dicembre l'Istituto superiore di sanità ha pubblicato un report ampio sull'incidenza delle scuole sull'epidemia, possiamo supporre che il dibattito istituzionale delle scorse ore ruotasse attorno a quello. **Peccato che di evidenza, nel report, ne compaia poca**: 3.173 focolai documentati e chiaramente riconducibili all'ambito scolastico, che costituirebbero circa il 2% di quelli totali.» ... «Ma i focolai scolastici sono, per ammissione stessa del report, sottostimati, e non abbiamo alcuna idea di quanto». (Paolo Giordano, La giusta scelta per le scuole: ripartire (subito) dai dati, *Corriere della sera*, 5 gennaio 2020).
- L'**Invalsi** svolgerà il testing consueto nei mesi di marzo-maggio prossimi. Sarebbe utile che predisponesse un progetto specifico per il confronto delle performance degli studenti con dad e con didattica in presenza.
- Iniziative altre; Sapienza; Anvur.

# I ventuno indicatori del Ministero della Salute (D.M. 30 aprile 2020)

- Sei riguardano la **capacità di monitoraggio**; due riferiti alle residenze sanitarie assistite sono opzionali. Gli altri quattro sono **banali e poco rilevanti**.
- Sei riguardano la **capacità di accertamento diagnostico**; uno, il tempo medio intercorso fra la data di inizio dei sintomi e quella di isolamento è opzionale; tre sono **velleitari**.
- Nove riguardano l'**intensità di trasmissione del virus** e la **tenuta dei servizi sanitari**; due sono opzionali (uno è il numero di accessi al pronto soccorso con quadri sindromici riconducibili a Covid-19, che sarebbe di interesse); quattro sono riferiti ai contagiati, con prevedibile collinearità; gli altri **tre sono effettivamente rilevanti**: tasso di occupazione di posti letto in area medica per pazienti Covid-19; tasso di occupazione di posti letto in terapia intensiva per pazienti Covid-19; Rt (con i limiti già segnalati).

*Non c'è alcun tentativo di ricercare corrispettivi empirici maggiormente pertinenti e probanti; non sono indicati pesi nonostante gli indicatori abbiano sostanza diversa ; il trattamento è alquanto arbitrario*

# Dalla naïveté alla black box

Come funziona l'analisi del rischio epidemico delle Regioni

## ANALISI DEL RISCHIO DI EPIDEMIA NON CONTROLLATA E NON GESTIBILE

21 indicatori → 16 non opzionali (compilati)

### VERIFICA DI QUALITÀ DEL DATO (4 INDICATORI)

#### PROBABILITÀ DI DIFFUSIONE

- **Aumento dei casi rispetto alla settimana precedente** (2 indicatori da due flussi Mds -DG PREV e ISS)
  - **Rt >1 sulla data di inizio sintomi** (1 indicatore)
  - **Aumento dei focolai rispetto alla settimana precedente** (1 indicatore flusso settimanale con Regioni/PA):
    - analisi della dimensione dei focolai
    - valutazione di qualità del contact tracing; Numero di nuovi casi non associati a catene di trasmissione note (1 indicatore)
  - **Criterio di Autovalutazione**
- 5 indicatori di cui 4 principali (almeno 2 segnali)**

#### IMPATTO

- **Criterio 1: Aumento di casi in soggetti >50aa**
  - **% occupazione dei posti letto in area medica** (1 indicatore – dati Mds DG PROG)
  - **% occupazione dei posti letto in terapia intensiva** (1 indicatore – dati Mds DG PROG)
  - **Criterio qualitativo: focolai in popolazioni vulnerabili** (dati delle Regioni/PA ed epidemic intelligence)
- 2 indicatori + 2 criteri**

#### RESILIENZA TERRITORIALE

- **Aumento della % di positività al tampone** e monitoraggio % casi diagnosticati in ospedale ed in altri contesti (1 indicatore)
  - **Tempi brevi tra data inizio sintomi e diagnosi** (1 indicatore)
  - **Numero sufficiente di risorse umane** (2 indicatori valutati insieme)
  - **Capacità di effettuare per tutti i nuovi casi una regolare indagine epidemiologica con ricerca dei contatti stretti** (1 indicatore)
  - **Criterio aggiuntivo: Eventuali segnali di criticità** aggiuntive riportati dalle Regioni vengono analizzati e considerati secondo un criterio di precauzione
- 4 indicatori + 1 criterio aggiuntivo (almeno 2 segnali)**

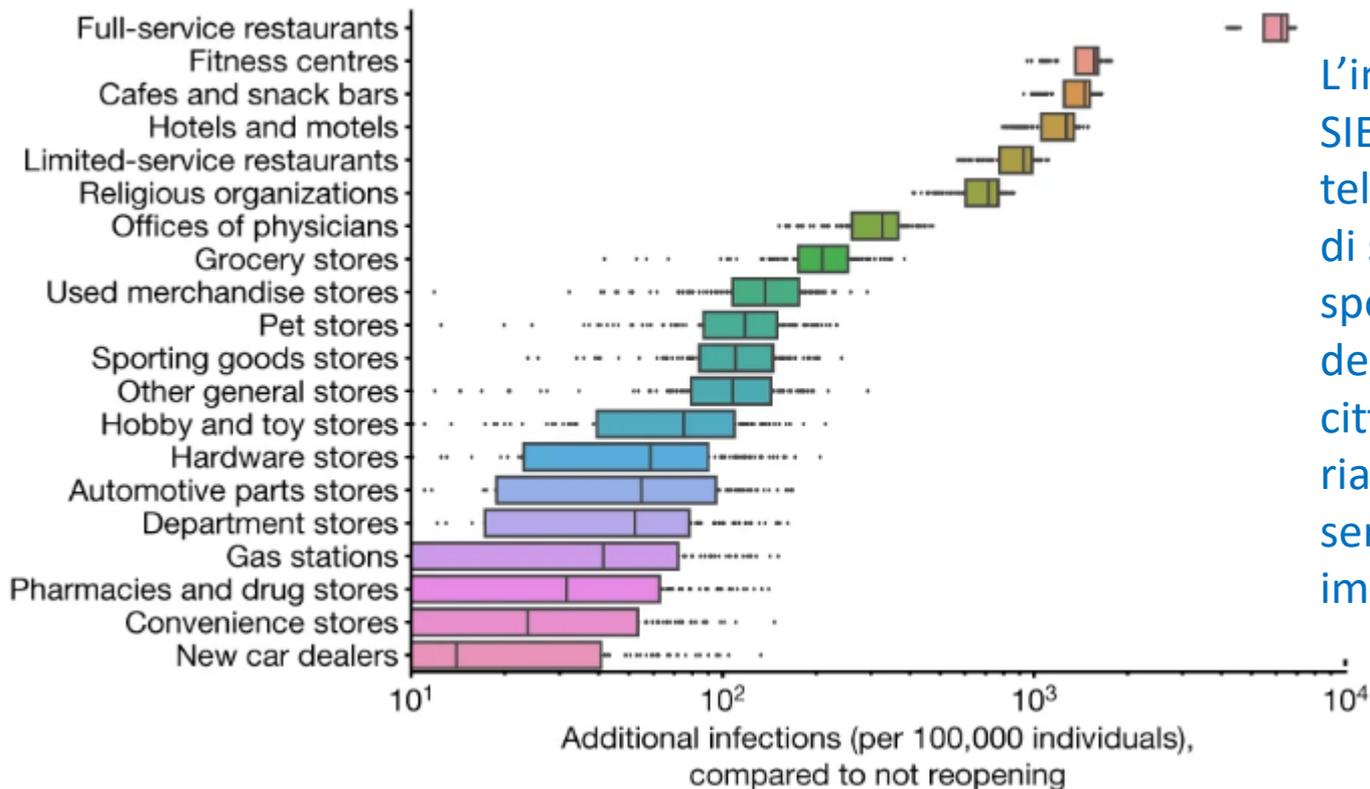
### VALUTAZIONE

# L'impiego auspicabile delle tecnologie dell'informazione e dei *big data*

- Le **nuove fonti (*big data*)** possono fornire un **contributo prezioso**:
  - sensori per rilevare il tasso di **occupazione nei trasporti pubblici**;
  - dati della telefonia mobile per stimare la dimensioni dei **flussi di spostamenti** tra territori nel corso della giornata e durante i giorni della settimana e anche le presenze in specifici luoghi (**assembramenti**);
  - *scanner data* per studiare come si siano modificate le **abitudini di consumo** nella grande distribuzione;
  - il *Daily Social Mood Index* che l'Istat ha sviluppato nel 2018 sulla base di un campione esteso di Twitter rappresenta un buon esempio di sperimentazione utile.
- Un recente articolo su *Nature* presenta i risultati di *dynamyc network models* per simulare la diffusione di SARS-CoV-2 in **dieci** nelle maggiori **aree metropolitane degli Stati Uniti** a un livello territoriale molto fine, basato sui dati della telefonia mobile.

*Occorre costruire partnership strategiche tra le società che possiedono quei dati e le istituzioni pubbliche che hanno la responsabilità della gestione sanitaria dell'epidemia, in modo da trasformarli in bene pubblico*

# L'impiego auspicabile delle tecnologie dell'informazione e dei *big data*



L'integrazione di modelli SIER con i dati della telefonia mobile consente di stimare l'impatto degli spostamenti sulla dinamica dei contagi all'interno delle città e gli effetti della riapertura o meno dei servizi alle famiglie e alle imprese

Chang, S. et al. (2020). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*.

# Valutazione della campagna vaccinale

- **Successo straordinario e inimmaginabile della ricerca scientifica:** vaccini 'validati' sono somministrati in tutto il mondo, a meno di un anno dall'inizio della pandemia.
- **Siamo protagonisti di una gigantesca sperimentazione sul campo:** la validazione di Pfizer è avvenuta su un campione complessivo di circa 40.000 soggetti (tra vaccinati e placebo). Le infezioni registrate sono state 170, 8 fra i vaccinati e 162 fra le persone trattate con placebo. L'esito della sperimentazione è stato positivo ma i numeri sono molto piccoli.
- **Grande investimento pubblico** in partnership con aziende e multinazionali: la gestione delle vaccinazioni dovrebbe assicurare **inclusività** ed **equità** tipiche di un bene comune.
- Nel nostro Paese la dimensione e la dinamica delle somministrazioni in questa prima fase appaiono incoraggianti, seppur emergano differenze nel territorio. Bisogna **assicurare lo stesso diritto alla salute** a tutta la popolazione.

## Alcune domande

- E' stata disegnata una **base di dati nazionale** che registri la popolazione vaccinata, unitamente alle sue caratteristiche anagrafiche e socio-economiche essenziali? È importante per comprendere l'effettivo avvicinamento alla c.d. immunità di gregge.
- Soprattutto, è stata disegnata un'**indagine che monitori nel corso del tempo le proprietà immunologiche dei diversi vaccini** su un campione di soggetti selezionati in modo da essere rappresentativi della popolazione e delle caratteristiche potenzialmente collegate? **Sarebbe bello che questo studio lo facessimo in Italia**, ma da qualche altra parte del mondo le risposte verranno comunque.

# Data set pubblici di dati individuali

- I dati raccolti dovrebbero costituire *data-set pubblici*; i dati individuali dovrebbero essere resi anonimi, comunque non riconoscibili.
- I dati sono invece tenuti riservati: problema di *gelosie istituzionali* o di *scarsa qualità*?
- Non vengono stimulate la ricerca e l'interdisciplinarietà.
- Molti scienziati delle discipline più diverse chiedono con forza che i dati elementari vengano messi a disposizione per la ricerca. *Finora nessuna risposta.*

# Conclusioni

- L'**incertezza** è ineludibile ma **può essere** fortemente **ridotta**.
- Dati raccolti e integrati secondo un **disegno organico**.
- Contrastare **comunicazione senza informazione**.
- Promuovere **trasparenza, credibilità e fiducia**.
- Collegare **organizzazione e responsabilità**.
- Coinvolgere la **statistica** e la **statistica ufficiale**.
- *Recovery Fund*: occasione per stimolare la **virtuosità informativa-digitale** a partire dai dati su Covid-19.
- Il **cammino è ancora lungo**.

# Sito-bibliografia essenziale

Alleva G., Arbia G., Falorsi P.D., Zuliani A. (2020), A sample approach to the estimation of the critical parameters of the SARS-CoV-2 epidemics: an operational design.

<http://arxiv.org/abs/2004.0606>, accettato da *Journal of Official Statistics (JOS)*.

Alleva G. e Zuliani A. (2020), Coronavirus: chiarezza sui dati, *Bancaria*, ISSN: 0005-4623

Alleva G. e Zuliani A. (2020), La lotta al virus ha bisogno dei dati, Lettera al Corriere della sera, 18 ottobre 2020.

Brauner J.M. et al (2020), Inferring the effectiveness of government interventions against COVID-19, *Science* 10.1126/science.abd9338.

Bucci E. (2021), L'orrida conta dei morti per Covid può aiutarci a capire se e come il virus si sta ritirando dal mondo, *Il Foglio*, 15 gennaio 2021.

Chang, S. et al. (2020), Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*, <https://doi.org/10.1038/s41586-020-2923-3>.

Giordano P. (2021), La giusta scelta per le scuole: ripartire (subito) dai dati, *Corriere della sera*, 5 gennaio 2021.

Parisi, G. (2020). L'indice Rt è inaffidabile, Blog di Huffington post, 15 novembre 2020, [https://www.huffingtonpost.it/entry/covid-indice-rt\\_it\\_5fb05a2ec5b6c5f3d2f7f156](https://www.huffingtonpost.it/entry/covid-indice-rt_it_5fb05a2ec5b6c5f3d2f7f156).

Parisi, G. (2020). Servono provvedimenti drastici adesso, *Scienza in rete*, 23/10/2020.

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

Grazie per l'attenzione