

STRATEGIE DI SEQUENZIAMENTO PER L'IDENTIFICAZIONE DELLE  
VARIANTI DI SARS-CoV-2 ED IL MONITORAGGIO DELLA LORO  
CIRCOLAZIONE IN ITALIA - *INDICAZIONI AD INTERIM*

VERSIONE N. 2 DEL 15-03-2022

## Sommario

Premessa .....	2
Obiettivo specifico 1 - Stimare la prevalenza delle principali varianti circolanti in Italia tramite indagini “Flash survey” .....	2
Obiettivo specifico 2 - Stabilire un flusso continuo di sequenziamenti che consenta di Identificare la comparsa di varianti nuove o già note di SARS-CoV-2 di interesse di sanità pubblica sul territorio italiano .....	3
Appendice.....	8

## Tabelle e Figure

<b>Figura 1 e Tabella 1.</b> Probabilità di identificare almeno un campione con una specifica variante con un campione settimanale di 500 sequenze per diverse dimensioni campionarie al variare della prevalenza attesa e del livello di diffusione dell’epidemia.....	4
<b>Figura 2 e Tabella 2.</b> Probabilità di identificare almeno un campione con una specifica variante mediante un campione settimanale di 750 sequenze per diverse dimensioni campionarie al variare della prevalenza attesa e del livello di diffusione dell’epidemia.....	5
<b>Figura 3 e Tabella 3.</b> Probabilità di identificare almeno un campione con una specifica variante con un campione settimanale di 1000 sequenze per diverse dimensioni campionarie al variare della prevalenza attesa e del livello di diffusione dell’epidemia .....	6
<b>Tabella 4.</b> Numero di campioni settimanali da sequenziare in ciascuna Regione/PA .....	7

## Premessa

La sorveglianza genomica dei virus è di grande importanza per identificare precocemente l'eventuale emergenza di nuove varianti virali e monitorare la loro circolazione sul territorio. Il sequenziamento dell'intero genoma (WGS) o almeno il sequenziamento completo o parziale del gene S dovrebbe essere eseguito per confermare l'infezione con una specifica variante.

L'obiettivo fissato dall'ECDC consiste nel fare un numero di sequenziamenti tali da garantire l'identificazione di varianti che circolino con una prevalenza del 5% o inferiore. Attualmente l'Italia sequenzia una percentuale di campioni positivi che garantisce la richiesta ECDC ed infatti, nei report ECDC che monitorano le attività di sequenziamento, l'Italia è sempre collocata tra i Paesi europei che rispettano tale livello di performance. La numerosità dei sequenziamenti settimanali effettuati in Italia dallo scorso aprile aveva infatti una probabilità del 100% di identificare varianti di interesse che circolano con una prevalenza superiore all'1% e quindi ben al di sotto della soglia richiesta da ECDC.

Alla luce di queste evidenze, al fine di ottimizzare le modalità di sequenziamento in modo da ottenere dati che garantiscano **un flusso continuo di informazioni (sul quale si basa l'identificazione precoce di nuove varianti)** e, allo stesso tempo, **la stima ad elevata precisione della loro prevalenza**, si suggeriscono due distinti flussi basati su diversi criteri di campionamento e obiettivi specifici.

Questo documento, da considerarsi nella sua versione *ad interim*, potrà subire delle ulteriori revisioni in relazione al quadro epidemiologico in rapida evoluzione, e alla circolazione di specifiche varianti di interesse in sanità pubblica o di nuove varianti di SARS-CoV-2.

### Obiettivo specifico 1

*Stimare la prevalenza delle principali varianti circolanti in Italia tramite indagini "Flash survey".*

Per questo obiettivo si rimanda alle note tecniche utilizzate in tutto l'anno 2021 e 2022 (<https://www.epicentro.iss.it/coronavirus/sars-cov-2-monitoraggio-varianti-indagini-rapide>) per le cosiddette "Flash survey", che hanno l'obiettivo di stimare la prevalenza delle varianti virali circolanti in Italia, in collaborazione con la Fondazione B. Kessler, il Ministero della Salute, le Regioni/PPAA.

Il campionamento è casuale e stratificato per Regione/PA in un giorno concordato del mese. La dimensione campionaria per Regione/PA viene calcolata separatamente a livello di macro area (4 macroaree - nomenclatura NUTS): Nord-Ovest (Piemonte, Valle d'Aosta, Liguria, Lombardia), Nord-Est (Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna), Centro (Toscana, Umbria, Marche, Lazio), Sud e Isole (Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sardegna, Sicilia).

Il campione necessario  $n$ , per ciascuna macroarea per poter osservare una variante con prevalenza  $p$  sul territorio nazionale in una popolazione di ampiezza  $N$  (popolazione dei positivi notificati) e

stimare questa prevalenza con una precisione  $\varepsilon$  e livello di confidenza  $(1 - \alpha)\%$  è dato dalla seguente formula

$$n \geq \frac{N z_{\alpha/2}^2 p(1-p)}{\varepsilon^2 (N-1) + z_{\alpha/2}^2 p(1-p)}$$

Tenendo conto del fatto che sul territorio circolano varianti con diverse prevalenze, si possono stimare prevalenze intorno a 1%, 10% o 50% con precisione rispettivamente intorno a 0.9%, 2.7% e 4.6% nelle 4 macro-aree considerate.

Inoltre, con l'ampiezza campionaria calcolata come sopra riportata è possibile osservare in ogni macro-regione varianti che circolano intorno allo 0.5-1% con un livello di confidenza del 95%.

Si specifica che l'ampiezza campionaria in ciascuna macroarea viene poi ridistribuita nelle Regioni e PPAA corrispondenti in base alla percentuale di casi notificati in uno dei giorni della settimana precedente sul totale di casi notificati nella macroarea di riferimento.

## Obiettivo specifico 2

*Stabilire un flusso continuo di sequenziamenti che consenta di identificare la comparsa di varianti nuove o già note di SARS-CoV-2 di interesse di sanità pubblica sul territorio italiano.*

Il campionamento, in questo caso, non è casuale, non riguarda cioè tutta la popolazione dei casi positivi, ma mirato a particolari gruppi di popolazione:

1. soggetti con manifestazioni cliniche severe che richiedano ospedalizzazione (30-40%);
2. soggetti ricoverati in terapia intensiva (20-30%);
3. soggetti con reinfezioni (definiti dalla nota DGPrev del 24 agosto 2021) (minimo 5-10%);
4. soggetti immunocompromessi (5-10%)
5. si raccomanda di includere una quota di soggetti di rientro da Paesi che risultino attenzionati dagli Organismi Internazionali per l'emergenza di nuove varianti di interesse di sanità pubblica (5-10%)

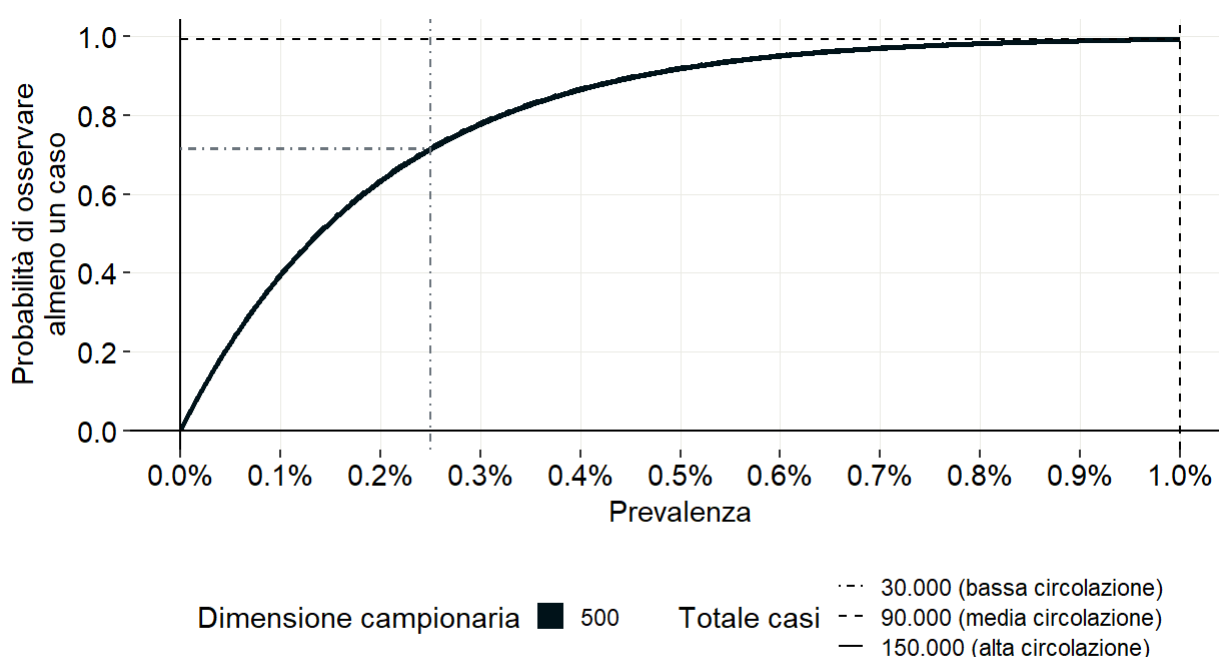
Nel caso i cui non si riuscisse a raggiungere alcune delle numerosità relative alle categorie sopra descritte, (in particolare per la categoria 4 e la categoria 5) le categorie restanti dovranno essere rimodulate proporzionalmente per arrivare al 100%, facendo il possibile per raggiungere la numerosità campionaria minima indicata per Regione/PPAA.

Si specifica, inoltre, che forme cliniche atipiche di COVID-19 siano auspicabilmente oggetto di sequenziamento e opportunamente segnalate in Note nella piattaforma della sorveglianza integrata.

È da sottolineare che per effettuare un sequenziamento è necessario partire da un campione processato per indagine in real-time PCR. Questa realtà può avere un impatto sulle stime di prevalenza e sulle modalità di campionamento.

Le figure e le tabelle seguenti riportano la probabilità di identificare almeno un campione con una specifica caratteristica (ad esempio una variante di interesse emergente) sulla base della prevalenza attesa stimata fino all'1% del livello di diffusione della infezione da Sars-CoV-2 (qui ipotizzato di 30.000, 90.000 e 150.000 infezioni settimanali, casi corrispondenti a bassa - 50 casi per 100.000 abitanti, intermedia - 150 casi per 100.000 abitanti e alta diffusione - 250 casi per 100.000 abitanti) e di tre diverse dimensioni campionarie qui ipotizzate. Per valori di prevalenza più elevati (>1%) la probabilità di identificazione è ancora maggiore.

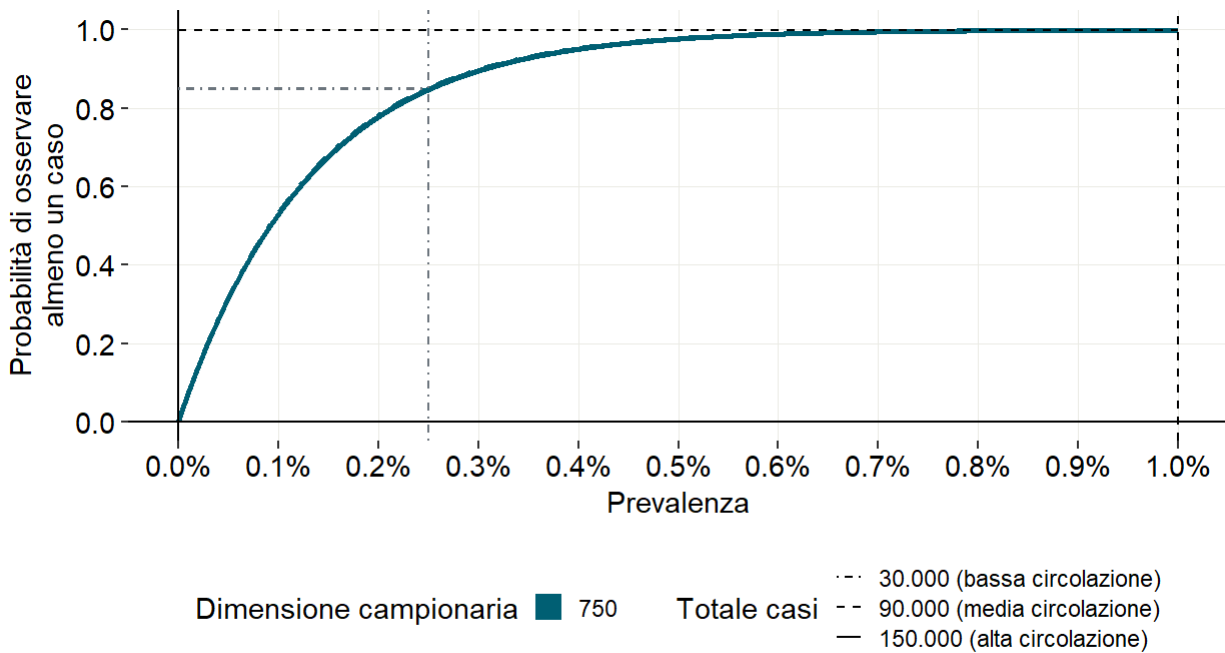
**Figura 1 e Tabella 1.** Probabilità di identificare almeno un campione con una specifica variante con un campione settimanale di 500 sequenze per diverse dimensioni campionarie al variare della prevalenza attesa e del livello di diffusione dell'epidemia.



Per popolazioni (casi settimanali) più grandi di 100.000 le curve sono sostanzialmente sovrapponibili a quella generata da una popolazione pari a 100.000

Dimensione campionaria	Totale casi (numerosità popolazione)	Prevalenza	Probabilità di osservare almeno un caso
500	30.000	0.25%	0,717
500	30.000	1%	0,994
500	90.000	0.25%	0,715
500	90.000	1%	0,994
500	150.000	0.25%	0,715
500	150.000	1%	0,993

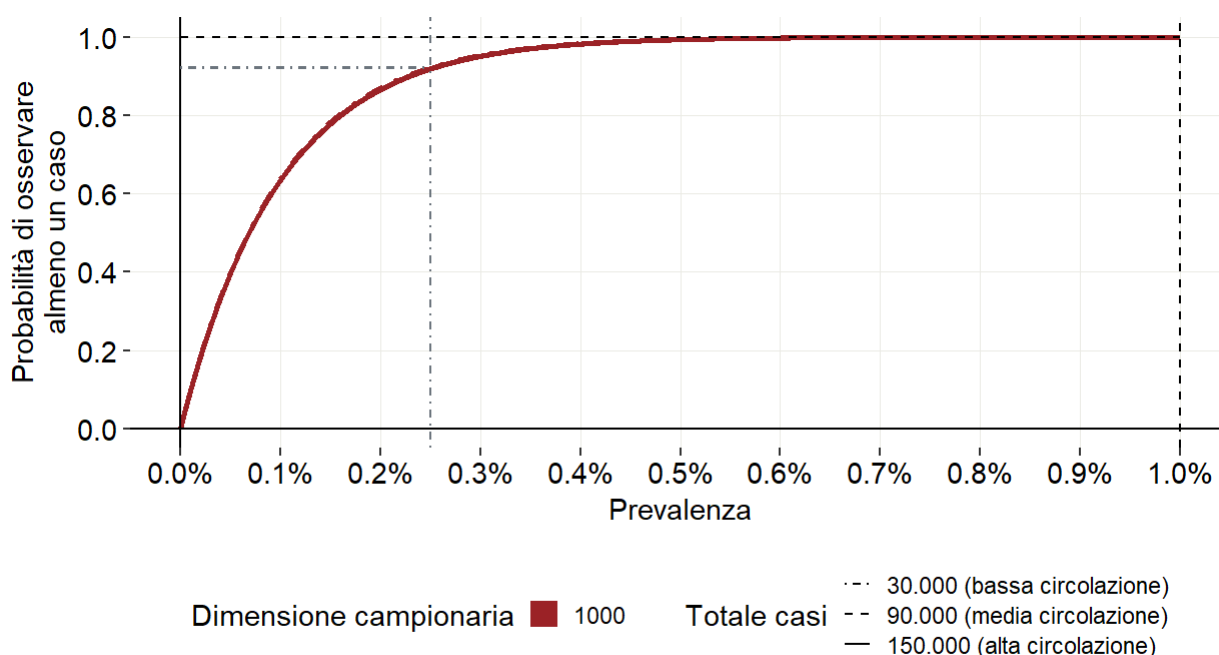
**Figura 2 e Tabella 2.** Probabilità di identificare almeno un campione con una specifica variante mediante un campione settimanale di 750 sequenze per diverse dimensioni campionarie al variare della prevalenza attesa e del livello di diffusione dell'epidemia.



Per popolazioni (casi settimanali) più grandi di 100.000 le curve sono sostanzialmente sovrapponibili a quella generata da una popolazione pari a 100.000

Dimensione campionaria	Totale casi (numerosità popolazione)	Prevalenza	Probabilità di osservare almeno un caso
750	30.000	0.25%	0,851
750	30.000	1%	1,000
750	90.000	0.25%	0,848
750	90.000	1%	0,999
750	150.000	0.25%	0,848
750	150.000	1%	0,999

**Figura 3 e Tabella 3.** Probabilità di identificare almeno un campione con una specifica variante con un campione settimanale di 1000 sequenze per diverse dimensioni campionarie al variare della prevalenza attesa e del livello di diffusione dell'epidemia.



Per popolazioni (casi settimanali) più grandi di 100.000 le curve sono sostanzialmente sovrapponibili a quella generata da una popolazione pari a 100.000

Dimensione campionaria	Totale casi (numerosità popolazione)	Prevalenza	Probabilità di osservare almeno un caso
1.000	30.000	0.25%	0,922
1.000	30.000	1%	1,000
1.000	90.000	0.25%	0,919
1.000	90.000	1%	1,000
1.000	150.000	0.25%	0,919
1.000	150.000	1%	1,000

Si osserva che con le diverse dimensioni campionarie settimanali qui ipotizzate si ha una probabilità maggiore del 99% di identificare almeno un caso di variante con una prevalenza sul territorio italiano di almeno l'1%. La probabilità aumenta all'aumentare della dimensione campionaria e al diminuire del livello di diffusione, anche se questi effetti sembrano sostanzialmente trascurabili con una prevalenza dell'1%. Per varianti con una prevalenza attesa di 0.25%, si apprezzano differenze di probabilità di identificazione tra le diverse dimensioni campionarie qui ipotizzate mentre l'effetto del livello di diffusione sembra essere molto più trascurabile. Si sottolinea che con un campione di 1.000 sequenze settimanali, indipendentemente dai livelli di circolazione ipotizzati, la probabilità di intercettare varianti/mutazioni con prevalenza di almeno lo 0,25% è >90%, contro circa una probabilità di 85% con 750 sequenze e di poco più del 70% con 500 sequenze. Tali risultati sono confermati sia con livelli di circolazione più bassi (fino a 5 casi/100.000 abitanti per settimana) che più elevati.

La suddivisione per Regione/PA del numero dei campioni da sequenziare è basata sulla popolazione residente in ciascuna Regione/PA. In tabella 4 si riporta la numerosità dei campioni da genotipizzare per Regione/PA per le tre diverse dimensioni campionarie ipotizzate.

**Tabella 4.** Numero di campioni settimanali da sequenziare in ciascuna Regione/PA.

Regione/PA	Popolazione 2021	% sequenze settimanali proporzionali a popolazione	dimensione campionaria settimanale ipotizzata		
Abruzzo	1.281.012	2.2	11	16	22
Basilicata	545.130	0.9	5	7	9
Calabria	1.860.601	3.1	16	24	31
Campania	5.624.260	9.5	47	71	95
E. Romagna	4.438.937	7.5	37	56	75
FVG	1.201.510	2.0	10	15	20
Lazio	5.730.399	9.7	48	73	97
Liguria	1.518.495	2.6	13	19	26
Lombardia	9.981.554	16.9	84	126	169
Marche	1.498.236	2.5	13	19	25
Molise	294.294	0.5	2	4	5
P.A. Bolzano	534.912	0.9	5	7	9
P.A. Trento	542.166	0.9	5	7	9
Piemonte	4.274.945	7.2	36	54	72
Puglia	3.933.777	6.6	33	50	66
Sardegna	1.590.044	2.7	13	20	27
Sicilia	4.833.705	8.2	41	61	82
Toscana	3.692.865	6.2	31	47	62
Umbria	865.452	1.5	7	11	15
V. d'Aosta	124.089	0.2	1	2	2
Veneto	4.869.830	8.2	41	62	82
<b>TOTALE</b>	<b>59.236.213</b>		<b>500</b>	<b>750</b>	<b>1000</b>

**Nota:** Le dimensioni campionarie qui stratificate a livello regionale sono basate su una ipotesi di distribuzione omogenea del virus sul territorio nazionale; nel caso di assenza o quasi assenza di



*circolazione in qualche regione, le sequenze da effettuare dovranno essere adattate al valore più prossimo alla disponibilità dei campioni raccolti.*

Inoltre si specifica di:

- eseguire il sequenziamento entro 7 gg lavorativi, ponendo particolare attenzione alla necessità di **completare in tempo reale le informazioni relative alle sequenze depositate in I-Co-Gen con i principali dati relativi al caso oltre alla data di prelievo del tampone**;
- **si raccomanda nel frattempo di inserire i risultati relativi alle varianti di interesse in sanità pubblica nella piattaforma della sorveglianza integrata COVID-19 dell'ISS attualmente in uso**;
- **caricare tutte le sequenze effettuate sulla piattaforma I-Co-Gen autorizzando l'invio automatico alla piattaforma internazionale GISAID.**
- **La settimana in cui ricadrà la flash survey non si sovrapporrà al campionamento relativo all'obiettivo specifico 2.**
- Le attività di sequenziamento verranno rimborsate da ISS in base alle risorse disponibili grazie al finanziamento previsto dall'art. 34bis della Legge 23 Luglio 2021 n. 106 (Modificazione apportate in sede di conversione al Decreto Legge 25 maggio 2021 n. 73).

## APPENDICE

### Avvertenze per i caricamenti batch su I-Co-Gen

- Qualche minuto dopo aver eseguito il caricamento in batch. si può ispezionare l'ultimo file rapporto\_[data] nella cartella elaborati, per vedere se ci sono stati problemi con il batch.
- Si può controllare dentro l'interfaccia web della piattaforma I-Co-Gen se i campioni sono stati creati e le elaborazioni avviate (in momenti di molteplici caricamenti. ci possono essere dei ritardi in questa fase).
- Si riceve una mail di fine analisi per ogni campione inviato.

Collegamento al manuale: <https://aries.iss.it/static/images/IRIDA21-ICoGen.pdf>.

Si prega di contattare l'indirizzo mail [helpdesk.icogen@iss.it](mailto:helpdesk.icogen@iss.it) per eventuali problemi inerenti al caricamento delle sequenze.