COVID-19: perché si effettuano due o più ”tamponi”? Quanti test sono necessari a ritenere sano o guarito un individuo?

Max Pierini*

May 18, 2020

1 Introduzione

La *prevalenza* di una malattia infettiva corrisponde alla percentuale di popolazione affetta ovvero alla probabilità che un individuo di tale popolazione scelto a caso sia affetto dalla malattia [1].

Dal punto di vista bayesiano la prevalenza corrisponde alla probabilità **a priori** $P(M)$ di essere malato. La probabilità dunque **a priori** di non essere affetto dalla malattia $P(\overline{M})$ sarà pari a $1 - P(M)$.

Durante un'epidemia la prevalenza, per ovvi motivi, si modifica: una percentuale nettamente superiore di popolazione è affetta dalla malattia ovvero la probabilità di essere malati aumenta.

Un test diagnostico qualitativo (come il tampone naso-faringeo) ha due possibilità di esito: positivo $\oplus$ oppure negativo $\ominus$.

La probabilità di ottenere un test positivo per un individuo malato rappresenta la *sensibilità* del test [1]

$$P(\oplus|M) = \text{sensibilità} \tag{1}$$

la situazione ideale sarebbe quindi $P(\oplus|M) = 1$ ovvero ottenere 100% di test positivi su tutti i malati.

La probabilità di ottenere un test negativo per un individuo sano (non affetto dalla malattia in questione) rappresenta invece la *specificità* del test [1]

$$P(\ominus|\overline{M}) = \text{specificità} \tag{2}$$

la situazione ideale sarebbe quindi $P(\ominus|\overline{M}) = 1$ ovvero ottenere 100% di test negativi su tutti i sani.

Prevalenza, sensibilità e specificità influiscono sulla probabilità **a posteriori** di essere malato o sano in seguito a risultato positivo o negativo del test.

*info@maxpierini.it

2 Test e Teorema di Bayes

Il teorema di Bayes (3) ci spiega il motivo e ci indica come calcolare queste probabilità.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3}$$

dove $P(B)$ può essere calcolato come

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) \tag{4}$$

Applichiamo il teorema di Bayes per calcolare la probabilità di essere malato in seguito a risultato positivo di un test diagnostico:

$$P(M|\oplus) = \frac{P(\oplus|M)P(M)}{P(\oplus)} \tag{5}$$

Al numeratore conosciamo tutti i termini: il primo è la sensibilità (1) e il secondo, se dell'individuo non si sa nulla ed è scelto a caso, è la prevalenza (anche nel corso di un'epidemia). Al denominatore invece troviamo la probabilità a priori di avere un test positivo, che non conosciamo direttamente, ma possiamo calcolare grazie alla (4).

$$P(\oplus) = P(\oplus|M)P(M) + P(\oplus|\overline{M})P(\overline{M}) \tag{6}$$

dove $P(\oplus|\overline{M})$ ovvero la probabilità di avere un test positivo su individuo sano è $P(\oplus|\overline{M}) = 1 - P(\ominus|\overline{M})$ e, come già sappiamo, $P(\overline{M}) = 1 - P(M)$. Dunque la (5) diventa:

$$P(M|\oplus) = \frac{P(\oplus|M)P(M)}{P(\oplus|M)P(M) + (1 - P(\ominus|\overline{M}))(1 - P(M))} \tag{7}$$

ovvero

$$P(M|\oplus) = \frac{\text{sensibilità} \cdot \text{prevalenza}}{\text{sensibilità} \cdot \text{prevalenza} + (1 - \text{specificità}) \cdot (1 - \text{prevalenza})} \tag{8}$$

Applicando invece il teorema di Bayes per calcolare la probabilità di non essere malato in seguito ad un test negativo otteniamo

$$P(\overline{M}|\ominus) = \frac{P(\ominus|\overline{M})P(\overline{M})}{P(\ominus)} \tag{9}$$

anche in questo caso conosciamo il numeratore (specificità e 1 - prevalenza). Il denominatore può essere espresso come

$$P(\ominus) = P(\ominus|\overline{M})P(\overline{M}) + P(\ominus|M)P(M) \tag{10}$$

dove $P(\ominus|M) = 1 - P(\oplus|M)$. Sostituendo nella (9) otteniamo quindi

$$P(\overline{M}|\ominus) = \frac{P(\ominus|\overline{M})(1 - P(M))}{P(\ominus|\overline{M})(1 - P(M)) + (1 - P(\oplus|M))P(M)} \tag{11}$$

ovvero

$$P(\overline{M}|\ominus) = \frac{\text{specificità} \cdot (1 - \text{prevalenza})}{\text{specificità} \cdot (1 - \text{prevalenza}) + (1 - \text{sensibilità}) \cdot \text{prevalenza}} \tag{12}$$

Abbiamo dunque ora tutti gli strumenti necessari per calcolare la probabilità a posteriori di essere malati (o sani) in seguito ad esito del test diagnostico.

3 Test quasi ideale

Prendiamo a titolo d'esempio un buon test ("quasi" ideale) con specificità e sensibilità al 99.5%, supponendo un prevalenza del 10%:

$$P(M) = .1 \quad , \quad P(\oplus|M) = .995 \quad , \quad P(\ominus|\overline{M}) = .995 \tag{13}$$

alle ascisse metteremo la probabilità a priori di essere malato da 0 a 1 e alle ordinate la probabilità a posteriori di essere malato dato il risultato del test (figura 1). La linea blu rappresenta la probabilità a posteriori di essere malato in seguito a test positivo $P(M|\oplus)$ mentre la linea rossa la probabilità a posteriori di essere malato in seguito a test negativo $1 - P(\overline{M}|\ominus)$. La linea tratteggiata verticale indica la prevalenza $P(M)$.

Poniamo $p_\ominus = .05$ come limite diagnostico per un individuo sano, ovvero una probabilità di essere malato $P(M|\ominus) < .05$ (5%) è sufficiente a dichiarare sano un individuo in seguito a test negativo.

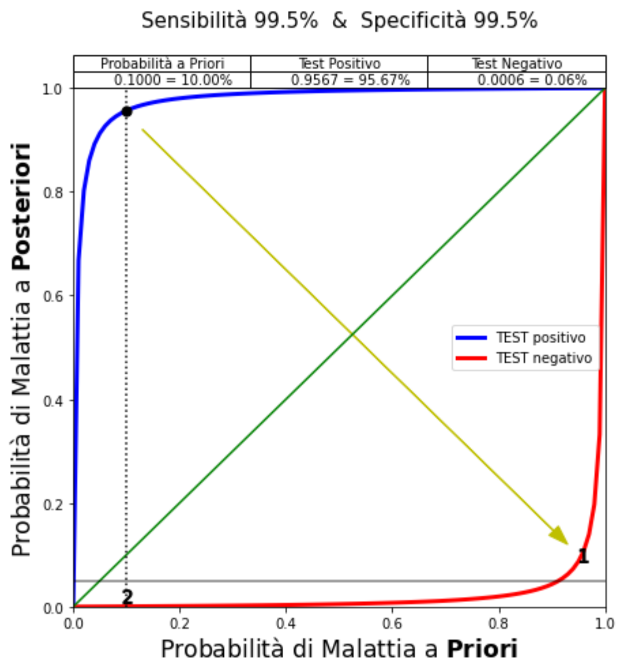

Figure 1: Esempio di test "quasi" ideale.

$$P(M|\ominus) = 0.0006$$

Notiamo come $P(M|\ominus)$, ovvero la probabilità a posteriori di essere malato dopo un test negativo, sia nettamente inferiore a $p_\ominus$. Dunque è sufficiente un solo test negativo per ritenere sano un individuo di cui non si abbiano precedenti informazioni (la sua probabilità a priori era solo la prevalenza).

$P(M|\oplus)$, ovvero la probabilità a posteriori di essere malato dopo un test positivo per un individuo nelle stesse condizioni, risulta superiore al 95% e diventa la nuova *probabilità a priori* per quel soggetto in caso di un test successivo (punto 1 sulla curva rossa). Si può calcolare che $P(M|\ominus)$ laddove la probabilità a priori del soggetto non sia la prevalenza ma $P(M) = P(M|\oplus) = 95.67\%$ è superiore al 5% infatti

$$P(M|\ominus) = 1 - \frac{P(\ominus|\overline{M})(1 - P(M|\oplus))}{P(\ominus|\overline{M})(1 - P(M|\oplus)) + (1 - P(\oplus|M))P(M|\oplus)} = 0.1000$$

da cui ricaviamo che un test negativo in un individuo che abbia precedentemente ricevuto un test positivo, non è sufficiente a ritenerlo sano ma ne sarà necessario un altro. Si può calcolare che la probabilità massima a priori per la quale sia sufficiente un solo test negativo per ritenere sano l'individuo (alla condizioni poste) è pari alla risoluzione per $P(M)$ di

$$p_\ominus = 1 - P(\overline{M}|\ominus) \tag{14}$$

ovvero

$$p_{\ominus} = 1 - \frac{P(\ominus|\overline{M})(1-P(M))}{P(\ominus|\overline{M})(1-P(M)) + (1-P(\oplus|M))P(M)} \quad (15)$$

da cui si ricava facilmente che $P(M)_{max} = 0.9128$.

Al di sotto di questa probabilità a priori, con questi valori di sensibilità e specificità, servirà un solo test negativo per escludere, stabilito $p_{\ominus} < .05$, che l'individuo sia malato.

Notiamo anche come, a parità di sensibilità e specificità, la prevalenza (ovvero la probabilità di malattia a priori) influisca sulla probabilità di malattia a posteriori. Ad esempio per una malattia rara come la Sindrome di Cushing endogena con prevalenza in Europa di un 1 caso du 26'000 [2] ovvero

$P(M) = 0.000038 = 0.0038\%$, la probabilità di malattia a posteriori in seguito a test positivo (con sensibilità e specificità di .995) sia

$P(M|\oplus) = 0.0076 = 0.76\%$ dunque molto bassa anche se notevolmente superiore rispetto alla probabilità a priori.

In questo caso, a parità di sensibilità, servirebbe un test altamente specifico, ad esempio con $P(\ominus|\overline{M}) = .9999$ col quale si otterrebbe una probabilità a posteriori per test positivo pari a

$P(M|\oplus) = 0.2768 = 27.68\%$

4 Sensibilità e Specificità

Mantenendo fissa la sensibilità a .995, variamo la specificità e viceversa

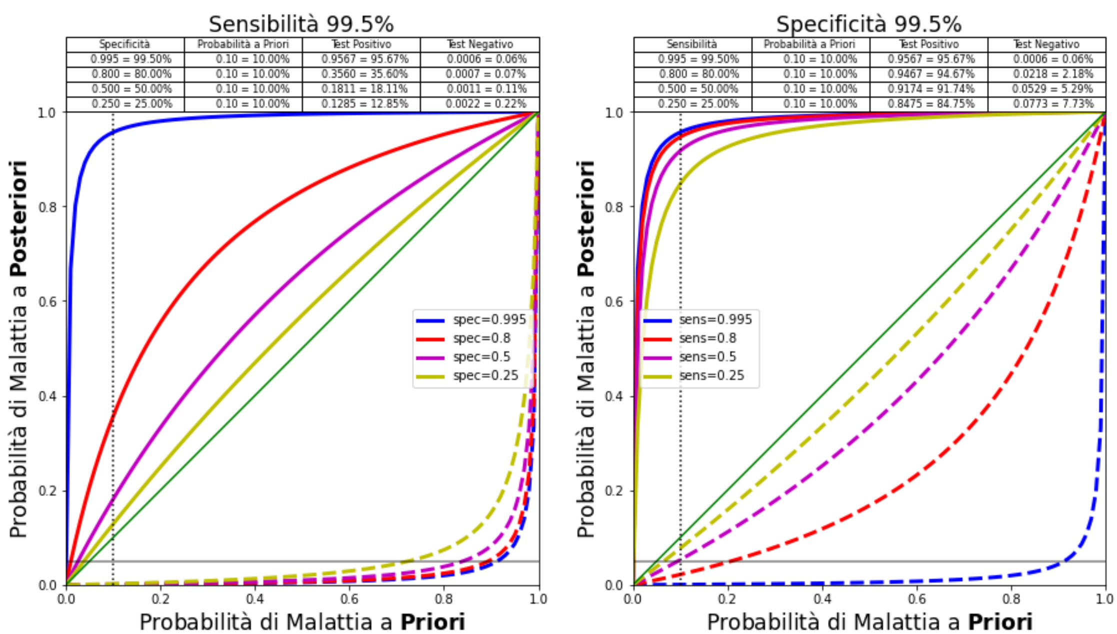

Figure 2: Rapporto tra specificità e sensibilità. Le linee piene indicano $P(M|\oplus)$, le linee tratteggiate $P(M|\ominus)$.

Notiamo come valori differenti di specificità a parità di sensibilità (o viceversa) abbiano effetti sia sulla probabibilità a posteriori per test positivo che per test negativo sebbene

- variazioni nella specificità $P(\ominus|\overline{M})$ abbiano maggior effetto su $P(M|\oplus)$
- variazioni nella sensibilità $P(\oplus|M)$ abbiano maggior effetto su $P(\overline{M}|\ominus)$

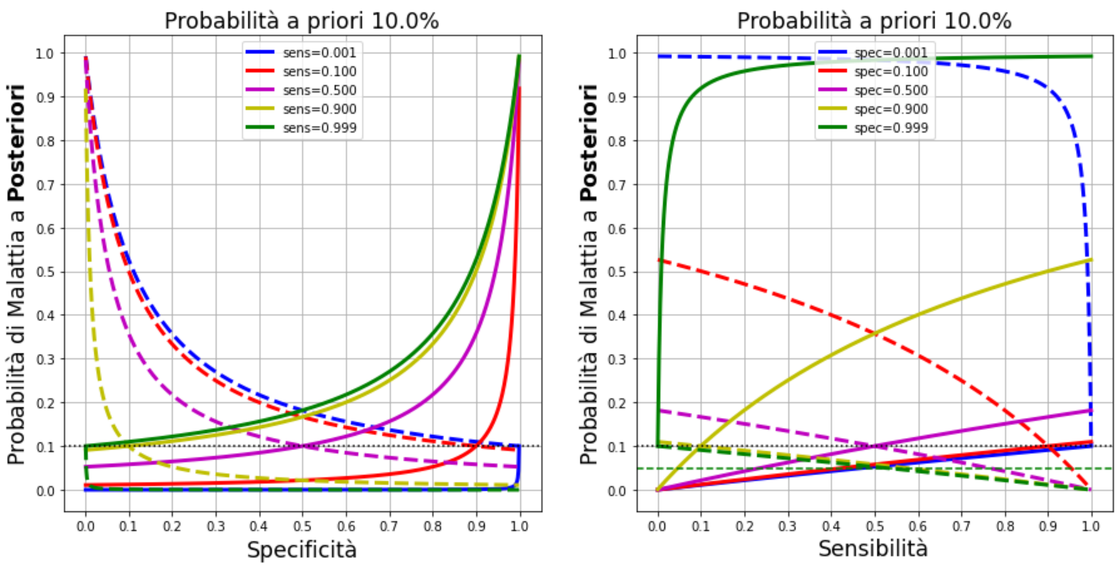

Figure 3: Rapporto tra specificità e sensibilità. Le linee piene indicano $P(M|\oplus)$, le linee tratteggiate $P(M|\ominus)$. Si nota come sia necessario che i test abbiano determinate caratteristiche per poter essere utili.

Si nota come sensibilità e specificità siano strettamente interconnesse. In particolare, data una probabilità di malattia a priori $P(M)$, si vuole che un test abbia

- una probabilità di malattia a posteriori dato test positivo almeno $P(M|\oplus) > P(M)$
- una probabilità di malattia a posteriori dato test negativo almeno $P(M|\ominus) < P(M)$

Per trovare i minimi requisiti di un test diagnostico dunque, data la probabilità di malattia a priori, si può risolvere il sistema a due equazioni

$$\begin{cases} P(M|\oplus) > P(M) \\ P(M|\ominus) < P(M) \end{cases}$$

ma avendo stabilito $p_\ominus < .05$ per $P(M|\ominus)$ si può anche assumere $p_\oplus > .5$ per $P(M|\oplus)$ ovvero che la probabilità di malattia a posteriori dato test positivo sia almeno superiore al 50%

$$\begin{cases} P(M|\oplus) > p_\oplus \\ P(M|\ominus) < p_\ominus \end{cases}$$

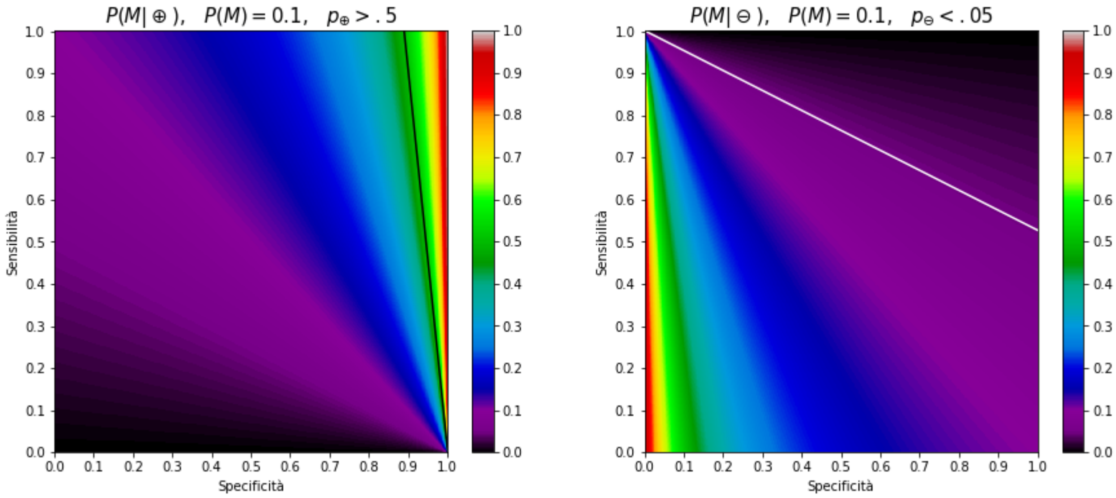

Figure 4: Requisiti per $P(M) = .1$. L'area a destra della linea nera nell'immagine di sinistra è il requisito minimo per $P(M|\oplus) > .5$. L'area a destra della linea bianca nell'immagine di destra è il requisito minimo per $P(M|\ominus) < .05$.

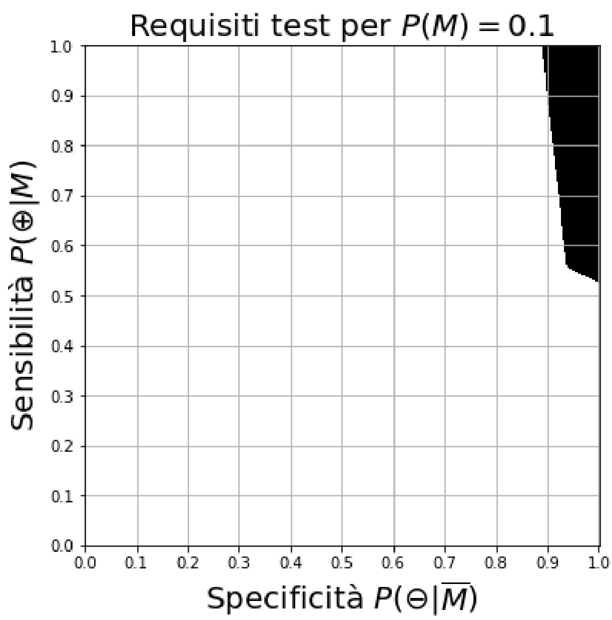

Figure 5: L'area nera indica i requisiti necessari $P(M|\oplus) > .5$ e $P(M|\ominus) < .05$ per un test con $P(M) = .1$.

5 COVID-19 e tamponi naso-faringei

Studi recenti suggeriscono che i tamponi naso-faringei usati per la diagnosi di COVID-19 (RT-PCR SARS-CoV-2 RNA test) abbiano sensibilità $P(\oplus|M) = 0.777$ e specificità di $P(\ominus|\overline{M}) = 0.988$ [3] e che la prevalenza di COVID-19 in Italia (in fase pandemica) sia circa pari a $P(M) = 0.13$ [4] [5] [6].

Verifichiamo che il test abbia i requisiti richiesti (figura 6): il punto giallo indica i parametri di sensibilità e specificità dei test in esame per $P(M) = .13$, i test hanno quindi caratteristiche accettabili.

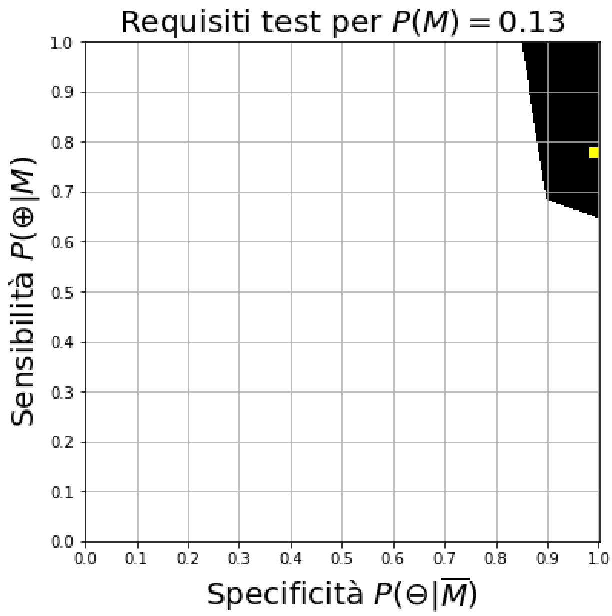

Figure 6: L'area nera indica i requisiti necessari $P(M|\oplus) > .5$ e $P(M|\ominus) < .05$ per un test con $P(M) = .13$. Il punto giallo indica i parametri di sensibilità e specificità dei test RT-PCR SARS-CoV-2 RNA test per COVID-19: sensibilità $P(\oplus|M) = 0.777$ e specificità di $P(\ominus|\overline{M}) = 0.988$

Vediamo come si modificano le curve delle probabilità a posteriori impostando questi tre parametri (figura 7).

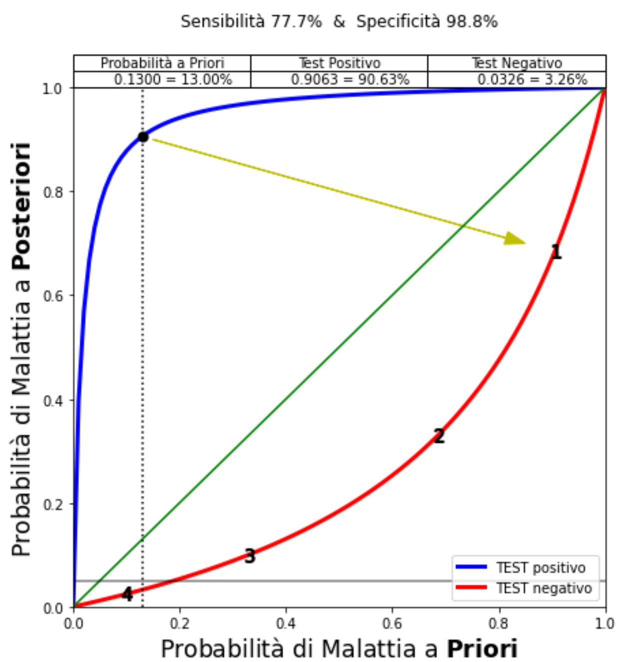

Figure 7: Probabilità di malattia COVID-19 a posteriori per test RT-PCR SARS-CoV-2 RNA.

Anche in questo caso dunque è sufficiente un test negativo per ritenere sano un soggetto su cui non si abbiano precedenti informazioni.

$$P(M|\ominus) = 0.0326 < p_\ominus = 0.05$$

Se invece un soggetto risultasse positivo (punto nero in figura 7), la sua probabilità di essere malato passerebbe da 0.13 (prevalenza stimata di COVID-19 in Italia in fase pandemica) a

$$P(M|\oplus) = 0.9063 > p_\oplus = .5$$

Vediamo quindi la probabilità a posteriori per test negativo $P(M|\ominus)$ con la probabilità a priori di un soggetto precedentemente risultato positivo (punto 1 in figura 7)

$$P(M|\ominus)_1 = 1 - \frac{P(\ominus|\overline{M})(1-0.9063)}{P(\ominus|\overline{M})(1-0.9063) + (1-P(\oplus|M))0.9063} = 0.6859$$

Dunque un tampone negativo non è sufficiente, il soggetto ha ancora un'elevata probabilità di essere infetto, supponiamo che ne venga fatto un secondo (punto 2 in figura 7)

$$P(M|\ominus)_1 = 1 - \frac{P(\ominus|\overline{M})(1-0.6859)}{P(\ominus|\overline{M})(1-0.6859) + (1-P(\oplus|M))0.6859} = 0.3302$$

Ancora superiore, 33.02% di probabilità di essere infetto non è accettabile per ritenere guarito il paziente. Effettuiamo un terzo test (punto 3 in figura 7)

$$P(M|\ominus)_1 = 1 - \frac{P(\ominus|\overline{M})(1-0.3302)}{P(\ominus|\overline{M})(1-0.3302) + (1-P(\oplus|M))0.3302} = 0.1001$$

10.01% nuovamente superiore a $p_\ominus$, dunque effettuiamo un quarto test (punto 4 in figura 7)

$$P(M|\ominus)_1 = 1 - \frac{P(\ominus|\overline{M})(1-0.1001)}{P(\ominus|\overline{M})(1-0.1001) + (1-P(\oplus|M))0.1001} = 0.0245$$

2.45% di probabilità di essere infetto è un valore che possiamo ritenere accettabile per considerarlo sano. Servirebbero quindi quattro tamponi nasofaringei negativi per escludere la possibilità che un soggetto precedentemente risultato positivo per COVID-19 sia considerabile attualmente guarito $P(M|\ominus) < .05$.

Calcolando in questo caso la massima probabilità a priori $P(M)_{max}$ data la quale sia sufficiente un solo tampone negativo per considerare sano il soggetto, otteniamo

$$P(M)_{max} = 0.1891$$

ovvero per probabilità a priori $P(M^*) > 18.91\%$ servirà sicuramente più di un tampone.

References

[1] M. Porta, *A dictionary of epidemiology.* Oxford university press, 2014.

[2] "Sindrome di cushing e malattia di cushing: cosa sono?" [Online]. Available: https://www.osservatoriomalattierare.it/errore/38-descrizioni/sezioni/5838-sindrome-di-cushing-e-malattia-di-cuschin-cosa-sono

[3] N. S. Padhye, "Reconstructed diagnostic sensitivity and specificity of the rt-pcr test for covid-19," *medRxiv*, 2020.

[4] Z. Ceylan, "Estimation of covid-19 prevalence in italy, spain, and france," *Science of The Total Environment*, p. 138817, 2020.

[5] M. A. Vollmer, S. Mishra, H. J. T. Unwin, A. Gandy, T. A. Mellan, V. Bradley, H. Zhu, H. Coupland, I. Hawryluk, M. Hutchinson *et al.*, "A sub-national analysis of the rate of transmission of covid-19 in italy," *medRxiv*, 2020.

[6] S. Flaxman, S. Mishra, A. Gandy, H. Unwin, H. Coupland, T. Mellan, H. Zhu, T. Berah, J. Eaton, P. Perez Guzman *et al.*, "Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on covid-19 in 11 european countries," 2020.