



Universiteit
Leiden

The Netherlands

Prediction of contralateral breast cancer: statistical aspects and prediction performance

Giardiello, D.

Citation

Giardiello, D. (2022, September 8). *Prediction of contralateral breast cancer: statistical aspects and prediction performance*. Retrieved from <https://hdl.handle.net/1887/3455362>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3455362>

Note: To cite this publication please use the final published version (if applicable).

RIASSUNTO IN ITALIANO

Il tumore al seno o alla mammella (chiamato anche carcinoma mammario invasivo o infiltrante) è il tumore più frequente tra le donne di tutto il mondo. Questa patologia, sebbene diagnosticata ad un numero sempre crescente di donne, nel corso degli anni è diventata meno letale. Mentre nel 1960 la sopravvivenza a 10 anni dalla diagnosi era di circa 4 donne su 10, attualmente più di 8 donne su 10 riescono a sopravvivere in Europa a questa patologia. Questo sostanziale miglioramento è dovuto principalmente alla diagnosi precoce, resa possibile dagli esami mammografici periodici, e alle modalità di trattamento sempre più avanzate e personalizzate. Se da un lato l'anticipazione diagnostica e la sopravvivenza a questa malattia sono migliorate nel tempo, dall'altro un numero sempre maggiore di donne è potenzialmente a rischio di sviluppare un secondo tumore al seno opposto, conosciuto anche come tumore al seno controlaterale. Il tumore al seno controlaterale è un altro tumore primario (e non una recidiva del primo tumore alla mammella) ed è il più frequente secondo tumore che può essere diagnosticato tra le donne con un tumore alla mammella primario, rappresentando circa il 50% di tutti i secondi tumori. In media, circa 4 o 5 donne su 100 sviluppano un tumore anche nel rimanente seno sano entro 10 anni dalla diagnosi del primo, il che comporta una riduzione di durata della sopravvivenza rispetto alle donne con una diagnosi di un singolo tumore. Per questa ragione, donne con un tumore alla mammella primario o ad alto rischio di sviluppare un tumore alla mammella anche nel seno sano possono scegliere di effettuare la mastectomia preventiva controlaterale, ovvero la rimozione del seno sano dopo che l'altro seno è stato colpito da un tumore. In donne con un elevato rischio di sviluppare un tumore al seno controlaterale, come quelle con una chiara predisposizione genetica alla malattia (ad esempio portatrici di una mutazione dei geni *BRCA1* e *BRCA2* o nel gene *CHEK2*), la mastectomia preventiva è una delle strategie preventive più considerate tra medici e pazienti. Nonostante l'efficacia della mastectomia preventiva non sia chiaramente provata soprattutto in assenza di predisposizioni genetiche, un numero crescente di donne a basso rischio di sviluppare un tumore al seno controlaterale decidono di sottoporsi alla mastectomia preventiva. Una stima del rischio oggettiva e basata sulle caratteristiche individuali delle pazienti può guidare la scelta riguardante le misure preventive da adottare nella pratica clinica. Ad esempio, pazienti in cui viene riconosciuto un rischio elevato possono scegliere, dopo opportuna valutazione e discussione con una équipe di professionisti, di rimuovere completamente il seno sano attraverso la mastectomia, oppure possono optare di effettuare ulteriori trattamenti o screening mammari personalizzati. Al contrario, pazienti con basso rischio possono evitare di rimuovere il seno sano salvaguardandosi da eventuali effetti collaterali dovuti all'intervento chirurgico e potenziali impatti di natura psicologica. L'obiettivo principale di questa tesi di dottorato è stato quello di sviluppare un modello statistico che stimi il rischio di sviluppare un tumore al seno

controlaterale, di valutarne la sua capacità predittiva e di stabilire la sua potenziale utilità nel supportare e migliorare le scelte da intraprendere nella pratica clinica tra medico e paziente (**capitolo 1**).

Nel **capitolo 2** abbiamo sviluppato e validato un modello di rischio per il tumore al seno controlaterale, da noi denominato PredictCBC. I dati, di popolazione o ricavati da 20 studi clinici eseguiti principalmente in Europa, Stati Uniti d'America e Australia, provengono complessivamente da più di 100,000 donne con diagnosi di carcinoma mammario diagnosticato tra il 1990 e il 2013. Il modello PredictCBC fornisce una stima individualizzata del rischio di tumore al seno controlaterale a 5 e a 10 anni dalla diagnosi del carcinoma mammario primario, utilizzando informazioni individuali di ogni singola paziente quali: le caratteristiche del primo tumore al seno e dei corrispondenti trattamenti; la familiarità, intesa come una precedente diagnosi di tumore al seno tra i parenti di primo grado; la presenza di mutazione nei geni *BRCA1/2*. Abbiamo evidenziato come la mastectomia (preventiva) del seno sano potrebbe essere non giustificata perfino in alcune pazienti portatrici della mutazione nei geni *BRCA1/2*, specialmente tra coloro che hanno altre caratteristiche favorevoli. D'altro canto, misure che prevenivano o possano diagnosticare precocemente il tumore al secondo seno come mammografie più frequenti e personalizzate potrebbero essere utili anche alle donne senza alcuna predisposizione genetica, qualora il rischio sia elevato per la presenza di altre caratteristiche sfavorevoli.

Nel **capitolo 3** abbiamo valutato la capacità previsionale del modello PredictCBC con quella di altri due modelli, denominati "la formula di Manchester" e il modello "CBCrisk", anch'essi sviluppati per stimare il rischio di tumore al seno controlaterale e attualmente disponibili in letteratura. La "formula di Manchester" è una formula euristica che stima il rischio di sviluppare il tumore nel corso della vita nel rimanente seno sano ed è stato messo a punto usando informazioni ricavate da una revisione sistematica precedentemente pubblicata. Lo strumento "CBCrisk", invece, è stato sviluppato utilizzando dei dati di studi provenienti da una casistica raccolta negli Stati Uniti. La comparazione fra questi tre strumenti è stata effettuata sui 20 studi precedentemente accennati in merito allo sviluppo del nostro modello. Come risultato, abbiamo verificato che mediamente tutti e tre gli strumenti forniscono una accuratezza previsiva moderata. Abbiamo inoltre osservato una considerevole eterogeneità delle prestazioni previsive tra i tre diversi strumenti in relazione alle differenze tra i vari studi per caratteristiche delle pazienti incluse e per *incidenza* di malattia. Da tali evidenze abbiamo dedotto che maggiori conoscenze biologiche, una migliore caratterizzazione clinica della malattia e più dettagliate informazioni genetiche potrebbero portare a prevedere meglio il rischio di tumore al seno controlaterale e conseguentemente a migliorare le decisioni cliniche. Anche la presenza di informazioni cliniche incomplete può compromettere la capacità

previsionale dei modelli, ed è quindi auspicabile che in futuro la loro raccolta avvenga con maggiore accuratezza.

Nel **capitolo 4**, abbiamo esteso il modello PredictCBC sviluppato nel **capitolo 2** includendo ulteriori importanti informazioni associate alla patologia come informazioni relative allo stile di vita (ad esempio, l'indice di massa corporea), fattori riproduttivi (ad esempio, il numero totale di nascite dopo la gestazione), ed informazioni genetiche come la presenza della mutazione nel gene *CHEK2* e il valore del *rischio poligenico*. È stato sviluppato e validato un modello statistico più avanzato e aggiornato del precedente (denominato PredictCBC-2.0) utilizzando dati più aggiornati ed estesi che hanno incluso più di 200,000 pazienti. Abbiamo dimostrato che le ulteriori informazioni genetiche incluse nel nuovo modello forniscono una stima del rischio più accurata e potrebbero quindi migliorare la scelta delle strategie preventive attualmente previste.

Nel **capitolo 5**, abbiamo comparato il rischio di tumore al seno controlaterale tra le pazienti diagnosticate con il carcinoma mammario (invasivo o chiamato anche infiltrante) e il carcinoma duttale in situ. Quest'ultimo tipo di carcinoma è considerato una precancerosi, ovvero un fenomeno che potrebbe predisporre allo sviluppo di una forma invasiva. Utilizzando il registro tumori dei Paesi Bassi contenenti informazioni dettagliate riguardanti le pazienti, abbiamo stimato che il rischio di tumore al seno controlaterale è leggermente più alto tra le pazienti con il carcinoma duttale in situ rispetto alle pazienti con il carcinoma mammario invasivo. Questo leggero aumento del rischio di sviluppare un tumore al seno controlaterale tra le pazienti con il carcinoma duttale in situ potrebbe essere attribuito al fatto che nei Paesi Bassi queste pazienti non vengono trattate con terapie mirate (come la chemioterapia) che tendono in maniera primaria ad aumentare la probabilità di guarigione e a ridurre principalmente il rischio di potenziali recidive e metastasi, ma anche a contenere il rischio di un tumore al seno sano. Questo non implica che queste pazienti dovrebbero essere trattate con queste terapie mirate, soprattutto in virtù del fatto che il carcinoma mammario in situ non rappresenta un rischio per la vita e le terapie potrebbero avere effetti collaterali che superano i benefici. Una più dettagliata e oggettiva quantificazione del rischio di tumore al seno controlaterale potrebbe essere utile anche per queste pazienti per poter stabilire strategie di prevenzione più mirate.

Guide pratiche per la valutazione delle prestazioni di previsione dei modelli di rischio di sopravvivenza

L'ambito della ricerca scientifica che si occupa della stima del rischio ha come obiettivo quello di sviluppare modelli previsivi accurati che possono essere applicabili nella pratica clinica. Un modello per la previsione del rischio può essere sviluppato utilizzando la regressione, una tecnica statistica che cerca di comprendere un fenomeno utilizzando

una serie di fattori (chiamati anche predittori). Ad esempio, se il fenomeno di interesse è la comparsa di una patologia (come, ad esempio, il tumore al seno controlaterale), si potrebbe essere interessati a comprendere come questa patologia sia legata all'età. In questo caso, l'occorrenza di tumore al seno controlaterale rappresenta il fenomeno di interesse mentre l'età è il predittore con il quale proviamo a comprendere meglio questo fenomeno. In molti studi sul cancro, come ad esempio gli studi sul tumore al seno, i ricercatori sono interessati a studiare il verificarsi di un determinato evento in un prefissato periodo di tempo. Nel caso del tumore al seno controlaterale, il fenomeno di interesse è il tempo (misurato in anni) trascorso dalla diagnosi del primo tumore al seno al secondo in un prefissato arco temporale (ad esempio a 5 o 10 anni dalla diagnosi del primo tumore). Questo fenomeno configura un capitolo della statistica definito come "analisi della sopravvivenza", per il quale sono stati predisposti appropriati strumenti di elaborazione. Secondo tale approccio, quando si studia un fenomeno che potrebbe accadere nel corso del tempo in un determinato gruppo di individui, per alcuni individui quel determinato evento potrebbe non accadere in un prefissato periodo di tempo. In questo caso il tempo è definito come tempo censorizzato. Un altro scenario possibile si verifica quando un altro fenomeno possa precludere il verificarsi dell'evento oggetto di studio. Per esempio, se si è interessati a studiare il verificarsi del tumore al seno in un prefissato periodo di tempo dopo la diagnosi del primo tumore, alcune pazienti potrebbero purtroppo non sopravvivere (per causa diretta o una causa non legata alla patologia di interesse). In questo caso, in queste pazienti non si potrà mai diagnosticare il tumore nel rimanente seno sano dato che un altro fenomeno (chiamato anche evento o rischio competitivo) si verifica durante il corso della loro vita. I più comuni modelli di regressione per l'analisi della sopravvivenza in assenza o in presenza di fenomeni/rischi/eventi competitivi sono, rispettivamente, il modello di regressione di Cox e il modello di Fine e Gray. Questi modelli possono essere utilizzati per prevedere il verificarsi di un determinato evento di interesse in un prefissato periodo di tempo. Per esempio, prevedere il rischio di tumore al seno controlaterale nei successivi 5 o 10 anni dalla diagnosi del primo tumore al seno. Una volta che un modello di previsione del rischio è stato sviluppato, è fondamentale valutarne la sua capacità e accuratezza previsiva. In primo luogo, la capacità previsiva di un modello viene valutata nella stessa popolazione utilizzata per sviluppare tale modello. Questo processo è definito come validazione o valutazione interna. La validazione esterna, invece, si riferisce alla valutazione della capacità e accuratezza di un modello previsivo in una popolazione diversa rispetto a quella per il quale il modello è stato sviluppato. La valutazione esterna permette di comprendere le ragioni per cui un modello è o non è generalizzabile in altre popolazioni o in altri scenari. Se la capacità previsiva di un modello fosse sufficientemente accurata, il modello previsivo potrebbe essere utile al fine di migliorare i processi decisionali nella pratica clinica (come, ad esempio, la scelta riguardante la mastectomia controlaterale preventiva). Un ulteriore obiettivo di questa tesi è stato di fornire delle guide pratiche per

valutare la capacità previsiva di modelli di regressione per l'analisi della sopravvivenza in assenza o presenza di fenomeni/rischi/eventi competitivi utilizzando dati reali nell'ambito del tumore al seno.

Nel **capitolo 6**, abbiamo fornito una guida pratica che aiuti a comprendere come costruire un modello previsivo per fenomeni il cui studio rientra nell'analisi di sopravvivenza. Successivamente abbiamo descritto come valutare la capacità previsiva e come stimare la potenziale utilità di tali modelli nella pratica clinica, utilizzando dati riguardanti pazienti con tumore al seno, e abbiamo reso disponibile il corrispondente codice software per due linguaggi ampiamente utilizzati in ambito statistico, R e SAS.

Nel **capitolo 7**, abbiamo fornito una descrizione dettagliata e comprensibile delle attuali misure utilizzate per valutare la capacità previsiva di un modello di analisi della sopravvivenza in presenza di rischi competitivi. In particolare, abbiamo illustrato come sviluppare un modello per il calcolo del rischio di un fenomeno in presenza di eventi competitivi, come calcolare le diverse misure utili per valutarne capacità predittiva e utilità clinica, utilizzando dati reali di pazienti con tumore al seno e fornendo il corrispondente codice in R. Sia il **capitolo 6** che il **capitolo 7** sono stati scritti per conto di una iniziativa internazionale denominata STRATOS, che ha l'obiettivo di fornire guide e documentazioni accurate riguardanti argomenti di natura analitica ad un pubblico non specializzato.

Definizioni:

***Incidenza:** rappresenta la proporzione di individui che vengono colpiti da una determinata patologia in un determinato periodo di tempo. Misura il numero di nuovi casi nel periodo di tempo e individua il rischio (cioè la probabilità) che ha un individuo di contrarre una determinata patologia in quel periodo di tempo.*

***Rischio poligenico:** è una misura che quantifica il potenziale rischio di sviluppare delle patologie basate sui geni e si calcola combinando gli effetti di un gran numero di varianti genetiche presenti nel genoma di ogni singolo individuo.*