



La funzione dell'hazard

The hazard function

Enzo Coviello,¹ Guido Miccinesi,² Donella Puliti², Eugenio Paci² e il gruppo dello Studio IMPATTO³

¹ Unità di statistica ed epidemiologia ASL Bari

² Centro per lo studio e la prevenzione oncologica, Firenze

³ L'elenco dei componenti il gruppo di lavoro dello Studio IMPATTO è in ultima pagina

Corrispondenza: Enzo Coviello, via Imbriani 186, 70055 Minervino Murge (BA); tel fax 080 3357867; e-mail: enzo.coviello@alice.it

Riassunto

Per descrivere l'occorrenza di un evento negli studi di coorte è abituale calcolare il tasso di incidenza e presentare la funzione di sopravvivenza. Rispetto a questi stimatori, la rappresentazione grafica della funzione dell'hazard ha il vantaggio di mostrare le variazioni dell'occorrenza dell'evento nel corso del periodo di osservazione che spesso rivestono interesse non trascurabile. Inoltre, nel confronto di soggetti con differenti caratteristiche, la funzione dell'hazard è un utile supporto alla verifica degli assunti e all'interpretazione dei risultati ottenuti con un modello di Cox. Questo contributo presenta il metodo per la stima della funzione dell'hazard e un esempio di applicazione di tale metodo a un caso reale costituito dai dati di sopravvivenza dei casi di tumore della mammella raccolti nello studio IMPATTO per la valutazione dell'efficacia dei ser-

vizi di screening mammografico. È illustrato il collegamento fra questa funzione e gli stimatori più spesso applicati e evidenziato il suo ruolo nei modelli di regressione adoperati nell'analisi della sopravvivenza. Nell'esempio, dalla stima della funzione dell'hazard emerge che nel primo anno dalla diagnosi le pazienti colpite da tumore della mammella mostrano un tasso di mortalità più basso di quello che si registra negli anni successivi e che la differenza degli hazard fra casi invitati e quelli non invitati al test di screening resta approssimativamente costante durante il tempo di osservazione indagato, due importanti rilievi che dimostrano l'utilità dell'impiego di questa funzione nell'analisi degli studi di coorte.

(*Epidemiol Prev* 2007; 31(6): 346-51)

Parole chiave: funzione dell'hazard, studi di coorte, analisi di sopravvivenza.

Abstract

In cohort studies the event occurrence is usually described by the incidence rate and the survivor function. In comparison with these estimators the plot of the hazard function has the advantage to show the variations of the occurrence of the event along the period of observation, which often are important to be highlighted. Furthermore, when comparing individuals with different characteristics, the hazard function is a valuable support to check the assumption and to interpret the results of a Cox regression model. This paper illustrates the method for estimating the hazard function and an example is given from a real case by using the survival data of the breast cancers collected in the IMPACT study, aimed to detect the

efficacy of the mammographic screening program. The relationship between the usual estimators and the hazard function is shown and its role in the survival regression modelling is emphasized. In the example the estimate of the hazard function allows to point out that the mortality rate of breast cancer in the first year after the diagnosis is lower than later and that the difference between the hazards of the invited cases and those of the not invited cases is approximately constant along the whole 10 years follow up, two important remarks both demonstrating the usefulness of the application of this function in the analysis of cohort studies.

(*Epidemiol Prev* 2007; 31(6): 346-51)

Key words: hazard function, cohort studies, survival analysis.

Introduzione

Nell'ambito degli studi di coorte è abituale calcolare come misura di occorrenza di malattia il tasso di incidenza, ottenuto dal rapporto fra il numero di nuovi casi (eventi) rispetto agli anni-persona accumulati dai soggetti in osservazione per un determinato intervallo di tempo chiamato periodo a rischio.¹ L'equivalente di questa quantità nell'ambito dell'analisi della sopravvivenza è l'hazard o tasso istantaneo. Questo termine è utilizzato per indicare il concetto del rischio di evento al tempo t condizionale al fatto che il soggetto sia sopravvissuto fino al tempo t .² L'hazard è la generalizzazione del tasso di incidenza al caso di intervalli brevissimi di tempo e corrisponde alla sua stima istante per istante lungo tutto il periodo a rischio. Quindi, mentre il tasso di incidenza è espresso dal valore numerico di un'unica misura, è utile riferirsi all'hazard rappresentandone graficamente l'andamento nel corso del tempo di osservazione.

Lo scopo di questo contributo è quello di illustrare come l'uso di questa funzione si colleghi agli stimatori più usualmente applicati negli studi di coorte e di sopravvivenza e come l'hazard possa rivelarsi un'utile prospettiva per descrivere l'occorrenza di un evento in modo più dettagliato e informativo di quanto consentito dal tasso di incidenza.

Nella prima parte si introduce il metodo per il calcolo della funzione dell'hazard e si presenta in breve lo studio Impatto. Nella seconda parte i dati raccolti in questo studio sono utilizzati per illustrare un esempio di applicazione del metodo a un caso reale.

Stima kernel-smoothed dell'hazard

Il punto di partenza per la stima dell'hazard è l'azzardo cumulativo $H(t)$. Lo stimatore di Nelson-Aalen è un modo semplice, disponibile nella generalità dei software, per stimare $H(t)$,

una step-function con incrementi a ogni tempo di occorrenza dell'evento. La quantità $\Delta H(t_i) = H(t_i) - H(t_{i-1})$ è una stima di $h(t)$ ai tempi di accadimento degli eventi.³

Una stima dell'hazard, $h(t)$, in ogni istante compreso fra il primo e l'ultimo evento osservato, può essere ottenuta lasciando gli incrementi dell'azzardo cumulativo attraverso diversi metodi di smoothing. A questo scopo si possono adoperare splines o forme di regressione locale, ma il metodo più semplice e ampiamente utilizzato è il metodo kernel.⁴ Questo essenzialmente consiste in una media pesata dei $\Delta H(t_i)$, dove i pesi sono definiti in base alla distanza fra il tempo in cui $h(t)$ è stimato e i tempi in cui $\Delta H(t_i)$ sono osservati. Più questa distanza è piccola maggiore sarà il contributo di uno specifico $\Delta H(t_i)$ alla stima di $h(t)$. I pesi sono definiti nell'ambito di una finestra temporale di ampiezza b per cui i $\Delta H(t_i)$ osservati in corrispondenza di tempi maggiori di $t + b$ o minori di $t - b$ hanno pesi uguali a 0 e non contribuiscono alla stima di $h(t)$.⁵

Sia pertanto

$$x = \frac{t - t_i}{b}$$

i pesi possono essere allora calcolati con l'uso di molteplici funzioni kernel di cui una di applicazione particolarmente diffusa è quella di Epanechnikov:⁶

$$K(x) = 0,75 (1 - x^2) \quad \text{per } -1 \leq x \leq 1 \quad \text{e } K(x) = 0$$

La stima di $h(t)$ è allora data da:

$$h(t) = b^{-1} \sum_{i=1}^D K(x) \Delta H(t_i)$$

dove la somma avviene per tutti i D tempi in cui un evento è osservato (adattata da: Klein e Moeschberger 2003, 167, equazione 6.2.4).

La stima di $h(t)$ dipende dalla scelta dell'ampiezza di b , un parametro rispetto al quale si misura la distanza dei $\Delta H(t_i)$ osservati. Una piccola finestra temporale produrrà stime molto variabili eccessivamente influenzate da poche stime locali. Il grafico della funzione dell'hazard assume allora un aspetto frastagliato con picchi e avvallamenti privi di reale significato. Una finestra troppo ampia invece può occultare significative variazioni dell'hazard in tratti particolari del periodo di osservazione. Sono stati proposti metodi per la scelta ottimale di b ,^{7,8} ma nella pratica si preferisce farsi guidare dall'aspetto visivo delle curve ottenute utilizzando diversi valori di b e scegliendo quello che consente di ottenere grafici che raggiungono il miglior compromesso fra i due comportamenti estremi delineati in precedenza.

Una particolare attenzione merita il fatto che nella parte iniziale e finale del periodo di osservazione la stima kernel-smoothed di $h(t)$ può essere distorta in basso se si utilizzano i

pesi kernel sopra indicati. Nella parte iniziale questo si verifica per le stime di $h(t)$ in corrispondenza di tempi inferiori al valore di b . In questa regione per ottenere stime corrette dell'hazard è necessario ricorrere a pesi kernel modificati, detti asimmetrici in quanto tengono conto del fatto che nessun evento è osservabile prima del tempo 0 e che quindi si dispone di un range ristretto di dati. Lo stesso si verifica all'altro estremo del periodo di osservazione quando le stime avvengono in corrispondenza di tempi che distano dall'ultimo evento osservato meno dell'ampiezza di b . Per maggiori dettagli concernenti la stima degli intervalli di confidenza per valutare la precisione della stima di $h(t)$ e le formule dei pesi kernel asimmetrici si rinvia al testo di Klein e Moeschberger.³

Nella prima parte dell'esempio si mostra l'applicazione di questo metodo alla situazione più semplice in cui per descrivere l'occorrenza dell'evento si deriva dallo stimatore dell'azzardo cumulativo di Nelson-Aalen la stima kernel-smoothed dell'hazard.

Adattando ai dati un modello di Cox è possibile stimare l'azzardo cumulativo baseline, corrispondente all'azzardo cumulativo dei soggetti in cui tutte le covariate incluse nel modello hanno valore 0. La tecnica precedentemente illustrata consente di derivare da questo una stima kernel-smoothed dell'hazard baseline. Inoltre, utilizzando le stime dell'azzardo cumulativo baseline e dell'hazard ratio, è possibile calcolare l'azzardo cumulativo predetto dal modello per uno specifico valore delle covariate⁹ e da questo ottenere la corrispondente stima dell'hazard predetto dal modello.

La seconda parte dell'esempio fa vedere come, nel confronto di soggetti con differenti caratteristiche, la funzione dell'hazard possa rivelarsi utile alla verifica degli assunti e all'interpretazione dei risultati ottenuti dopo aver adattato ai dati un modello di Cox.

Lo studio IMPATTO¹⁰

Lo studio IMPATTO raccoglie i dati dei Registri tumori o di patologia attivi nel territorio italiano, che sono in grado di fornire informazioni di popolazione sulle caratteristiche del tumore alla diagnosi e sulla modalità di identificazione, intendendo con questo termine se la diagnosi è avvenuta nell'ambito di un programma di screening organizzato.

Obiettivo dello studio è la valutazione dell'efficacia dei servizi di screening mammografico e dell'impatto che hanno avuto sulla popolazione generale in termini di: andamento dell'incidenza e della stadiazione, uso della chirurgia conservativa nelle aree coperte da screening e riduzione della mortalità per tumore della mammella.

Lo studio include tutti i casi di tumore della mammella, sia in situ che invasivo, diagnosticati tra il 1988 ed il 2001 in donne tra i 40 ed i 79 anni. Tutti i casi sono stati classificati per dimensione e stato linfonodale in base al sistema di stadiazione TNM. Inoltre, per ciascun caso, sono state raccolte anche informazioni sul grading, sul tipo di intervento chirurgico, sul

numero di linfonodi asportati e sul numero dei positivi e sull'effettuazione del linfonodo sentinella. Tutti i casi sono stati seguiti per la valutazione dello stato in vita e della mortalità per causa specifica al 31 dicembre 2005.

Attraverso l'incrocio con gli archivi della popolazione invitata allo screening, ciascun caso è stato classificato in merito alla modalità diagnostica in una delle seguenti categorie:

- cancro identificato al primo test di screening (SD)
- cancro identificato ad un test di screening ripetuto (SD)
- cancro in donne con almeno un test di screening negativo prima dell'incidenza (NSD)
- cancro in donne invitate dal programma e mai rispondenti (NSD)
- cancro in donne non ancora invitate (NSD).

Per ogni caso è stata ricostruita l'intera storia di screening con la prima data di invito e tutte le date dei test di screening eseguiti dalla donna.

E' stata costruita una Banca dati che raccoglie complessivamente 41.427 casi di tumore della mammella incidenti tra il 1988 ed il 2001 in donne tra i 40 ed i 79 anni residenti nelle aree dei seguenti Registri tumori: Bologna, Ferrara, Firenze, Modena, Palermo, Parma, Ragusa, Reggio Emilia, Torino, Umbria e Veneto.

Descrizione dell'occorrenza dell'evento

L'esempio considera i dati dello studio Impatto relativi a 5892 casi diagnosticati tra il 1991 e il 2000 in donne con età compresa fra 50 e 59 anni. In base ai dati di follow-up aggiornati al 31 dicembre 2005 si dispone di 39.769,6 anni-persona durante i quali sono stati osservati 860 decessi attribuibili a tumore della mammella.

La misura di occorrenza riferita a tutto il periodo di osservazione è il tasso di mortalità per tumore della mammella pari al 21,6‰ (tutti i tassi e gli hazard nell'articolo sono

riportati in anni-persona). La stima dell'hazard, a sinistra nella figura 1, è stata ottenuta, a partire dallo stimatore dell'azard cumulativo di Nelson-Aalen applicando la tecnica del kernel-smoothing con una finestra temporale di 3 anni. Subito dopo la diagnosi il tasso è di poco superiore al 12‰, mostra una fase di rapido incremento fino a raggiungere l'acme intorno a 3 anni dalla diagnosi con un valore di poco inferiore al 25‰. In seguito l'hazard decresce e mostra oscillazioni di minore ampiezza fra 5 e 10 anni dalla diagnosi con valori prossimi al 20‰. Il grafico riporta anche una linea tratteggiata orizzontale in corrispondenza del tasso di mortalità sopra specificato e cinque punti che indicano le stime del tasso di mortalità calcolate dividendo il follow-up in intervalli biennali, collocati al centro del rispettivo intervallo. Come può osservarsi la funzione dell'hazard è molto vicina alle stime intervallari del tasso di incidenza. Un test χ^2_4 che saggia l'ipotesi di tassi di mortalità costanti nei 5 bienni è uguale a 20,98, $p < 0,001$, indicando che nel corso del follow-up esiste una variazione dell'occorrenza dei decessi statisticamente significativa di cui è necessario tener conto nella descrizione del fenomeno.

La probabilità di sopravvivenza stimata con il metodo di Kaplan-Meier è illustrata nella figura 1 a destra. L'81% delle pazienti affette da tumore della mammella è viva dopo dieci anni dalla diagnosi. Anche questa funzione esprime l'andamento dell'occorrenza della mortalità durante tutto il corso del periodo di osservazione. Il valore più basso dell'hazard nella prima fase del follow-up corrisponde alla minore inclinazione del primo tratto della funzione di sopravvivenza. Negli anni successivi l'inclinazione della curva di sopravvivenza si accentua e resta approssimativamente costante analogamente a quanto osservato per l'hazard che, dopo una fase di incremento, si assesta a un livello più alto relativamente stabile.

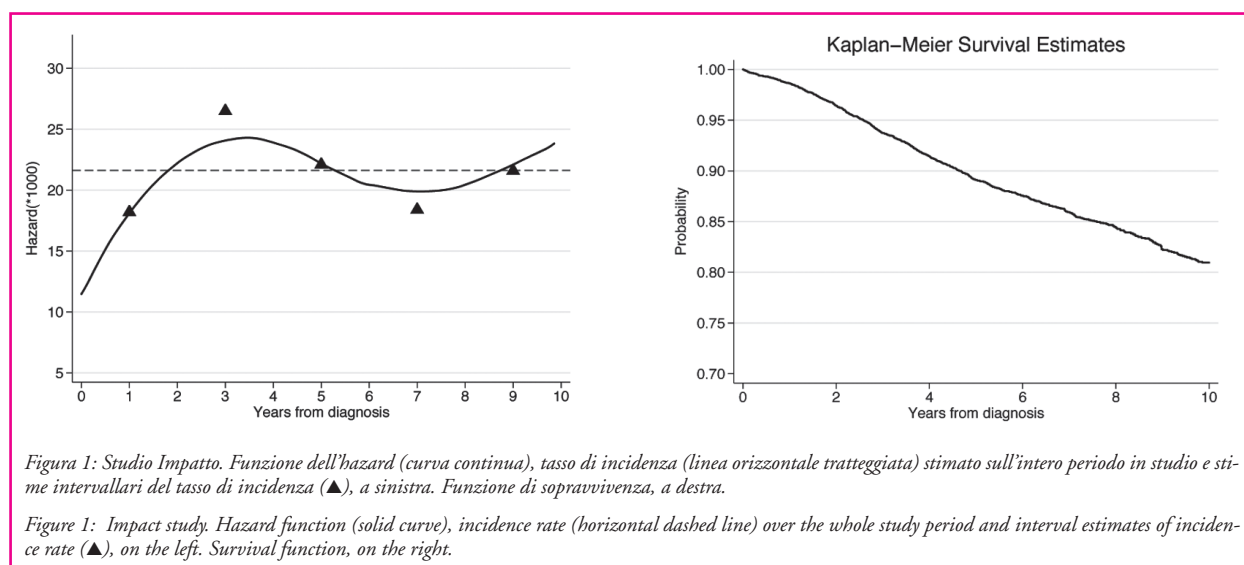


Figura 1: Studio Impatto. Funzione dell'hazard (curva continua), tasso di incidenza (linea orizzontale tratteggiata) stimato sull'intero periodo in studio e stime intervallari del tasso di incidenza (▲), a sinistra. Funzione di sopravvivenza, a destra.

Figure 1: Impact study. Hazard function (solid curve), incidence rate (horizontal dashed line) over the whole study period and interval estimates of incidence rate (▲), on the left. Survival function, on the right.

Confronto fra soggetti

Nelle donne in età compresa fra 50 e 59 anni il tasso di mortalità calcolato fra i casi invitati e non invitati è risultato rispettivamente uguale a 17,3 e 26,8‰ con un rate ratio pari a 0,64 a vantaggio dei casi invitati. La figura 2 mette a confronto le funzioni dell'hazard calcolate a partire dallo stimatore dell'azzardo cumulativo di Nelson-Aalen nei due gruppi (linee continue). I casi invitati hanno un tasso di mortalità inferiore a quello dei non invitati per tutta la durata del follow-up.

Il grafico è stato realizzato su scala logaritmica. Su questa scala la distanza fra le curve è l'espressione grafica dell'effetto (hazard ratio) dell'invito al test di screening sulla mortalità per tumore della mammella. Le funzioni dell'hazard stimate nel gruppo dei casi invitati e dei casi non invitati riproducono lo stesso aspetto della funzione relativa al pool complessivo dei casi riportata nella precedente sezione e la loro distanza resta approssimativamente costante durante tutto il corso del periodo di osservazione. Questo rilievo è un modo immediato per il controllo grafico della proporzionalità degli hazard, assunto la cui verifica è essenziale quando si stima l'effetto di una variabile con un modello di Cox. Infatti, in accordo a quanto illustrato dal grafico, il risultato del test di Grambsch e Therneau,¹¹ che saggia l'ipotesi della proporzionalità degli hazard, è non significativo: $\chi^2_1 = 0,05$, $p = 0,83$.

Dopo aver adattato un modello di Cox è possibile derivare, come descritto nella precedente sezione, la stima degli hazard predetti da modello. Nella figura 2 le curve degli hazard predetti dal modello sono indicate come linee tratteggiate. Essendo state generate secondo l'assunto di proporzionalità degli hazard imposto dal modello, la loro distanza è rigorosamente costante su scala logaritmica. La discreta sovrapposizione delle curve degli hazard ottenute non parametricamente e di quelle predette dal modello secondo l'assunto di proporzionalità indica il buon adattamento del modello ai dati analizzati.

Nella figura 3 è illustrato il confronto fra i casi affetti da tumore in stadio I rispetto a quelli in stadio II o più avanzato. Qui le curve degli hazard stimati in modo non parametrico a partire dall'azzardo cumulativo di Nelson-Aalen (linee continue) hanno una distanza che tende a ridursi nel corso del follow-up. Il grafico perciò suggerisce che l'effetto dello stadio clinico non possa essere descritto da un unico hazard ratio costante per tutto il follow-up ma che ci troviamo di fronte a un effetto non proporzionale che cambia nel tempo. Adattando ai dati un modello di Cox l'hazard ratio corrispondente ai casi di tumore in stadio avanzato è uguale a 7,86. In questo caso, però, il risultato del test di Grambsch e Therneau, $\chi^2_1 = 12,27$, $p < 0,001$, rigetta l'ipotesi di proporzionalità degli hazard e conferma perciò che il precedente hazard ratio non è una misura valida dell'effetto dello stadio sulla mortalità per tumore della mammella.

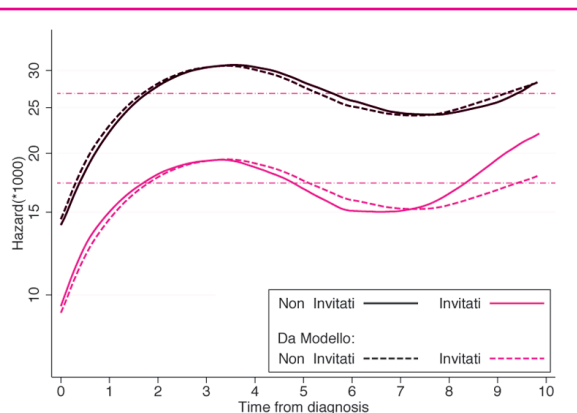


Figura 2: Studio Impatto. Funzione dell'hazard nei casi invitati e non invitati al test di screening (linee continue). Le rette orizzontali tratteggiate corrispondono al tasso di mortalità stimato nei due gruppi. Le curve tratteggiate corrispondono alle stime degli hazard da modello.

Figure 2: Impact study. Hazard function in cases invited and not invited to the screening test (solid curves). Horizontal dashed straight lines are drawn at the level of the incidence rates in both groups. Dashed curves correspond to the hazard estimates predicted by a Cox model.

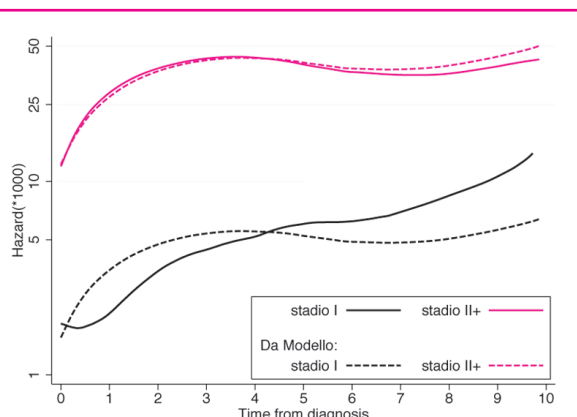


Figura 3: Studio Impatto. Funzione dell'hazard dei casi affetti da tumore della mammella in stadio I e stadio avanzato (linee continue). Le linee tratteggiate corrispondono alle stime dell'hazard predette dal modello.

Figure 3: Impact study. Hazard function in breast cancer cases in stage I and II+ (solid curves). Dashed curves correspond to the hazard estimates predicted by a Cox model.

Le linee tratteggiate corrispondenti agli hazard stimati da modello in base all'assunto dell'effetto proporzionale della variabile stadio sottostimano la reale differenza degli hazard all'inizio del follow-up e la sovrastimano alla fine dando in questo caso l'espressione grafica dello scarso adattamento del modello ai dati analizzati.

Discussione

Una larga parte degli studi che raccolgono informazioni sull'incidenza di malattia o di decesso dispone di informazioni

a livello individuale molto precise sulla data di inizio e di conclusione del periodo di rischio. Rispetto al dettaglio di questa informazione riportare come misura di occorrenza di un evento il tasso di incidenza relativo a tutto il periodo del follow-up è insoddisfacente. In questo modo si assume implicitamente che l'occorrenza del fenomeno sia costante e che quindi la distribuzione degli eventi nel tempo non sia in alcun modo collegata al processo della patologia in studio.

La probabilità di sopravvivenza descrive il verificarsi dell'evento in relazione al tempo di osservazione. Questa funzione non è però espressione diretta della misura di occorrenza dell'evento che può solo essere desunta dall'inclinazione della curva, un dettaglio visivamente non ben valutabile.

E' possibile descrivere le variazioni dell'occorrenza dell'evento dividendo il follow-up in intervalli e calcolando il tasso di incidenza all'interno di ciascuno di essi, ma in questo modo è più difficile apprezzare nel suo complesso l'andamento del tasso di incidenza e le differenze che emergono dal confronto di gruppi di pazienti con specifiche caratteristiche.

Nel caso esemplificato la rappresentazione grafica della funzione dell'hazard nel gruppo di casi invitati e non invitati ha fatto immediatamente emergere due importanti dettagli. Nel primo anno dalla diagnosi le pazienti colpite da tumore della mammella hanno un tasso di mortalità più basso di quello che si registra negli anni successivi. Il secondo aspetto è che la differenza degli hazard fra casi invitati e quelli non invitati al test di screening rimane approssimativamente costante durante tutto il periodo di osservazione. La minore mortalità dei casi invitati in età compresa fra 50 e 59 anni è perciò un vantaggio stabile che non subisce decrementi nel corso del follow-up indagato. Questa è una importante osservazione perché tende a escludere la possibilità che il vantaggio dell'esecuzione dello screening sia solo dovuto al lead time bias cioè all'apparente miglioramento della sopravvivenza indotto dalla mera anticipazione della diagnosi senza che la morte per cancro sia effettivamente posticipata o evitata. Nei programmi di screening del tumore della mammella la stima del lead time medio è pari a circa 4 anni.¹² In mancanza di reali benefici dello screening, una volta decorso questo tempo, le funzioni dell'hazard dei due gruppi dovrebbero tendere ad avvicinarsi fino a incrociarsi, mentre quelle stimate nei dieci anni di osservazione di questo studio non supportano tale tendenza.

Entrambi questi rilievi sarebbero sfuggiti se si fosse indicato solo il tasso di mortalità e sarebbero emersi con minore evidenza ed efficacia comunicativa dalla illustrazione della probabilità di sopravvivenza o dalla esposizione dettagliata dei tassi di mortalità per intervalli di osservazione.

La funzione dell'hazard ha un ruolo centrale nei modelli di regressione applicati nell'analisi della sopravvivenza. L'esempio ha mostrato come la rappresentazione grafica degli hazard illustri efficacemente l'effetto di una variabile nel modificare il tasso di mortalità e dia la percezione immediata della

esistenza o meno della proporzionalità degli hazard. Tale assunto è alla base del modello di Cox e apposite estensioni della sua formulazione di base trovano applicazione in caso di hazard non proporzionali.¹¹ La possibilità di ottenere con questo modello una stima dell'azzardo cumulativo corrispondente al livello 0 (baseline) e a specifici valori delle covariate presenti nel modello consente inoltre di ottenere stime kernel-smoothed dell'hazard previste da modello e di confrontarle con quelle che si possono calcolare non parametricamente per un apprezzamento a colpo d'occhio della bontà di adattamento del modello ai dati.

Il confronto fra la sopravvivenza dei casi invitati e dei casi non invitati al test di screening ha preso in considerazione la mortalità causa-specifica per tumore della mammella. Poiché l'invito al test di screening non ha effetto sulle altre cause di morte è giustificato attendersi che un confronto che prenda in considerazione le altre cause di morte come evento competitivo dia risultati simili a quelli illustrati.¹³

E' importante soffermarsi sul fatto che il periodo di osservazione di uno studio può essere diversamente definito in relazione all'origine da cui parte la misura del tempo. Nell'esempio svolto l'origine è rappresentata dalla data della diagnosi per cui il periodo di osservazione corrisponde al tempo dall'esordio della malattia, ma è possibile stabilire altri riferimenti per l'origine come ad esempio la data di nascita, che corrisponde alla misura del tempo in termini di età, o l'inizio dell'esposizione, scelta spesso interessante nell'ambito degli studi di epidemiologia occupazionale. La rappresentazione grafica dell'hazard consente di studiare l'associazione fra l'occorrenza dell'evento e una particolare scala temporale e può fornire indicazioni per la scelta dell'asse temporale di base da utilizzare nell'analisi e del modo in cui includere le altre scale come variabili tempo-dipendenti del modello di regressione.¹⁴

Nell'esempio l'inclusione nel modello della variabile stadio clinico con effetti che variano nel corso del tempo determina una modifica della forma della funzione dell'hazard stimata in modo non parametrico. Analogamente, se la funzione dell'hazard stimata lungo l'asse temporale del tempo dalla diagnosi modifica la sua forma in seguito all'inclusione nel modello della variabile età, si deve prendere in considerazione la presenza di un'interazione fra la scala temporale del tempo dalla diagnosi con quella del tempo dalla nascita.

Amplia, inoltre, l'utilità del kernel-smoothing la possibilità di impiegare questa tecnica anche in modelli dove la mortalità della coorte in studio è confrontata con quella di una popolazione di riferimento per ottenere stime del relative mortality ratio^{14,15} o dell'hazard in eccesso¹⁶ per tutto il periodo di osservazione.

La scelta della finestra temporale ha rilevante importanza per la stima kernel-smoothed dell'hazard. Oltre a farsi guidare dall'aspetto complessivo del grafico, è importante considerare la numerosità dei casi di cui si dispone. Con un'al-

ta numerosità è possibile utilizzare finestre temporali ristrette mentre nel caso opposto è necessario ricorrere a finestre temporali più ampie. Questa considerazione induce a ritenere ragionevole l'uso di finestre temporali di ampiezza variabile nel corso del follow-up, scegliendo ad esempio una finestra più stretta all'inizio del follow-up quando è maggiore la numerosità dei casi e più larga alla fine quando, per effetto degli eventi e delle censure, il numero dei soggetti ancora in osservazione può essersi ridotto sensibilmente e la stima dell'hazard può essere affetta da una notevole imprecisione rilevabile dall'ampiezza degli intervalli di confidenza.⁸

In realtà, rispetto a questa utilità e versatilità di applicazioni, è da chiedersi perché l'illustrazione della funzione dell'hazard non abbia trovato ampio impiego in campo epidemiologico. Finora questo è in parte stato dovuto alla scarsa disponibilità di software in grado di stimarla. Attualmente potrebbe avere maggiore importanza la poca consuetudine dei ricercatori con questa funzione, cosa che si spera di aver migliorato con questo contributo (tutte le analisi sono state effettuate con Stata ver. 9, a richiesta possono essere forniti dal primo degli autori i codici dei comandi utilizzati per ottenere i risultati).

Conflitti di interesse: nessuno

Elenco del gruppo di lavoro dello studio IMPATTO

Responsabile: Eugenio Paci

D. Puliti, P. Falini, M. Zappa, I. Esposito, E. Crocetti, CSPO Firenze
C. Naldoni, A. C. Finarelli, P. Sassoli de' Bianchi, Regione Emilia-Romagna; S. Ferretti, Registro Tumori Ferrara; Gian Piero Baraldi, Breast Cancer Screening Programme, Ferrara; M. Federico, C. Cirilli, Registro Tumori Modena; R. Negri, Azienda USL Modena; V. De Lisi, P. Sgargi, Registro Tumori Parma; A. Traina, M. Zarcone, Registro di Patologia Palermo; A. Cattani, N. Borciani, Azienda USL Reggio Emilia; L. Mangone, Registro Tumori Reggio Emilia; F. Falcini, A. Ravaoli, R. Vattiato, A. Colamartini, Registro Tumori Romagna; M. Serafini, B. Vitali, P. Bravetti, Azienda USL Ravenna; F. Desiderio, D. Canuti, C. Fabbri, Azienda USL Rimini; A. Bondi, C. Imolesi, Azienda USL Cesena; N. Collina, P. Baldazzi, Azienda USL di Bologna Area Nord; M. Manfredi, C. Petrucci, G. Saguatti, Azienda USL di Bologna Area Città; N. Segnan, A. Ponti, G. Del Mastro, C. Senore, A. Frigerio, S. Pitarella, CPO Piemonte; S. Patriarca, R. Zanetti, Registro Tumori Piemonte; M. Vettorazzi, M. Zorzi, Istituti Oncologico Veneto; Padova; R. Tumino, A. Sigona, Registro Tumori Azienda Ospedaliera Ragusa; G. La Perna, C. Iacono, oncoibla-u.o. Oncologia, Azienda Ospedaliera Ragusa; F. Stracci, F. La Rosa Registro Tumori Umbro; M. Petrella, I. Fusco Moffa, Azienda USL Perugia.

Bibliografia

1. Rothman KJ, Greenland S. Measures of disease frequency. In: Rothman KJ, Greenland S. (eds). *Modern Epidemiology*. Philadelphia, Lippincot Williams & Wilkins, 1998. Vol. 1, 29-46.
2. Marubini E, Valsecchi MG. Distribution functions for failure time T. In: Marubini E, Valsecchi MG (eds). *Analysing Survival Data from Clinical Trials and Observational Studies*. Chichester, John Wiley & Sons, 1995. Vol. 1, 143-170.
3. Klein JP, Moeschberger ML. Topics in univariate estimation. In: Klein JP, Moeschberger ML (eds). *Survival Analysis. Techniques for censored and truncated data*. New York, Springer-Verlag, 2003. Vol 1: 165-200.
4. Wang JL. Smoothing Hazard Rates. In: P. Armitage, T. Colton (eds). *Encyclopedia of Biostatistics*, 2nd Edition. London, Wiley, 2005.
5. Ramlau-Hansen H. Smoothing counting process intensities by means of kernel functions. *Annals of Statistics* 1983; 11: 453-66.
6. Hastie TJ, Tibshirani RJ. Smoothing. In: Hastie TJ, Tibshirani RJ (eds). *Generalized Additive Models*. London, Chapman & Hall, 1990. Vol 1: 9-38.
7. Salgado-Ugarte IH, Pérez-Hernández M. Exploring the use of variable bandwidth kernel density estimators. *The Stata Journal* 2003; 3: 125-32.
8. Müller HG, Wang JL. Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics* 1994; 50: 61-76.
9. Hosmer DW jr, Lemeshow S. Regression models for survival data. In: Hosmer DW jr, Lemeshow S (eds). *Applied Survival Analysis*. New York, John Wiley & Sons, Inc, 1999. Vol 1: 87-111.
10. Federici A. *Screening: profilo complesso di assistenza*. Roma, Il Pensiero Scientifico Editore, 2007.
11. Therneau TM, Grambsch PM. Testing proportional hazards. In: Therneau TM, Grambsch PM (eds). *Modeling Survival Data. Extending the Cox Model*. New York, Springer-Verlag, 2000. Vol 1: 127-52.
12. Paci E, Duffy SW, Giorgi D, et al. Population-based breast cancer screening programmes: estimates of sensitivity, overdiagnosis and early prediction of the benefit. In: Duffy SW, Hill C, Esteve J (eds). *Quantitative methods for the evaluation of cancer screening*. London, Arnold, 2001.
13. Fine JP, Gray RJ. Proportional hazards model for the subdistribution of a competing risks. *Journal of the American Statistical Association* 1999; 94: 496-509.
14. Breslow NE, Day NE. *Fitting models to continuous data*. IN Breslow NE, Day NE (eds). *Statistical Methods in Cancer Research*. Vol II - The Design and Analysis of Cohort Studies. Lyon, IARC Scientific publication No. 87, 1987. Vol 1: 177-229.
15. Andersen PK, Borch-Johnsen K, Deckert T et al. A Cox regression model for relative mortality and its application to diabetes mellitus survival data. *Biometrics* 1985; 41: 921-32.
16. Buckley JD. Additive and multiplicative models for relative survival rates. *Biometrics* 1984; 40: 51-62.