



## Stima dell'incidenza del carcinoma mammario attraverso il flusso dei ricoveri ospedalieri: confronto con i dati dei Registri tumori

## Cancer incidence estimation by hospital discharge flow as compared with cancer registries data

Stefano Ferretti,<sup>1</sup> Stefano Guzzinati,<sup>2</sup> Paola Zambon,<sup>2</sup> Gianfranco Manneschi,<sup>3</sup> Emanuele Crocetti,<sup>3</sup> Fabio Falcini,<sup>4</sup> Stefania Giorgetti,<sup>4</sup> Claudia Cirilli,<sup>5</sup> Monica Pirani,<sup>5</sup> Lucia Mangone,<sup>6</sup> Enza Di Felice,<sup>6</sup> Vincenzo Del Lisi,<sup>7</sup> Paolo Sgargi,<sup>7</sup> Carlotta Buzzoni,<sup>3</sup> Antonio Russo,<sup>8</sup> Eugenio Paci<sup>3</sup>

<sup>1</sup> Registro tumori della Provincia di Ferrara, Università di Ferrara.

<sup>2</sup> Registro tumori del Veneto, Istituto oncologico veneto - IRCCS, Padova.

<sup>3</sup> Registro tumori toscano, Unità operativa di epidemiologia clinica e descrittiva, Istituto per lo studio e la prevenzione oncologica, Firenze.

<sup>4</sup> Registro tumori della Romagna, Istituto scientifico romagnolo per lo studio e la cura dei tumori, Meldola, Forlì-Cesena.

<sup>5</sup> Registro tumori della provincia di Modena, Centro oncologico modenese, Modena.

<sup>6</sup> Registro tumori reggiano, Unità di epidemiologia, Dipartimento di sanità pubblica, Azienda sanitaria locale di Reggio Emilia.

<sup>7</sup> Registro tumori della Provincia di Parma, Unità operativa di oncologia, Azienda ospedaliera-Università di Parma.

<sup>8</sup> Servizio di epidemiologia e statistica medica, Azienda ospedaliera S. Carlo Borromeo, Milano.

**Corrispondenza:** Stefano Ferretti, Registro tumori della Provincia di Ferrara, Dipartimento di medicina sperimentale e diagnostica, Sezione di anatomia, istologia e citologia patologica, via Fossato di Mortara 64B, 44100 Ferrara, tel. 0532 455513, fax 0532 248021; e-mail: frs@unife.it

### Cosa si sapeva già

■ Il progressivo miglioramento dei flussi sanitari informatizzati suggerisce l'opportunità di ottenere stime di incidenza per i tumori a partire dai soli ricoveri ospedalieri con maggiore rapidità rispetto alla latenza temporale dei Registri tradizionali e la possibilità di ottenere indicatori per le aree non coperte.

■ L'accuratezza dei dati ottenuti rappresenta l'elemento decisivo per la loro utilizzazione ai fini della programmazione e della verifica di impatto degli interventi sanitari.

### Cosa si aggiunge di nuovo

■ La concordanza osservata in termini di stime dell'incidenza tra Registri tumori e un algoritmo sviluppato sulle schede di dimissione ospedaliera nasconde l'effetto della combinazione crociata di problemi di sensibilità e specificità del flusso SDO.

■ Il numero dei casi falsi positivi e l'evidente perdita selettiva dei casi a peggiore prognosi pone seri limiti e notevoli ricadute sulla possibilità di utilizzazione dell'algoritmo sviluppato nello studio degli interventi sanitari, in particolare per quanto riguarda i percorsi assistenziali.

### Riassunto

**Obiettivo:** questo studio si è posto l'obiettivo di valutare l'accuratezza di un algoritmo basato sulle schede di dimissione ospedaliera (SDO) nella stima dell'incidenza del carcinoma mammario femminile in tre regioni italiane (Emilia-Romagna, Toscana e Veneto). Lo sviluppo e l'evoluzione di flussi sanitari informatizzati come quello delle SDO propone infatti questa possibile alternativa al più oneroso modello di funzionamento dei Registri tumori (RT) allo scopo di ottenere in tempi più rapidi una stima dell'incidenza delle neoplasie, anche al fine della valutazione di impatto delle procedure terapeutiche.

**Disegno e setting:** nelle aree coperte da Registri tumori di tre regioni italiane (Emilia-Romagna, Toscana e Veneto) è stata valutata la *performance* di un algoritmo di estrazione dei casi incidenti dei tumori della mammella femminile basato sulla diagnosi alla dimissione e sulla contestuale terapia chirurgica. La stima dell'incidenza così ottenuta è stata confrontata con i corrispondenti tassi di incidenza dei Registri

tumori. L'accuratezza dell'algoritmo SDO è stata inoltre verificata attraverso l'appaiamento dei casi SDO con l'archivio di Registro pertinente e attraverso l'analisi di sopravvivenza dei casi sfuggiti al modello.

**Risultati:** nel periodo di osservazione gli RT coinvolti hanno registrato 6.079 casi incidenti a fronte di 3.125.425 anni/persona. Il flusso SDO ha contestualmente individuato 6.000 casi. I tassi di incidenza dei due metodi (RT 194,5; SDO 192,0 x 100.000) non hanno mostrato differenze significative. I casi effettivamente concordanti sono però risultati 5.038, con una sensibilità e un valore predittivo positivo dell'algoritmo SDO pari rispettivamente a 82,9% e 84,0%, una quota di "falsi positivi" del 9,9% e una quota di casi non intercettati pari al 12,3% dell'incidenza RT. Questi ultimi sono prevalentemente casi non trattati chirurgicamente, a prognosi peggiore (sopravvivenza a 3 anni pari al 62,0% contro l'88,8% della casistica generale) e con una più alta quota proporzionale di pazienti anziane.

**Conclusioni:** le stime ottenute dall'algoritmo SDO risultano non molto dissimili dai dati già noti della letteratura. In realtà la concordanza dell'algoritmo con i dati degli RT nella stima globale dell'incidenza appare come la risultante dell'effetto crociato di carenze di sensibilità e specificità dell'algoritmo nell'individuazione dei casi. Ciò compromette l'affidabilità di questo modello nella valutazione di impatto dei percorsi cli-

nico-assistenziali nell'ottica di uno studio longitudinale su coorti di pazienti oncologici, che rappresenta oggi la prospettiva più promettente per la programmazione sanitaria. (*Epidemiol Prev* 2009; 33 (4-5): 147-53)

**Parole chiave:** registri tumori, schede dimissione ospedaliera, stima dell'incidenza dei tumori.

### Abstract

**Objective:** the study evaluates the accuracy of an algorithm based on hospital discharge data (HDD) in order to estimate breast cancer incidence in three Italian regions (Emilia-Romagna, Toscana and Veneto) covered by cancer registries (CR). The evolution of computer-based information systems in health organization suggests automatic processing of HDD as a possible alternative to the time-consuming methods of CR. The study intends to verify whether HDD quickly provides reliable cancer incidence estimates for diagnosis and therapy evaluations.

**Design and setting:** an algorithm based on discharge diagnosis and surgical therapy of hospitalized breast cancer patients was developed in order to provide breast cancer incidence. Results were compared with the corresponding incidence data of cancer registries. The accuracy of the automatic method was also verified by a direct record-linkage between HDD output and registries' files. The overall survival of cases lost to "HDD method" was analyzed.

**Results:** in the period covered by the study (3,125,425 person/year) CR enrolled 6,079 incident cases, compared to 6,000 cases record-

ed through the HDD flow. Incidence rates of the two methods (CR 194.5; HDD 192.0  $\times$  100.000) showed no statistical differences. However, matched cases by the two methods were only 5,038. The sensitivity of the HDD algorithm was 82.9% and its predictive positive value (PPV) was 84.0%. False positive cases were 9.9%. On the other hand, 12.3% CR incident cases were not identified by the algorithm: these were mainly made up of older women, not eligible for surgical therapy. Their three-years survival was 62.0% vs 88.8% of the whole incidence group.

**Conclusion:** HDD flow performance was similar to observations reported in the literature. The agreement between HDD and CR incidence rates is a result of a cross effect of both sensitivity and specificity limitations of the HDD algorithm. This can seriously impair the reliability of the latter method with regard to the evaluation of diagnostic and therapeutic strategies in cohort studies (i.e. the most effective approach to health setting in oncology). (*Epidemiol Prev* 2009; 33 (4-5): 147-53)

**Keywords:** cancer registries, hospital discharge data, cancer incidence estimate.

### Introduzione

I Registri tumori (RT) hanno rappresentato per un consistente periodo l'unica fonte in grado di fornire dati di incidenza su base di popolazione delle neoplasie, evitando le distorsioni da selezione altrimenti presenti in analoghe valutazioni costruite su casistiche cliniche o su archivi gestionali.<sup>1,2</sup> In Italia, negli ultimi decenni, il sistema sanitario si è progressivamente dotato di flussi informativi amministrativi il cui miglioramento, in termini di copertura e accuratezza, ne ha progressivamente incoraggiato l'utilizzazione anche ai fini valutativi epidemiologici e clinici.

I Registri tumori hanno essi stessi tratto notevole beneficio da questo processo di informatizzazione, che ha facilitato la ricognizione dei casi incidenti altrimenti affidata ad attività manuali (reperimento e consultazione delle cartelle cliniche, diagnosi microscopiche e certificati di decesso).<sup>3</sup> Ciò ha nel contempo portato a una generale riflessione critica sul ruolo dei Registri tumori, sia per le nuove prospettive di studio aperte da questa informatizzazione, sia nel senso della riconsiderazione dell'efficienza (costo/beneficio) dell'attività dei Registri, soprattutto sul terreno della tempistica di produzione e della qualità dei dati.<sup>4</sup>

L'incidenza delle neoplasie maligne è materia da tempo non più limitata ad approcci descrittivi, ma ormai accettata a pieno titolo tra gli indicatori fondamentali per studi di rischio

oncologico e impatto delle procedure di diagnosi e terapie a livello regionale e nazionale.<sup>5-7</sup> Si è quindi inevitabilmente prospettata la possibilità che i sistemi informativi sanitari correnti, opportunamente interrogati, siano ormai da soli in grado di garantire le informazioni tradizionalmente fornite dai Registri,<sup>8,9</sup> a livelli di qualità accettabili, con drastica riduzione dei tempi di produzione, più agevole estensione a gran parte o a tutto il territorio nazionale<sup>10,11</sup> e possibilità di allargamento a patologie non oncologiche di interesse.<sup>12</sup>

Indagini di questo genere, in parte eterogenee nei metodi e nelle finalità, sono presenti in letteratura e recentemente hanno documentato inevitabili discrepanze, soprattutto in termini di accuratezza, tra gli algoritmi basati sui dati di dimissione ospedaliera e l'incidenza tradizionalmente individuata dai Registri tumori, con ricadute più o meno consistenti sulla stima dell'incidenza o sulla possibilità di utilizzare i dati ai fini di programmazione o verifica delle politiche sanitarie.

Il presente studio si propone di confrontare, in termini di accuratezza e affidabilità, i dati di incidenza ottenuti attraverso un algoritmo di estrazione di dati applicato al flusso informativo delle schede di dimissione ospedaliera (SDO) rispetto alle casistiche dei Registri tumori, soprattutto nella prospettiva di utilizzazione di queste fonti nella programmazione dei percorsi di prevenzione e di assistenza dei pazienti oncologici. Le neoplasie della mammella femminile, scelte per questo raffronto, iden-

Codici diagnosi dimissione	Codici procedure chirurgiche concomitanti
174.0-9 Tumori maligni della mammella	8512 Biopsia a cielo aperto 8520 Asport./demol. SAI 8536 Altra mammectomia sottocutanea bilaterale 8541 Mx semplice monolat.
233.0 Carcinomi in situ della mammella	8521 Asport. locale 8522 Quadrantectomia 8542 Mx semplice bilat. 8543 Mx semplice allarg. monolat.
V10.3 Anamnesi personale di tumore maligno della mammella	8523 Mx subtotale 8524 Asportazione tx ectop. 8525 Asportazione capezzolo 8533 Mammectomia sottocutanea monolaterale + impianto protesi 8534 Altra mammectomia sottocutanea monolaterale 8535 Mammectomia sottocutanea bilaterale + impianto protesi 8544 Mx semplice allarg. bilat. 8545 Mx radicale monolaterale 8546 Mx radicale bilaterale 8547 Mx radicale monolat. allarg. 8548 Mx radicale bilat. allarg.

Tabella 1. Algoritmo SDO, criteri di estrazione dei casi.

Table 1. SDO Algorithm, case-eligibility criteria.

tificano una patologia a marcata tracciabilità da parte degli archivi informatizzati delle SDO in virtù di un percorso clinico omogeneo che prevede un'alta percentuale di ricoveri per intervento chirurgico; esse appaiono quindi in grado di fornire elementi di riflessione critica nelle condizioni di migliore *performance* delle tecniche di elaborazione automatica.

## Metodi

Lo studio è stato condotto sulla popolazione di tre regioni italiane (Emilia-Romagna, Toscana e Veneto) coperta dall'attività di Registri tumori. Dalle schede di dimissione ospedaliera dei pazienti residenti nelle aree in studio sono stati estratti i casi incidenti nel corso di un anno solare (2003 per il Veneto, 2004 per Emilia-Romagna e Toscana), definiti sulla base di un ricovero con diagnosi di dimissione principale o secondaria di neoplasia mammaria e contestuale intervento chirurgico principale o secondario correlato (tabella 1).

Tale casistica è stata successivamente confrontata con l'analogo archivio prodotto dal *pool* dei Registri, con approfondimenti nel contesto delle singole regioni. Tra i casi incidenti dei Registri sono stati inclusi, in deroga alle regole internazionali,<sup>13</sup> anche i tumori bilaterali e monolaterali successivi al primo, sincroni o metacroni e i carcinomi *in situ*. La fonte del Registro è stata considerata *gold standard* per la valutazione di specificità e sensibilità dell'algoritmo SDO, in quanto essa stessa

verso un *linkage* individuale con valutazione della sensibilità (% dei casi identificati dall'algoritmo SDO sul totale degli incidenti RT) e valore predittivo positivo del flusso SDO (PPV, % dei casi confermati dal RT tra gli identificati dall'algoritmo SDO). Dal gruppo dei casi discordanti sono stati successivamente scorporati quelli non riconducibili al medesimo anno di incidenza per motivi legati alla fisiologica latenza tra la prima diagnosi (data di incidenza del RT) e l'intervento chirurgico segnalato dalla SDO; in questo sottogruppo sono stati compresi i casi con SDO nell'anno in studio e data di incidenza RT nell'ultimo trimestre dell'anno precedente e i casi con incidenza RT nell'anno e SDO entro i tre mesi dell'anno successivo. I casi discordanti finali (al netto della latenza temporale) sono stati successivamente analizzati in ordine alla loro natura ("falsi positivi" e casi non intercettati dalla SDO) e al motivo della discrepanza rispetto ai dati in possesso del Registro. Tra i casi di tumore invasivo incidenti secondo la fonte RT e non individuati dall'algoritmo SDO è stato inoltre condotto un approfondimento sui tassi età specifici e sulla sopravvivenza a 12 e 36 mesi per l'intero *pool* e limitatamente a 30 mesi per il dettaglio delle singole aree, utilizzando per quest'ultima analisi il metodo attuariale (SPSS v. 8.0; SPSS Inc, Chicago IL) calcolato dalla data di incidenza del primo tumore mammario invasivo, con esclusione dei casi segnalati dal solo certificato di morte (DCO).

Regione	Algoritmo SDO			Registri tumori		
	n. casi	tassi*	IC 95%	n. casi	tassi*	IC 95%
Emilia-Romagna (2004)	2,988	199.0	191,9-206,3	3,047	202.9	195,8-210,3
Toscana (2004)	1,194	191.9	181,2-203,1	1,207	194.0	183,2-205,2
Veneto (2003)	1,818	181.5	173,3-190,1	1,825	182.2	174,0-190,8
Totale	6,000	192.0	187,2-196,9	6,079	194.5	189,6-199,5

Tabella 2. Confronti tra i tassi di incidenza.

Table 2. Comparisons between incidence rates.

Tabella 3. Casi discordanti.  
Table 3. Discordant cases.

	Emilia-Romagna		Toscana		Veneto		pool	
	n. casi	%	n. casi	%	n. casi	%	n. casi	%
<b>Casi falsi positivi SDO</b>								
casi prevalenti/recidive	211	68.5	84	68.3	131	79.4	426	71.5
les. benigne/altri tumori	60	19.5	34	27.6	22	13.3	116	19.5
errori compil. SDO	28	9.1	5	4.1	9	5.5	42	7.0
casi non definiti	9	2.9	0		3	1.8	12	2.0
Totale	308	100.0	123	100.0	165	100.0	596	100.0
<b>Casi persi dall'algorithm SDO (falsi negativi)</b>								
SDO assente	129	32.6	9	6.1	136	66.7	274	36.6
SDO tardive	9	2.3	5	3.4	0		14	1.9
SDO errate	115	29.0	36	24.3	47	23.0	198	26.5
DCI/DCO	2	0.5	30	20.3	11	5.4	43	5.7
SDO non eligibili	141	35.6	68	45.9	10	4.9	219	29.3
Totale	396	100.0	148	100.0	204	100.0	748	100.0

Sulla casistica di Toscana e Veneto è stata infine condotta una rivalutazione della concordanza non considerando le SDO di casi "prevalenti" secondo le regole internazionali (precedenti ricoveri per tumore mammario almeno nei 5 anni precedenti l'indagine).

Gli intervalli di confidenza (IC) dei tassi di incidenza sono stati calcolati secondo la distribuzione di Poisson, mentre per la sensibilità/PPV dell'algorithm e per la sopravvivenza sono stati utilizzati rispettivamente l'intervallo binomiale esatto e quello approssimato secondo la distribuzione normale.

## Risultati

Nel periodo e nelle tre regioni in esame (6.294.202 donne residenti,<sup>14</sup> di cui il 49,7% in carico agli RT per complessivi 3.125.425 anni/persona) i Registri tumori hanno segnalato 6.079 nuovi casi di neoplasie mammarie, con un tasso grezzo di 194,5 x 100.000 residenti. La sola fonte SDO ha evidenziato complessivamente 6.000 casi con un tasso di 192,0 x 100.000. In ciascuna regione e nel complesso del *pool* i tassi ottenuti attraverso i due metodi non hanno mostrato significative differenze (tabella 2).

Dall'integrazione (*linkage*) RT/SDO si è osservato però come soltanto 5.038 casi siano stati condivisi dai due flussi, con una sensibilità dell'algorithm SDO dell'82,9% (IC 95% 81,9-83,8) e un PPV dell'84,0% (IC 95% 83,0-84,9). I casi con reale discrepanza tra fonte SDO e RT, al netto cioè dei discordanti per registrazioni nei trimestri a cavaliere tra l'anno precedente e successivo, sono risultati in totale 1.378, divisi tra casi segnalati dalla fonte SDO in assenza di neoplasie mammarie secondo il RT (596 casi, 9,9% della intera casistica SDO), casi recuperati dalla fonte SDO e precedentemente persi dal RT

(34 casi, 0,6% della casistica SDO) e casi incidenti per il RT e non segnalati dalla SDO (748 casi, 12,3% della casistica RT). Limitatamente alla fascia di età dai 45 ai 74 anni la percentuale di casi persi alla SDO (331 casi) è risultata dell'8,2%.

Tra i casi discordanti (tabella 3), il dettaglio dei 596 casi «falsi positivi SDO» segnala un 71,5% di casi incidenti in anni precedenti (prevalenti con interventi su recidive), il 19,5% di errate segnalazioni (tumori benigni o altre lesioni), un 7% di errori nella compilazione della SDO (anagrafici, codifiche) e una piccola residua quota (12 casi, 2%) senza sufficienti evidenze e ancora in corso di accertamento da parte degli RT.

I 748 casi incidenti secondo gli RT e "persi" dalla sola rilevazione SDO comprendono, con alcune differenze tra le diverse aree, casi con SDO assente o non attribuibile (36,6%), SDO oltre il primo trimestre dell'anno successivo (1,9%), SDO errate (identificativi, codifiche: 26,5%), casi recuperati dal certificato di morte dall'RT (5,7%), casi generalmente non eligibili per intervento chirurgico o con indicazione di lesione benigna/incerta (29,3%). Si tratta peraltro di casi di cui gli RT possiedono conferma microscopica nell'86,4% dei casi (65,9% istopatologica, 20,5% citopatologica) a fronte di circa il 95% della casistica nazionale (91% istologica, 4% citologica).<sup>15</sup>

Il confronto tra la casistica incidente globale del carcinoma mammario nelle aree e nel periodo considerato e le lesioni non rilevate dall'algorithm SDO mostra inoltre un comportamento differente per quanto riguarda l'incidenza per classi di età,

	12 mesi				30 mesi			
	pool incidenza		persi alla SDO		pool incidenza		persi alla SDO	
	SOPRAVVIVENZA	IC 95%	SOPRAVVIVENZA	IC 95%	SOPRAVVIVENZA	IC 95%	SOPRAVVIVENZA	IC 95%
Emilia-Romagna	96.4	95,6-97,1	81.4	77,4-85,5	89.4	87,9-90,9	66.9	62,0-71,9
Toscana	95.5	94,2-96,8	73.2	65,6-80,8	89.0	86,9-91,1	54.7	46,2-63,3
Veneto	96.3	95,3-97,2	81.2	75,1-87,2	90.4	88,9-91,8	68.6	61,4-75,8
pool	96.2	95,6-96,7	79.7	76,6-82,8	89.8	88,9-90,7	64.9	61,2-68,6

Tabella 4. Sopravvivenza totale e dei pazienti persi alla rilevazione SDO per Regione.  
Table 4. Survival of all cases and of those lost by SDO algorithm, by Region.

incidenza algoritmo SDO	criteri protocollo		esclusioni prevalenti	
	n. casi	tassi	n. casi	tassi
	3,012	185.5 (178,9-192,2)	2,800	172.4 (166,1-179,0)
casi falsi positivi SDO	n. casi	% SDO	n. casi	% SDO
	288	9.6	137	4.9
casi persi dall'algoritmo SDO (falsi negativi)	n. casi	% RT	n. casi	% RT
	352	11.6	391	12.9

con una più marcata rappresentazione proporzionale nelle età avanzate dei casi "persi" (figura 1). Considerando infine il comportamento clinico di queste ultime neoplasie, l'analisi dei tumori invasivi evidenzia una sopravvivenza a 12 e 30 mesi dalla diagnosi rispettivamente del 79,7% e del 64,9%, contro il 96,2 e l'89,8% della casistica generale, con dati omogenei tra le diverse regioni in studio (tabella 4). Lo svantaggio in termini prognostici appare uniformemente distribuito in tutto il periodo di follow up considerato e con differenze statisticamente significative a 36 mesi fra *pool* (88,8%), casi persi alla SDO di tutte le età (62,0%) e nella fascia 45-74 anni (71,9%; figura 2). L'esclusione dalla casistica SDO nell'anno in studio dei casi con analogo ricovero e intervento per neoplasia mammaria negli anni precedenti effettuato a scopo esplorativo in due regioni (Toscana e Veneto), ha evidenziato una notevole diminuzione (di oltre il 50%) dei casi «falsi positivi SDO» a fronte però di un aumento (dell'11% circa) della quota dei casi incidenti (fonte RT) non rilevata dall'algoritmo (tabella 5).

## Discussione

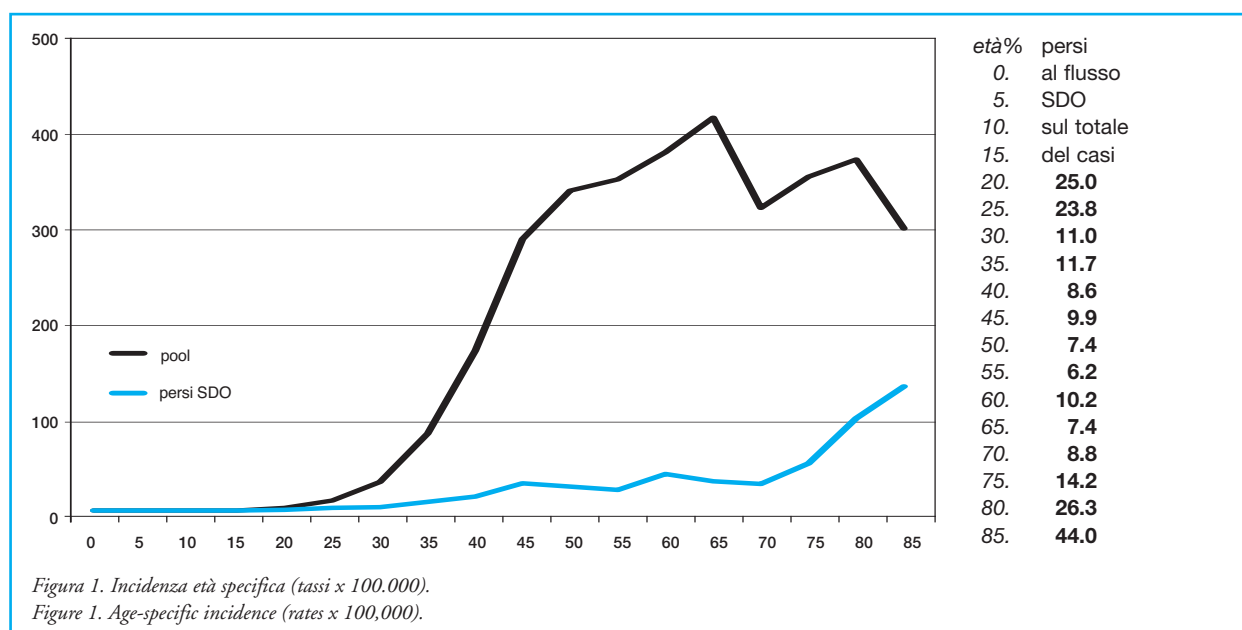
Lo studio ha preso in considerazione la casistica incidente in tre regioni italiane caratterizzate da una buona copertura da

parte dei Registri tumori e conseguentemente da un gran numero di casi incidenti di

tumori maligni della mammella femminile, nell'intento di valutare la *performance* di un sistema di valutazione dell'incidenza basato sul solo flusso delle SDO nelle migliori condizioni potenziali (patologia diffusa, buona tracciabilità). La metodologia adottata per l'estrazione dei dati segue criteri già documentati in letteratura.<sup>16</sup>

Preliminarmente a considerazioni di merito sui dati, occorre evidenziare alcune caratteristiche intrinseche del flusso SDO che identifica strutturalmente una serie di prestazioni sanitarie, a fronte di un Registro tumori che rappresenta invece un archivio di patologie. La fonte SDO è necessaria a un Registro di neoplasie, ma non è automaticamente riconducibile a esso: il trattamento di tumori multipli o multicentrici può essere effettuato infatti nell'ambito dello stesso ricovero (e generare una sola segnalazione), così come interventi in tempi diversi su di un'unica lesione possono generare più documenti. Non è inoltre possibile desumere con sicurezza, soprattutto negli organi pari, molteplicità tumorali sulla base di codici di interventi chirurgici bilaterali, per l'assenza di un appaiamento intervento-lesione non previsto dalla procedura, né è generalmente consentito dalla codifica della patologia (ICD-9 CM). Ne consegue l'impossibilità di abbinamento univoco

parte dei Registri tumori e conseguentemente da un gran numero di casi incidenti di tumori maligni della mammella femminile, nell'intento di valutare la *performance* di un sistema di valutazione dell'incidenza basato sul solo flusso delle SDO nelle migliori condizioni potenziali (patologia diffusa, buona tracciabilità). La metodologia adottata per l'estrazione dei dati segue criteri già documentati in letteratura.<sup>16</sup>



tra ricovero e patologia con l'accuratezza richiesta da una rilevazione di incidenza: le stesse procedure di *record linkage* risentono inevitabilmente di queste discrepanze. La trasformazione di una serie di prestazioni in una sequenza di lesioni richiede perciò procedure di approfondimento anche su altre fonti (come la diagnostica anatomo-patologica), che rientrano di fatto nei tradizionali e più onerosi schemi di lavoro di un Registro tumori. Diversamente da altri approcci sulla fonte SDO presenti in letteratura che considerano soltanto il primo ricovero con intervento per tumore, analogamente alle regole internazionali seguite dai Registri, il presente studio ha considerato anche i tumori mammari successivi al primo, mono/ bilaterali, sincroni o metacroni e le lesioni *in situ*.

La maggiore complessità di questo approccio appare più accurata nel documentare una possibile molteplicità di tumori che, ancorché esclusi dalle casistiche ufficiali di incidenza, costituiscono comunque ulteriori patologie per le pazienti e un carico assistenziale aggiuntivo per l'organizzazione sanitaria. Date queste premesse, un primo risultato evidenzia una buona corrispondenza delle stime di incidenza del «metodo SDO» rispetto ai dati RT, complessivamente e nel dettaglio delle singole realtà territoriali, migliore di altre esperienze italiane.<sup>17</sup> Si tratta tuttavia solo di una concordanza apparente tra due casistiche notevolmente diverse, realizzata in virtù della sommaria di discordanze di segno opposto. In realtà l'algoritmo SDO appare gravato da una bassa specificità che porta all'arruolamento di un casistica in cui, all'approfondimento anatomo-patologico e clinico, quasi un caso su dieci non risulta essere una neoplasia mammaria.

Ai fini dello studio dei percorsi diagnostici e terapeutici un secondo e più grave problema è rappresentato dal mancato reperimento di una consistente serie di lesioni, analogamente a quanto riportato in altri contributi presenti in letteratura. La sensibilità e il PPV dell'algoritmo (82,9% e 84,0%) sono risultati infatti non troppo dissimili da quanto riscontrato da altri autori, in particolare in Italia (Piemonte) da Baldi et al. (76,7% e 92,6%), in Francia da Ganry et al. (85% e 97% per le donne con prima diagnosi di ospedalizzazione, 94% e 78%

con seconda diagnosi) e negli Stati Uniti da Warren et al. (62% e 88%).<sup>9,16,18</sup>

L'analisi supplementare in due regioni escludendo le segnalazioni SDO di pazienti con diagnosi principale di tumore mammario in anni anteriori a quello in studio è stata condotta per consentire un ulteriore approfondimento comparativo con la letteratura, orientato a un maggior controllo di eventuali casi "prevalenti". Fermo restando il rischio di escludere possibili casi incidenti sulla base di precedenti SDO errate (accertate in quote non trascurabili nella casistica in studio), questo metodo offre una migliore specificità (con dimezzamento dei casi "falsi positivi" per miglior controllo dei prevalenti) pagata però con un'ulteriore perdita di sensibilità (aumento dei casi "persi alla SDO" di oltre l'11%).

La conseguenza più critica nell'uso del solo algoritmo SDO è rappresentata in sostanza dalla perdita selettiva di tumori in pazienti anziane e a peggiori prognosi, per l'impossibilità dell'algoritmo di intercettare donne non trattate chirurgicamente per problemi di età, gravità del quadro clinico o stadio della malattia alla diagnosi, ma anche per incompleta o errata compilazione delle SDO o per approfondimenti diagnostici conclusi successivamente alla dimissione ospedaliera. Ciò appare chiaramente attraverso l'analisi della sopravvivenza osservata a breve-medio termine. La prognosi peggiore generalmente osservata tra le pazienti non individuate dall'algoritmo

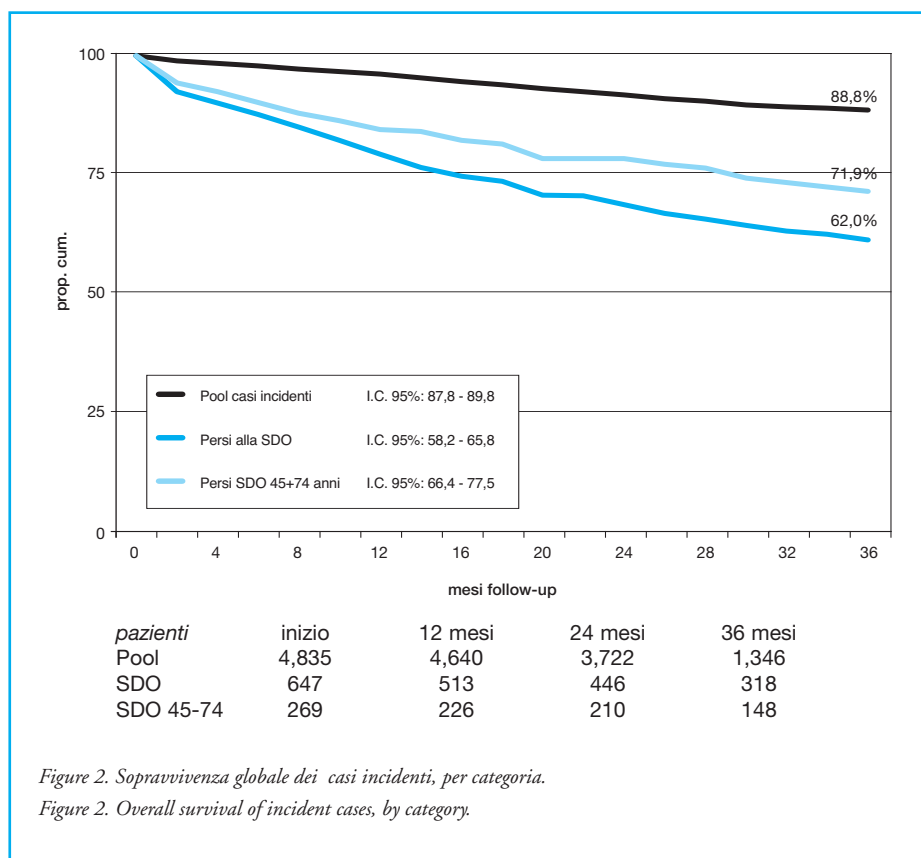


Figure 2. Sopravvivenza globale dei casi incidenti, per categoria.  
 Figure 2. Overall survival of incident cases, by category.

SDO è confermata anche nel gruppo tra i 45 e i 74 anni di età, di particolare interesse, per lo studio dell'impatto dei programmi di screening. Questa classe di età mostra infatti livelli intermedi e significativamente differenti rispetto sia all'intero gruppo di «persi SDO», sia ai valori espressi dall'intera casistica incidente dei Registri tumori.

### Conclusioni

A fronte delle potenziali premesse a suo favore (economia e pronta disponibilità dei dati rispetto alla latenza e al maggior lavoro dei Registri), l'algoritmo di intercettazione dei casi dal solo flusso SDO non sembra soddisfare i criteri di accuratezza richiesti dall'esigenza di identificare coorti di pazienti incidenti per approfondirne i percorsi diagnostico-assistenziali, ovvero ciò che rappresenta attualmente il contributo più importante dei Registri tumori. L'uso della sola fonte SDO, se da un lato può stimare tassi di incidenza fortuitamente concordanti con quelli prodotti dai RT, dall'altro presenta di fatto consistenti problemi di specificità, sensibilità e distorsione che ne vanificano l'uso ai fini di sanità pubblica, soprattutto orientata alla valutazione di terapie e follow up.

I problemi osservati sulle neoplasie mammarie, patologie di chiara definizione e ad alta tracciabilità, fanno inoltre prevedere ben più gravi difficoltà per altri tipi di tumore in cui l'identificazione dell'incidenza comprende una più alta quota di casi da definire attraverso attente analisi integrate e retrospettive, spesso non supportate da livelli diagnostici isto-citopatologici o che non prevedono ricoveri ospedalieri.

La "lentezza" dei Registri tumori, più volte evidenziata, è del resto in gran parte dovuta proprio alla necessità dell'integrazione di più fonti (anagrafiche, anatomo-patologiche, cliniche, di mortalità), operazione oggi ancora gravata da problemi di standardizzazione e codifica condivisa dei dati, superabili solo al prezzo di un rilevante lavoro manuale.<sup>19,20</sup>

La documentazione delle schede di dimissione ospedaliera, insieme agli altri flussi informativi sanitari, rappresenta invece un formidabile strumento di indagine in una logica di mutua integrazione, particolarmente in patologie complesse o all'interno di programmi articolati. Gli screening oncologici rappresentano, tra gli altri, l'esempio tipico della necessità di approfondimento multidisciplinare a cui l'attività di registrazione può rispondere in maniera soddisfacente solo integrando l'enorme potenziale informativo che già oggi gli archivi sanitari correnti offrono in termini di ricoveri, assistenza ambulatoriale e domiciliare, prescrizioni farmaceutiche, diagnosi anatomo-patologiche, strumentali e molto altro ancora.

In questo senso, nel sistema informativo sanitario italiano, tutti gli sforzi per migliorare l'integrabilità dei vari flussi informativi già esistenti appaiono ragionevolmente più produttivi rispetto a interventi di interrogazione "spinta" di una sola fonte. L'esperienza quotidiana dei Registri tumori può in questo senso rappresentare un contributo garantendo già in questa logica, e attraverso l'unione delle diverse fonti, la copertura di tut-

ta la storia clinica dei pazienti (dalla diagnosi alla terapia con percorsi, tempi e modalità): se ciò è condizionato da tempi necessariamente non brevissimi, i Registri sono d'altro canto in grado di fronteggiare richieste tempestive di stime geografiche o temporali con l'affidabilità che deriva loro dalle robuste serie storiche ormai disponibili, validate a livello internazionale.

Ulteriori sforzi in questo senso saranno probabilmente da prevedere per permettere a questo patrimonio, ancora in larga parte inesplorato, di divenire strumento ordinario di programmazione razionale di tutto l'approccio preventivo e clinico in oncologia.

**Conflitti di interesse:** nessuno.

### Bibliografia

1. Wagner G. History of cancer registration. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG (eds). *Cancer registration: principles and practice*. Lyon, IARC Sc. Publ. N. 95, 1991.
2. Storm HH. Cancer registries in epidemiological research. *Cancer Causes Control* 1996; 7: 299-301.
3. Ferretti S, Giacomini A. *Tecniche di registrazione dei tumori*, Milano, Inferenze, 2008.
4. Zanetti R, Rosso S. Riuscirà l'epidemiologia a salvare i registri tumori? *Epidemiol Prev* 2002; 26(4): 198-202.
5. Terracini B, Zanetti R. A short history of pathology registries, with emphasis on cancer registries. *Soz Präventivmed* 2003; 48(1): 3-10.
6. Parkin DM. The role of the registries in cancer control. *Int J Clin Oncol* 2008; 13: 102-11.
7. Parkin DM. The evolution of population-based cancer registry. *Nat Rev Cancer* 2006; 6: 603-12.
8. Rostgaard K, Holst H, Mouridsen H et al. Do clinical databases render population-based cancer registries obsolete? The example of breast cancer in Denmark. *Cancer Causes Control* 2000; 11: 669-74.
9. Ganry O, Taleb A, Peng J et al. Evaluation of an algorithm to identify incident breast cancer cases using DRGs data. *Eur J Cancer Prev* 2003; 12: 295-99.
10. Bolzan M, Conte G. Evaluation in terms of annual incidence rate of lung cancer within the ULSS of the Veneto region (Italy) from hospital discharge data (years 1980-1982). *Tumori* 1988; 74(6): 697-704.
11. Brackley ME, Penning MJ, Lesperance ML. In the absence of cancer registry data, is it sensible to assess incidence using hospital separation records? *Int J Equity Health* 2006; 5: 12.
12. Simonato L, Baldi I, Balzi D et al. Obiettivi, strumenti e metodi per un utilizzo epidemiologico di archivi sanitari elettronici correnti in diverse aree italiane. *Epidemiol Prev* 2008; 32(3): 5-14.
13. Fritz A, Percy C, Jack A et al. *International Classification of Diseases for Oncology, 3rd edition*. Geneva, World Health Organization, 2000.
14. Statistiche demografiche ISTAT ([www.demo.istat.it](http://www.demo.istat.it)).
15. AIRTUM Working Group. I tumori in Italia, rapporto 2006: incidenza, mortalità e stime. *Epidemiol Prev* 2006; 30(1 suppl 2): 62-63.
16. Baldi I, Vicari P, Di Cuozzo D et al. A high positive predictive value algorithm using hospital administrative identified incident cases. *J Clin Epidemiol* 2008; 61(4): 373-79.
17. Simonato L, Canova C, Corrao G et al. Ricerca e sviluppo di algoritmi: la definizione di alcune patologie neoplastiche. *Epidemiol Prev* 2008; 32(3 suppl): 94-96.
18. Warren JL, Feuer E, Potosky AL et al. Use of Medicare hospital and physician data to assess breast cancer incidence. *Med Care* 1999; 37: 445-56.
19. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods part I: comparability, validity and timeliness. *Eur J Cancer* 2009; 45: 747-55.
20. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: principles and methods part II: completeness. *Eur J Cancer* 2009; 45: 756-64.