

# SVILUPPO E VALIDAZIONE DI UN ALGORITMO PER L'IDENTIFICAZIONE DEI NUOVI CASI DI TUMORE MALIGNO DELLA MAMMELLA ATTRAVERSO L'UTILIZZO DEI FLUSSI AMMINISTRATIVI SANITARI REGIONALI

Faragalli Andrea<sup>1</sup>, Iommi Marica<sup>1</sup>, Prospero Emilia<sup>2</sup>, Sarti Donatella<sup>2</sup>, Peconi Chiara<sup>2</sup>, Carle Flavia<sup>1,3</sup>, Gesuita Rosaria<sup>1</sup>

*1-Centro di Epidemiologia, Biostatistica ed Informatica Medica, Università Politecnica delle Marche*

*2-Dipartimento di Scienze Biomediche e Sanità Pubblica, Sezione di Igiene, Medicina di Prevenzione e Sanità Pubblica, Università Politecnica delle Marche*

*3-Agenzia Regionale Sanitaria Regione Marche, Ancona*

## Introduzione

I flussi amministrativi sanitari regionali sono una fonte semplice ed economica per rilevare e monitorare patologie che richiedono un contatto con il Servizio Sanitario Nazionale (SSN).

Per ottenere misure epidemiologiche accurate è necessario implementare algoritmi di identificazione dei casi, collegando tra loro più flussi amministrativi, e procedere alla loro validazione mediante un confronto con fonti di dati di riferimento.

Nel caso del tumore della mammella (TM), la scheda di dimissione ospedaliera (SDO) è una delle fonti secondarie utilizzata per individuare i nuovi casi dal registro tumori (RT), che rappresenta lo standard di riferimento per la registrazione dei casi incidenti; nella Regione Marche, il registro è presente dal 2010.

## Obiettivi

Validazione dell'algoritmo di identificazione dei nuovi casi di TM basato sui flussi amministrativi sanitari regionali mediante il confronto con i nuovi casi riportati nel RT della Regione Marche per il 2017.

## Metodi

La popolazione target è rappresentata dalle donne maggiorenni residenti ed assistite nella Regione Marche nel 2017. I nuovi casi di TM sono stati identificati effettuando un record-linkage di tipo deterministico fra Anagrafe Assistiti, SDO e Specialistica Ambulatoriale. Sono state incluse le donne che presentavano per la prima volta nel 2017 una SDO (ricovero indice) con diagnosi primaria o secondaria di tumore maligno della mammella o carcinoma in situ della mammella (ICD-9-CM 174.\* , 233.0). Per ciascun caso di TM sono state individuate le prestazioni (da SDO o da Specialistica Ambulatoriale) di mammografia, ecografia mammaria, o biopsia mammaria nei dodici mesi precedenti la data del ricovero indice. La data indice è la data relativa alla prestazione più prossima al ricovero indice o, in assenza, la data del ricovero stesso. Sono state escluse le donne che alla data indice risultavano non residenti in Anagrafe Assistiti, le donne con almeno un ricovero per anamnesi di tumore della mammella o tumore non specifico della mammella o per malattia metastatica ab initio (ICD-9-CM V10.3, 238.3 e 239.3, 197\*, 198\*, 199\*) in diagnosi principale o secondaria nel ricovero indice o nel periodo compreso tra il 01/01/2011 e la data indice.

I nuovi casi di TM registrati nel RT (ICD-O-3 C50.\* /3, C50.\* /2) sono stati collegati attraverso un record linkage deterministico con quelli individuati dall'algoritmo utilizzando il codice identificativo unico del paziente. L'anno di diagnosi del RT è stato confrontato con l'anno identificato dall'algoritmo, al fine di valutare la capacità dei flussi amministrativi di datare correttamente i nuovi casi di TM. Sono stati stimati la percentuale di accordo (PA, rapporto fra i casi individuati contemporaneamente da RT e dall'algoritmo sul totale dei casi individuati da RT e dall'algoritmo), la sensibilità (rapporto fra i casi individuati contemporaneamente da RT e dall'algoritmo sul totale dei casi individuati da RT) e il valore predittivo positivo (VPP, rapporto tra i nuovi casi identificati contemporaneamente da RT e dall'algoritmo sul totale dei casi individuati dall'algoritmo) e i relativi intervalli di confidenza al 95%.

## Risultati

Nel 2017 sono state individuate dai flussi amministrativi 1645 donne dimesse per la prima volta con una diagnosi di TM; di queste, ne sono state escluse 68 poiché presentavano almeno un ricovero con diagnosi di anamnesi di tumore della mammella, 23 con almeno un ricovero per tumore non specifico della mammella e 65 con almeno un ricovero per malattia metastatica ab initio nel ricovero indice o nel periodo compreso tra il 01/01/2011 e la data indice. L'algoritmo ha individuato 1489 nuovi casi di TM nel 2017; per lo stesso anno, il RT riportava 1559 nuovi casi.

I casi incidenti correttamente identificati dall'algoritmo sono 1252 (PA = 69.7% IC95% 67.5-71.8, Sensibilità = 80.3% IC95% 78.2-82.2, PPV = 84.1% IC95% 82.1-85.9); per 122 donne l'anno di diagnosi non concordava tra le due fonti (Tabella 1); 113 sono state individuate come incidenti dal RT, ma escluse dall'algoritmo per i criteri di eleggibilità; 108 erano registrate nel RT ma non sono state tracciate dall'algoritmo; 201 sono state individuate solo dall'algoritmo e non erano presenti nel RT. Questi ultimi due gruppi sono stati oggetto di una seconda valutazione: del gruppo di 108 casi presenti solo nel RT, 4 casi sono stati rettificati perché 3 non erano residenti al momento della diagnosi e 1 caso era incidente nel 2002; del gruppo dei 201 casi identificati dall'algoritmo, 14 sono stati confermati come incidenti nel 2017 dal RT, 32 risultavano incidenti nell'anno successivo, mentre 122 erano casi di recidiva. Da questa seconda valutazione, i casi incidenti correttamente identificati dall'algoritmo sono risultati 1266 (PA = 70.6% IC95% 68.4-72.7, Sensibilità = 81.2% IC95% 79.2-83.1, PPV = 85.0% IC95% 83.1-86.8).

Dei casi presenti solo nel RT, per 33 è stata individuata la SDO relativa alla diagnosi per TM, ma il codice identificativo unico risultava erroneamente compilato, mentre, per 22 è stato erroneamente riportato un codice ICD-9-CM diverso da 174.\* o 233.0; 49 donne non risultavano tracciate dai flussi amministrativi, in quanto, come riportato da RT, avevano effettuato solo una biopsia (istologia), oppure erano decedute (scheda di morte).

### **Conclusioni**

I flussi amministrativi possono essere considerati fonti accurate per l'identificazione dei nuovi casi di TM in assenza di un registro di patologia. Oltre che per la conduzione di studi epidemiologici, questo algoritmo può essere utile come base per implementare registri di patologia nelle regioni in cui questi ultimi non sono presenti. La capacità di identificazione dell'algoritmo è condizionata dalla qualità e completezza del dato amministrativo; a tal fine può essere utile approfondire i fattori che determinano la discordanza fra le due fonti e convalidare ulteriormente l'algoritmo utilizzando dati di altre regioni.

**Tabella 1. Confronto dei casi incidenti di Tumore della Mammella identificati dall'algoritmo basato sui flussi amministrativi sanitari regionali ed il Registro Tumori della Regione (RT) Marche.**

	1° valutazione	2° valutazione
	n	n
Casi incidenti identificati dall'algoritmo	1489	1489
individuati nel registro tumori, incidenti nel 2017	1252	1266
individuati nel registro tumori, incidenti in un anno differente	122	277
<i>incidenti nel 2017 dall'algoritmo, registrati prima del 2017 dal RT</i>	36	159
<i>incidenti nel 2017 dall'algoritmo, registrati dopo il 2017 dal RT</i>	-	32
<i>incidenti nel 2017 dal RT, individuati dall'algoritmo prima del 2017</i>	51	51
<i>incidenti nel 2017 dal RT, individuati dall'algoritmo dopo il 2017</i>	35	35
individuati nel registro tumori, incidenti nel 2017, non rispettavano i criteri di inclusione	113	113
individuati solo dall'algoritmo	201	33 <sup>#</sup>
Casi incidenti identificati dal RT ma non tracciati dall'algoritmo	108	104*
Percentuale di accordo	69.7%	70.6%
Sensibilità	80.3%	81.2%
Valore Predittivo Positivo	84.1%	85.0%

\*3 non residenti al momento della diagnosi e 1 caso incidente nel 2002; # 14 casi confermati come incidenti nel 2017 dal RT, 32 risultavano incidenti nell'anno successivo, 122 erano casi di recidiva.