

Corso per operatori dei Registri Tumori – Reggio Emilia, 3-6 dicembre 2007

IL CONTROLLO DEI FLUSSI INFORMATIVI

Sandro Tognazzo

Registro Tumori del Veneto

Istituto Oncologico Veneto - IRCCS

INTRODUZIONE

Per ottenere informazioni sulla diagnosi, i trattamenti, lo stato vitale, la residenza e l'identità dei casi, un Registro tumori utilizza in maniera crescente archivi elettronici, provenienti da servizi ed strutture esterne e contenenti dati raccolti ed elaborati per finalità diverse dalla registrazione dei tumori.

PRINCIPALI FLUSSI D'INTERESSE PER UN REGISTRO TUMORI

FONTI DIAGNOSTICHE

- Schede di dimissione ospedaliera (SDO)
- Schede di morte ISTAT
- Referti anatomo patologici
- Referti di radioterapia, radiologia e di laboratorio

TRATTAMENTI E TERAPIE

- Radioterapia
- Prestazioni ambulatoriali
- Prescrizioni farmaceutiche

PRATICHE AMMINISTRATIVE

- Esenzioni ticket
- Anagrafe assistibili

Quando si acquisisce un flusso informativo, bisogna sempre cercare di **rispondere a due domande**:

- **è un flusso completo** ? ossia, comprende davvero tutti i soggetti contattati e l'attività svolta dalla struttura nel periodo di riferimento? e le informazioni che servono, sono sempre presenti?
- **contiene informazioni corrette**? ossia, quello che è registrato corrisponde al vero, è ragionevole e coerente?

Rispondendo a tali domande, possiamo stabilire l'utilità del flusso e la necessità di integrarlo con altre fonti informative.

STRUTTURA DI UN FLUSSO INFORMATIVO

In termini generali, un flusso informativo è composto da **unità di informazione**, denominate **record** o **entità** nella terminologia informatica.

Ogni unità è composto da una serie di elementi informativi o **“item”**, denominati **campi, variabili** o **attributi** nella terminologia informatica

STRUTTURA DI UN FLUSSO INFORMATIVO

Gli “item” possono essere:

- **numerici**: quantità, misurazioni, importi, date
- **non numerici ma codificati** secondo un sistema di classificazione (es. ICD-IX, ICD-O, SNOMED, DRG,...)
- **non numerici e non codificati**

Per gli “item” numerici, bisogna verificare che i valori indicati siano possibili e verosimili (**correttezza**).

Per gli “item” codificati, bisogna verificare che i codici siano:

- **validi**, cioè rientrino tra quelli previsti (**correttezza**);
- **appropriati**, cioè vengano usati col significato previsto dal sistema di codifica (**correttezza**)

Per tutti gli “item”, bisogna verificare se sono effettivamente riportati (**completezza**)

STRUTTURA DI UN FLUSSO INFORMATIVO

ESEMPI

SDO (Schede di dimissione ospedaliera)

un **record** = un **ricovero**

campi = <nome cognome> <nome> <data nascita> <comune nascita> ...
<n. accettazione> <reparto> <data ingresso> <data dimissione>
<diagnosi principale> <diag secondaria> ...

REFERTI ANATOMO PATOLOGICI

un **record** = un **referto**

Campi = <nome cognome> <nome> <data nascita> <comune nascita> ...
<n.referto> <data prelievo> <data refertazione> <tipo esame>
<materiale> <diagnosi microscopica> ...

(tipo campo: numerico; codificato; non codificato)

TIPI DI CONTROLLO EFFETTUABILI

	Completezza	Correttezza
ARCHIVIO COMPLESSIVO	si	no
ITEM NUMERICI	si	si
ITEM CODIFICATI	si	si
ITEM NON CODIFICATI	si	(si)

ALCUNI CONTROLLI TIPICI

	Completezza	Correttezza
ARCHIVIO COMPLESSIVO	<ul style="list-style-type: none">•N. record per anno e mese•Salti nella sequenza numeri progressivi•Numero rec. per tipo di esame/ tipo ricovero/ reparto...	
ITEM NUMERICI	Frequenza valori mancanti	<ul style="list-style-type: none">•Valori esterni al “range” possibile• Frequenza outlier (valori poco probabili)
ITEM CODIFICATI	<ul style="list-style-type: none">•Frequenza valori mancanti•Frequenza valori fuori formato	<ul style="list-style-type: none">•Codici inesistenti•Frequenza di uso di ciascun codice•Codici incompatibili con altri item (es. sesso–diagnosi)
ITEM NON CODIFICATI	come sopra	Caratteri non validi (es. presenza cifre su cognome)

NOTA SUI CONTROLLI DI COMPLETEZZA

Per l'**archivio complessivo**, servono informazioni storiche e/o esterne, sul numero atteso di eventi e/o individui per i flussi da validare.

Ad es., per valutare :

- un flusso schede di morte, si possono utilizzare i dati ISTAT sul movimento anagrafico dei Comuni
- un archivio di referti, si possono utilizzare i volumi storici riscontrati per lo stesso Servizio

NOTA SUGLI ITEM IDENTIFICATIVI PERSONALI

Comprendono la chiave anagrafica (nominativo, sesso, data e comune di nascita) ed i codici fiscale e sanitario.

Vanno controllati in modo non superficiale e con priorità, perché un grado di completezza scarso (ad es. nessun codice e mancanza di data e/o comune di nascita)

- rende difficili le ricerche manuali nell'archivio;
- può precludere l'eventuale integrazione automatica (effettuata mediante procedure di record linkage) con le informazioni della stessa persona contenute in altri archivi.

ESEMPIO 1 – CONTROLLI SUL FLUSSO SDO

RIEPILOGO GENERALE (tabella Controlli SDO-1)

- Viene prodotta **una tabella per ogni regime** di ricovero (ordinario, day hospital, non indicato), con una riga **per ciascun istituto ospedaliero** vengono calcolati i seguenti indicatori.
- Per le sole **SDO che riferiscono diagnosi di tumore** (ICD IX 140-239):
 - numero assoluto distinto per sesso;
 - percentuale sul numero totale SDO;
 - percentuale con codice sanitario valido (non mancante e con lunghezza 9 caratteri);
 - percentuale con cognome nome validi (entrambi non vuoti e diversi da “NP”);
 - percentuale con data nascita valida (giorno 1-31, mese 1-12, 1900<=anno <= anno riferimento); numero di mesi mancanti (nessuna SDO con data dimissione nel mese);
- Per le sole **SDO che NON riferiscono diagnosi di tumore** (ICD IX diverso da 140-239):
 - numero assoluto;
 - percentuale con codice sanitario valido (non mancante e con lunghezza 9 caratteri);
 - numero di mesi mancanti (nessuna SDO con data dimissione nel mese).

ESEMPIO 1 – CONTROLLI SUL FLUSSO SDO

RIEPILOGO GENERALE

DENOMINAZIONE	Numero SDO totale	CON DIAGNOSI DI TUMORE								SENZA DIAGNOSI DI TUMORE			
		Numero SDO	%	Maschi	Femmine	I	C/NOME	D NASCITA	Mesi mancanti	Numero SDO	%	errori su ID	Mesi mancanti
	14440	2782	19.3	1495	1287					11658	80.7	1	
	435	29	6.7	13	16				11	406	93.3		
	2164	265	12.2	146	119					1899	87.8		
	1092	6	0.5	1	5				12	1086	99.5		
	2726	346	12.7	156	190					2380	87.3		
	12939	1790	13.8	1003	787					11149	86.2		
	869	12	1.4	6	6				12	857	98.6		
	3251	275	8.5	141	134					2976	91.5		
	19308	2705	14.0	1426	1279					16603	86.0		
	298	6	2.0	1	5				12	292	98.0		
	11615	1473	12.7	815	658					10142	87.3	1	
	11275	1239	11.0	496	743					10036	89.0		
	9208	1584	17.2	590	994					7624	82.8	1	
	3152	368	11.7	196	172					2784	88.3		
	5222	305	5.8	138	167					4917	94.2		
	8486	988	11.6	433	555					7498	88.4		
	42029	8671	20.6	4232	4439	.				33358	79.4	5	
	6834	683	10.0	294	389	.				6151	90.0	1	

ESEMPIO 1 – CONTROLLI SUL FLUSSO SDO

RIEPILOGHI ANALITICI (tabella Controlli SDO-2)

Tali tabelle riguardano solo le **SDO che riferiscono diagnosi di tumore** (ICD IX 140-239).

Viene prodotta **una tabella per ciascun regime di ricovero e istituto ospedaliero, con una riga per ogni reparto** dove si riportano i seguenti indicatori:

- numero SDO con diagnosi maligne, pazienti maschi;
- numero SDO con sole diagnosi non maligne, pazienti maschi;
- codici ICD9 di tumore maligno primario più frequenti (primi due) tra i pazienti maschi;
- numero SDO con diagnosi maligne, pazienti femmine;
- numero SDO con sole diagnosi non maligne, pazienti femmine;
- codici ICD9 di tumore maligno primario più frequenti (primi due) tra i pazienti femmine;
- degenza media (tutti i pazienti);
- percentuale di diagnosi tumorali non valide o incompatibili con sesso (1);
- percentuale valori mancanti o non validi (lunghezza <> 9 caratteri) su codice sanitario;
- percentuale valori mancanti o non validi su campi data ingresso e data dimissione (2)
- percentuale valori mancanti su campo numero scheda;
- percentuale valori mancanti su campo modalità di dimissione;
- percentuale valori mancanti su campo comune di residenza;
- percentuale valori mancanti su campi DRG e importo;

Sempre per ciascun regime di ricovero e istituto ospedaliero, per ogni reparto viene inoltre prodotta la distribuzione per mese del numero di SDO, distinta per sesso.

(1) Sono non valide le diagnosi ICD IX 166-169, 176-178, 209; sono incompatibili con sesso maschile: diagnosi ICD IX 174, 179-184, 217-221, 233.0-233.3, 236.0-236.3; femminile: diagnosi ICD IX 175, 185-187, 222, 233.4-233.6, 236.4-236.6 .

(2) Sono valide le date con giorno 1-31, mese 1-12, anno (anno di riferimento)-(anno di riferimento – 1), data dimissione non precedente data ingresso, giorni di permesso inferiori ai giorni di degenza

ESEMPIO 2 – CONTROLLI SUI REFERTI ANATOMO-PATOLOGICI

Tra le tre fonti diagnostiche principali, è quella più complessa da validare.

Controlli di **completezza**:

- presenza di interruzioni nella serie dei numeri progressivi dei referti
- distribuzione per tipo di esame (citologia, istologia,...), mese ed anno del prelievo

Controlli di **correttezza** sulla codifica:

- individuazione codici non esistenti e personali
- distribuzione di frequenza codici topografici e morfologici
- correttezza sintattica della codifica (nel caso dello SNOMED)

Importante ma non semplice rilevare codifiche formalmente ammissibili ma usate in modo scorretto (es., tumori sistematicamente codificati con comportamento sbagliato)

ESEMPIO 2 – CONTROLLI SUI REFERTI ANATOMO-PATOLOGICI

La nomenclatura SNOMED permette di codificare con una sintassi precisa :

- TUTTE le informazioni di un referto
- qualsiasi tipo di referto, non solo quelli anatomo-patologici

La pratica reale di codifica generalmente utilizza in modo molto ridotto tali possibilità. Il tipo di frase sintatticamente corretta generalmente utilizzata in pratica è:

PQTMM AND PQTMM ...

T=topografia

obbligatorio

M=morfologia

obbligatorio

P=procedura diagnostica (biopsia, agoaspirato,...)

opzionale

Q=qualificatore (diagnosi probabile, possibile, differenziale,...)

opzionale

Morfologia	Frequenza	Percentuale	Morfologia	Frequenza	Percentuale	Morfologia	Frequenza	Percentuale
M8140/3	5.076	9,1%	M814X/3	491	0,9%	M8780/0	192	0,3%
M8211/0	4.631	8,3%	M8010/4	473	0,8%	M8740/0	191	0,3%
M8760/0	1.838	3,3%	M8000/6	471	0,8%	M8720/3	190	0,3%
M8070/3	1.823	3,3%	M8140/0	466	0,8%	M8052/0	186	0,3%
M8750/0	1.721	3,1%	M8832/0	463	0,8%	M8441/0	181	0,3%
M8130/2	1.691	3,0%	M8070/4	452	0,8%	M8240/1	171	0,3%
M8890/0	1.580	2,8%	M9730/3	404	0,7%	M8470/0	171	0,3%
M8130/3	1.467	2,6%	M9120/0	394	0,7%	M8890/3	163	0,3%
M8010/3	1.464	2,6%	M9621/3	385	0,7%	M8940/0	160	0,3%
M8140/6	1.332	2,4%	M9591/3	384	0,7%	M9823/3	155	0,3%
M9591/8	1.136	2,0%	M8312/3	381	0,7%	M8500/3	152	0,3%
M8263/0	1.119	2,0%	M8050/0	379	0,7%	M9084/0	137	0,2%
M8500/4	1.010	1,8%	M814Y/3	336	0,6%	M8130/5	136	0,2%
M8090/3	981	1,8%	M9590/3	335	0,6%	M9611/3	133	0,2%
M8850/0	888	1,6%	M8130/4	333	0,6%	M8520/2	130	0,2%
M9010/0	764	1,4%	M8261/1	319	0,6%	M9820/3	125	0,2%
M8010/6	748	1,3%	M8070/6	318	0,6%	M8000/3	120	0,2%
M9990/6	748	1,3%	M9632/3	316	0,6%	M8500/6	117	0,2%
M9591/6	735	1,3%	M8170/3	278	0,5%	M8020/3	113	0,2%
M8120/2	686	1,2%	M9612/3	273	0,5%	M8800/3	111	0,2%
M8120/4	682	1,2%	M8010/2	263	0,5%	M8720/6	107	0,2%
M8120/3	657	1,2%	M9656/3	257	0,5%	M8490/3	106	0,2%
M9650/8	644	1,2%	M8520/4	247	0,4%	M9540/0	103	0,2%
M8500/2	607	1,1%	M9940/3	236	0,4%	M814X/6	102	0,2%
M8810/0	563	1,0%	M9692/3	209	0,4%	ALTRI	7.394	13,2%
M9960/1	530	0,9%	M9620/3	206	0,4%	TOTALE	55.939	
M8050/3	503	0,9%	M9560/0	203	0,4%			

LISTA CODICI MORFOLOGICI SNOMED ANOMALI RISCONTRATI

Codice	Frequenza		Codice	Frequenza
M814X3	626		MY0188	7
M814X6	469		MY0550	11
M814Y3	220		MY4373	31
M814Y6	234		MY5204	3
M84011	2		MY6214	1
M87670	199		MY6216	2
M93607	4	recidiva di pinealoma ?	MY6218	3
M95313	3		MY6357	1
M95901	16	linfoma comportamento incerto?	MY7400	4
M97410	4		MYYYY14	2
MX0550	2		MYYYYYY	280

ESEMPIO 3 – SCHEDE DI MORTE

ANNO	N. SCHEDE PERVENUTE	N. DECESSI (ISTAT)
1	47.021	42.044
2	38.813	40.972

Nel primo anno, 12.561 schede duplicate

Schede effettive: 40.741