



Inserto BEN

Bollettino Epidemiologico Nazionale

Editoriale

Roberto Raschetti

Laboratorio di Epidemiologia e Biostatistica, ISS

"Each person in the world creates a book of life. This book starts with birth and ends with death. Its pages are made up of the records of the principal events in the life. Record linkage is the name given to the process of assembling the pages of this book into a volume"

HL Dunn, 1946

Il concetto di "record linkage" non è nuovo e nasce prima della diffusione capillare della tecnologia informatica.

I principali motivi di interesse risiedono nelle grandi dimensioni delle popolazioni di riferimento, nella sistematicità nei processi di acquisizione e nella loro integrabilità che consentono di aumentare il contenuto informativo dei data-base amministrativi (DBA). A fronte delle grandi potenzialità applicative occorre, però, sottolineare i problemi che si possono incontrare nell'utilizzo per finalità epidemiologiche dei DBA:

- i livelli di copertura: quali sono i livelli di copertura delle popolazioni target e, più precisamente, quali sono le caratteristiche dei soggetti non inclusi?
- la qualità dei dati: in questo caso è lecito attendersi che solo i dati che hanno un diretto impatto sul processo gestionale abbiano un livello di qualità alto e controllabile.

Le esperienze di realizzazione di sistemi informativi molto articolati e complessi in alcune realtà regionali, dove è possibile correlare archivi riferiti alle dimissioni ospedaliere, alla prescrizione di farmaci, alla mortalità, hanno offerto e possono continuare a offrire interessanti opportunità nella realizzazione, ad esempio, di grandi studi osservazionali. Tutto ciò nella consapevolezza dei limiti, non solo tecnico/scientifici, ma anche di quelli etici relativi soprattutto ai meccanismi di tutela della privacy. La navigabilità in insiemi eterogenei di dati personali a carattere sanitario non deve tradursi in questi casi nella serendipicità tipica della navigazione in Internet, ma deve essere regolata da un preciso protocollo di studio.

E proprio dell'utilizzo di dati personali provenienti da archivi eterogenei tratta questo fascicolo del BEN, in cui il primo articolo evidenzia come il collegamento tra diversi archivi informatizzati si basi sulla corretta definizione e sulla scelta di chiavi identificative individuali; il secondo articolo, presentando un uso epidemiologico delle schede di dimissioni ospedaliere, mostra utilizzo e limiti per studiare un problema specifico.

Studi dal territorio

UTILIZZO DI DIFFERENTI CHIAVI IDENTIFICATIVE DI SOGGETTI PRESENTI IN DIVERSI ARCHIVI

**Carlotta Sacerdote¹, Marco Dalmasso²,
Giovannino Ciccone¹, Moreno Demaria³
e Roberto Gnani²**

¹Epidemiologia dei Tumori, ASO San Giovanni Battista, CPO Piemonte, Torino

²Servizio di Epidemiologia, ASL 5, Torino

³Area di Epidemiologia Ambientale, ARPA, Torino

Da tempo l'analisi di informazioni di carattere socio-sanitario utilizza il collegamento di dati relativi alle persone fisiche presenti in archivi di differente provenienza, livello di aggiornamento, accessibilità e qualità (1). In assenza, almeno per la situazione italiana, di un univoco codice in grado di individuare ogni persona in modo indipendente dallo specifico settore di interesse, è utile collegare i diversi archivi informatizzati attraverso la definizione di chiavi identificative ricavate da sottoinsiemi, più o meno ampi, dei dati anagrafici disponibili.

In Italia, esperienze di utilizzo di chiavi identificative estratte da dati di tipo anagrafico sono state effettuate per il collegamento di informazioni relative alla stessa persona presenti su differenti archivi (2-4) o sulla medesima fonte informativa, come nel caso dell'individuazione di successivi ricoveri ospedalieri del medesimo assistito. Lo scopo del presente lavoro è di fornire semplici indicatori di validità di diverse chiavi identificative, tramite un esempio di utilizzo di differenti chiavi sperimentate su tre diversi archivi di interesse epidemiologico.

La valutazione delle caratteristiche di differenti chiavi univoche di identificazione è stata effettuata utilizzando i seguenti archivi:

- Coorte torinese del progetto EPIC (5): contiene informazioni anagrafiche, sugli stili di vita e sulle abitudini alimentari di 10 604 residenti a Torino. Il risultato atteso su questo archivio è che le chiavi identificative risultino univoche;
- Anagrafe del Comune di Torino: contiene i dati relativi a 1 944 080 cittadini iscritti all'Anagrafe comunale alla data del 31 dicembre 1998 (compresi i soggetti deceduti o emigrati a partire dal 1971). Anche utilizzando questo archivio il risultato atteso è che le chiavi identificative risultino univoche;

- Schede di Dimissione Ospedaliera (SDO) del Piemonte nell'anno 1997: includono 923 289 casi di ricovero registrati presso strutture ospedaliere del Piemonte. Mentre nei due archivi precedenti non possono esistere ripetizioni del medesimo soggetto, nel caso delle dimissioni questo evento è frequente. Per misurare la capacità delle differenti chiavi nel riconoscere correttamente questa condizione, è utilizzato il Codice Fiscale (CF) che, in 668 744 ricoveri (corrispondenti al 72,43% del totale), risulta indicato in modo completo e formalmente corretto. L'analisi è stata condotta sui 335 725 casi di ricoveri ripetuti che fanno riferimento a 122 884 differenti persone fisiche, identificate da uno specifico CF. Il risultato atteso utilizzando questo archivio è che le chiavi non risultino diverse per il medesimo soggetto.

Considerando l'impossibilità di utilizzare, almeno per quanto riguarda l'archivio delle dimissioni ospedaliere, l'informazione sul nome, vengono costruite le seguenti chiavi identificative:

- prime 4 lettere del cognome (sono eliminati spazi, apostrofi, accenti), 1 carattere per il sesso, codice del Comune di nascita e data di nascita in forma completa;
- analoga alla a), ma considerando solo le prime 3 lettere del cognome;
- analoga alla a), ma considerando le prime 4 lettere del cognome estratte con algoritmo tipo CF;
- analoga alla c), ma estraendo solo le prime 3 lettere del cognome.

La coorte EPIC e l'Anagrafe di Torino permettono di quantificare la quota di casi per i quali vengono generati, per soggetti diversi, valori non univoci della chiave identificativa; questo valore rappresenta una misura della scarsa specificità delle chiavi.

L'archivio delle dimissioni ospedaliere, dove si assume che i casi con lo stesso CF facciano riferimento alla medesima persona, permette di misurare la quota di casi per i quali non vengono generati gli stessi valori della chiave in corrispondenza del medesimo CF. Anche in questa situazione, questo valore misura la scarsa specificità della chiave.

Nella Tabella vengono presentati, espressi in termini di variazione percentuale, i margini di aumento (per gli archivi della coorte EPIC e dell'Anagrafe) e di diminuzione (per l'archivio delle SDO, rispettivamente per tutte le persone con ricoveri ripetuti e per le persone con ricoveri doppi) dell'indicatore di corretta identificazione al variare della chiave.

Aumentando il numero di caratteri considerati, nel caso dell'archivio della coorte EPIC, i guadagni in termini di corretta attribuzione appaiono contenuti, mentre, a causa della maggiore numerosità dell'archivio, risultano più consistenti nel caso dell'Anagrafe. Le diminuzioni dell'indicatore registrate utilizzando l'archivio delle SDO vanno nella direzione attesa: aumentando il numero di caratteri utilizzati, aumenta la probabilità di generare chiavi diverse per gli stessi soggetti.

Più interessante è il confronto fra le prestazioni registrate dalle distinte coppie di chiavi per le quali vengono utilizzati i primi caratteri del cognome (*a*, *b*) o la loro corrispondente versione costruita utilizzando l'algoritmo del CF (*c*, *d*). Anche in questa analisi, si osservano, sugli archivi EPIC e Anagrafe, aumenti dell'indicatore di corretta attribuzione nel caso si utilizzi il CF. All'opposto, nell'archivio delle SDO, l'utilizzo delle chiavi *c*, *d* fa registrare, rispetto alle chiavi *a*, *b*, consistenti diminuzioni dell'indicatore di corretta attribuzione. Queste variazioni possono essere attribuite alla scarsa qualità della compilazione del cognome, che interessa in maggior misura le consonanti piuttosto che le vocali (non utilizzate nel caso delle chiavi *c*, *d*).

Come atteso, il confronto fra le coppie di chiavi per le quali sono utilizzati 3 o 4 caratteri del cognome, evidenzia come, all'aumentare del numero di caratteri considerati, diminuisca la probabilità di assegnare la medesima chiave a persone distinte ma, parallelamente, aumenti la probabilità di considerare come distinte occorrenze che risultano relative al medesimo soggetto.

La probabilità di generare erroneamente chiavi identiche a partire da individui diversi è inferiore al 2,5% con qualunque chiave (Anagrafe). La percentuale di errore si riduce in media dell'1% utilizzando 4 caratteri invece che 3 e dello 0,5% utilizzando l'algoritmo del CF

Tabella - Variazione della percentuale di corretta attribuzione in rapporto alla modalità di estrazione dei caratteri del cognome e al numero di caratteri considerati

Archivio	Modalità di estrazione dei caratteri del cognome	3 caratteri utilizzati	4 caratteri utilizzati	Differenza percentuale
EPIC	Estrazione dei primi caratteri	99,85	99,91	+ 0,06
	Estrazione tipo Codice Fiscale	99,91	99,92	+ 0,02
	Differenza percentuale	+ 0,06	+ 0,01	
Anagrafe	Estrazione dei primi caratteri	97,71	98,85	+ 1,14
	Estrazione tipo Codice Fiscale	98,31	99,10	+ 0,79
	Differenza percentuale	+ 0,60	+ 0,25	
Schede di Dimissione Ospedaliera: soggetti con ricoveri multipli	Estrazione dei primi caratteri	96,06	95,94	- 0,12
	Estrazione tipo Codice Fiscale	95,95	94,96	- 0,99
	Differenza percentuale	- 0,11	- 0,98	
Schede di Dimissione Ospedaliera: soggetti con due ricoveri	Estrazione dei primi caratteri	96,85	96,76	- 0,09
	Estrazione tipo Codice Fiscale	96,82	96,08	- 0,74
	Differenza percentuale	- 0,03	- 0,68	

rispetto alla semplice estrazione dei primi caratteri del cognome. La probabilità di generare erroneamente chiavi diverse a partire da dati anagrafici dello stesso individuo (archivio SDO) risulta intorno al 4% per tutte le chiavi, tranne quando si utilizzano solo 3 caratteri con l'algoritmo del CF (errore intorno al 5%).

Le due chiavi che minimizzano entrambi gli errori sono quella che utilizza i primi 4 caratteri del cognome e quella che utilizza i primi 3 caratteri con algoritmo tipo CF. La scelta tra le due chiavi dipenderà dal tipo di dati che vengono linkati e dal grado di completezza delle variabili anagrafiche delle fonti informative utilizzate.

Riferimenti bibliografici

1. Rosso S. Archivi e liste di popolazione: accessibilità, completezza, aggiornamento. In: Atti del Convegno della Associazione Italiana di Epidemiologia, 1986.
2. Lagorio S, Forastiere F, Michelozzo P, et al. Accertamento delle cause di morte in studi di follow up: confronto di procedure utilizzabili in Italia. *Epid Prev* 1987; 31: 57-61.
3. Costa G, Demaria M, Bisanti L, et al. Uso di dati amministrativi per la ricerca epidemiologica. La consultazione dell'archivio dei codici fiscali per l'accertamento di esistenza in vita negli studi di coorte. *Epid Prev* 1988; 35: 40-6.
4. Costa G, Demaria M. Un sistema longitudinale di sorveglianza della mortalità secondo le caratteristiche socio-economiche, come rilevate ai censimenti di popolazione: descrizione e documentazione del sistema. *Epid Prev* 1988; 36: 37-47.
5. Riboli E, Kaaks R. The EPIC project: rationale and study design. *Int J Epidemiol* 1997; 26(suppl. 1): S6-14.

METODO PER IL MONITORAGGIO DELL'ENDEMIAM GOZZIGENA

Enrico de Campora¹, Sergio Lodato¹
e Renato Pizzuti²

¹Agenzia Regionale Sanitaria della Campania, Napoli

²Osservatorio Epidemiologico Regionale della Campania, Napoli

Il gozzo tiroideo rappresenta la più frequente manifestazione clinica della carenza nutrizionale di iodio (ID). Lo iodio è un elemento essenziale per la crescita e lo sviluppo umano e il gozzo endemico rappresenta una malattia d'adattamento, conseguente alla stimolazione cronica della ghiandola tiroide da parte della tireotropina. Nelle popolazioni esposte a carenza di iodio si possono avere anche altre manifestazioni, quali aumentata tendenza all'aborto, aumento della natalità, ipotiroidismo neonatale quasi sempre transitorio, deficit neuropsicologici e cognitivi minori fino al cretinismo, disturbi definiti come disordini da carenza iodica (1). Il gozzo è normalmente più frequente nelle zone collinose o montane dove i livelli di iodio nel suolo sono più bassi.

Le manifestazioni cliniche del gozzo sono in genere di scarso rilievo, e, nelle fasi iniziali, sono rappresentate da un semplice aumento di volume della ghiandola tiroidea. Nei gozzi di lunga durata si pos-

sono invece osservare evoluzioni di tipo nodulare e iperfunzione, che rappresentano le complicanze più frequenti del gozzo endemico (2).

La iodoprofilassi, in grado di eliminare le malattie da carenza iodica, è ancora inadeguata in Italia, in quanto il sale iodato consumato è meno del 3% di tutto il sale alimentare (3).

Attualmente per monitorare la prevalenza e la distribuzione territoriale della ID si effettuano periodici screening della popolazione in età scolastica. Le metodologie adottate sono: palpazione della tiroide, ecografia della tiroide e determinazione della ioduria.

Il gozzo viene definito endemico se si riscontra in più del 5% della popolazione totale o in più del 10% della popolazione in età scolare (2). In una review delle indagini epidemiologiche condotte in Italia negli ultimi 20 anni, in modo conforme alle raccomandazioni del WHO, è stato rilevato che la prevalenza del gozzo nella popolazione giovanile superava il 20%. In alcune località della Campania la prevalenza raggiungeva il 73% dei giovani esaminati (4).

In uno studio svolto nel 1993-96 è stata misurata l'escrezione urinaria di iodio (EUI) su campioni raccolti da 3 480 soggetti (sia adulti che bambini) di un vasto territorio della regione Campania che comprendeva aree urbane ed extraurbane. La media dell'EUI è risultata significativamente più bassa nei soggetti residenti in paesi della provincia di Benevento, Avellino e Caserta rispetto a Napoli (5). Gli autori concludevano che in molte aree della regione è presente una deficienza di iodio lieve-moderata e che quindi un programma di iodoprofilassi non fosse più rinviabile.

Le metodologie di screening attualmente in uso richiedono tuttavia un alto livello di standardizzazione delle tecniche per garantire precisione, accuratezza dei test e riproducibilità dei risultati. Inoltre, i costi delle campagne e le difficoltà organizzative connesse comportano rilevanti ostacoli all'effettuazione di periodici survey condotti dalle autorità sanitarie locali.

Allo scopo di valutare se la frequenza di ricoveri ospedalieri per "Gozzo tiroideo non tossico" possa fornire indicazioni utili alla localizzazione geografica delle aree endemiche sono state analizzate le Schede di Dimissione Ospedaliera (SDO); sono state selezionate le schede relative ai residenti in Campania ovunque ricoverati nel corso del 2000 (circa 1 191 000 in Campania e 93 000 in altre regioni), che riportassero come diagnosi principale i codici della ICD9 CM relativi al "Gozzo semplice o non specificato" (compresi tra 240.0 e 241.9). Come confronto sono stati inoltre estratti i ricoveri ordinari dell'anno 2000 avvenuti in Italia con le medesime diagnosi principali (<http://www.ministerosalute.it/linksanita/sdo/dati/agggregati/default.htm> - ultima consultazione 16.9.02).

Le popolazioni per il calcolo dei tassi sono quelle ISTAT relative all'anno 2000.

Nell'anno 2000 in Italia risultano 28 577 ricoveri ordinari con diagnosi principale di gozzo semplice, con un tasso di ricovero ordinario di 49,5 ricoveri ogni 100 000 abitanti. Il tasso corrispondente dei residenti in Campania, ottenuto con il metodo della standardizzazione di-

Tabella - Ricoveri per gozzo tiroideo, Campania 2000

ASL	Tassi per 100 000	Percentili	Rischio relativo
Benevento 1	217,4	> 95°	2,6
Salerno 3	189,3	> 90° < 95°	2,3
Napoli 4	184,4	> 75° < 90°	2,2
Napoli 5	171,8	= 75°	2,0
Napoli 1	171,8	> 25° < 75°	2,0
Napoli 3	155,7	> 25° < 75°	1,9
Avellino 2	145,8	> 25° < 75°	1,7
Napoli 2	127,5	> 25° < 75°	1,5
Salerno 2	118,5	> 25° < 75°	1,4
Avellino 1	116,5	= 25°	1,4
Caserta 1	105,7	< 25°	1,3
Caserta 2	103,4	< 25°	1,2
Salerno 1	84,0	< 25°	1,0

retta, è di 79,7 (1,6 volte il tasso italiano). Tra le persone cui è stato diagnosticato il gozzo semplice o non specificato in Italia, il 59% aveva una diagnosi principale da gozzo multinodulare non tossico (ICD-9 421.1); in Campania, il valore corrispondente è del 57%.

In Campania, i ricoveri totali documentati con diagnosi principale di gozzo tiroideo sono risultati 8 588 (0,7% di tutti i ricoveri dei cittadini campani), la spesa ospedaliera relativa per l'assistenza era, per l'anno 2000, pari a 10,5 milioni di euro. Il 48% dei ricoveri totali è costituito da ricoveri ordinari (3 883 soggetti con 1,1 ricoveri in media) e il 47% dei ricoveri ordinari ha subito un intervento chirurgico.

L'80% dei soggetti con almeno un ricovero ordinario è di sesso femminile, l'età media di 47,9 anni, il 3,2% dei pazienti aveva < 25 anni e 23 soggetti < 15 anni.

In Campania, la distribuzione geografica dei ricoveri è risultata essere molto disomogenea con una particolare concentrazione di casi nelle aree storicamente note per carenza di iodio nel suolo e alta incidenza di gozzo endemico. In particolare, in provincia di Benevento il tasso di ricovero totale per gozzo tiroideo è di 217 casi per 100 000 abitanti (1,5 volte la media regionale). Nella provincia di Salerno due differenti Aziende Sanitarie Locali (ASL) presentano una differenza tra i rispettivi tassi di 2,3 volte: la ASL Salerno 3, che comprende alcune zone montuose a Sud di Salerno, ha un tasso di 189,3 casi per 100 000 abitanti, mentre la ASL Salerno 1 che include la città e il suo hinterland ha un tasso di 84 (Tabella). Il rischio relativo, calcolato sul tasso minore (ASL Salerno 1) rivela che 7 ASL campane su 13 hanno un rapporto maggiore di 1,5, mentre solo 2 superano il 90° percentile.

L'utilizzo delle SDO per la valutazione della frequenza di alcune malattie è prassi consolidata pur con i limiti identificati da alcuni autori. Ovviamente il ricovero, in particolare quello ordinario, rappresenta solo un momento emergente della malattia che, nel caso del gozzo tiroideo, sembra essere legato alla comparsa delle complicanze (evoluzione nodulare) e alla necessità di asportazione chirurgica della ghiandola.

Per poter utilizzare tale metodica quale stimatore diretto del grado di endemia gozzigena sarebbe tuttavia necessario conoscere la proporzione di casi ospedalizzati sul totale dei casi di gozzo tiroideo semplice diagnosticati secondo i criteri WHO sul territorio in esame.

In ogni caso, se vi è una proporzione costante tra casi ambulatoriali e ospedalizzati, la Campania sembra confermarsi quale zona ad alta incidenza gozzigena visto il rapporto di 1,6 tra tasso standardizzato campano e quello nazionale.

I tassi di ricovero per gozzo tiroideo in Campania dimostrano, inoltre, una forte disomogeneità territoriale e la spesa ospedaliera per questa malattia è ancora rilevante nonostante alcune ASL abbiano intrapreso campagne di educazione sanitaria per la supplementazione alimentare con sale iodato.

Il sistema proposto può contribuire a identificare le zone con maggiore prevalenza di gozzo tiroideo e a indirizzare gli interventi locali di educazione sanitaria al fine di eradicare la malattia. Esso potrebbe inoltre essere utilizzato per monitorare, in particolare sui casi giovanili, il successo delle campagne di informazione sanitaria in modo più rapido ed economico della ripetizione degli screening scolastici.

Riferimenti bibliografici

1. Pinchera A, Rago T, Vitti P. Fisiopatologia della carenza iodica. *Ann Ist Super Sanità* 1998; 34(3): 301-5.
2. Macchia PE, Fenzi G. Gozzo endemico: quadro clinico ed evoluzione. *Ann Ist Super Sanità* 1998; 34(3): 307-10.
3. Aghini-Lombardi F, Antonangeli L. Legislazione sulla iodoprofilassi in Italia. *Ann Ist Super Sanità* 1998; 34(3): 363-6.
4. Nasti A, Pisano G, Mariano A, et al. Escrezione urinaria di iodio nella regione Campania. *Ann Ist Super Sanità* 1998; 34(3): 413-6.
5. Aghini-Lombardi F, Antonangeli L, Vitti P. Epidemiologia del gozzo endemico in Italia. *Ann Ist Super Sanità* 1998; 34(3): 311-4.

In breve

Dall'inizio del 2003 sono entrate in vigore le nuove definizioni di caso di malattia infettiva da adottare per la sorveglianza europea. Le definizioni hanno lo scopo di migliorare la confrontabilità dei dati raccolti nei diversi Paesi dell'Unione Europea nell'ambito della rete di sorveglianza epidemiologica decisa dal Parlamento Europeo nel 1998.

Per le nuove definizioni dei casi, si consiglia di consultare il documento: http://www.europa.eu.int/eurlex/pri/it/voj/dat/2002/l_086/l_08620020403it00440062.pdf

Comitato editoriale BEN

Donato Greco, Nancy Binkin, Paolo D'Argenio, Paola De Castro, Carla Faralli, Marina Maggini, Stefania Salmaso

Full English version is available at: www.ben.iss.it
e-Mail: ben@iss.it