

ATTENDIBILITÀ DEI DATABASE AMMINISTRATIVI NELLE INDAGINI EPIDEMIOLOGICHE IN AMBITO NEFROLOGICO: IL CASO DELL'INSUFFICIENZA RENALE CHE NECESSITA TERAPIA SOSTITUTIVA NELLA POPOLAZIONE DIABETICA

M. Nordio¹, F. Antonucci², M. Feriani³, A. Inio⁴, P. Marchini⁵

¹U.O. di Nefrologia e Dialisi, ULSS 15 "Alta Padovana", Camposampiero (PD)

²U.O. di Nefrologia e Dialisi, ULSS, Feltre (BL)

³U.O. di Nefrologia e Dialisi, Ospedale di Mestre, ULSS 12 Veneziana, Mestre (VE)

⁴Dipartimento di Prevenzione, ULSS 12 Veneziana, Venezia

⁵U.O. di Nefrologia e Dialisi, Ospedale Civile di Venezia, ULSS 12 Veneziana, Venezia

Reliability of administrative databases in epidemiological research: the example of end-stage renal disease requiring renal replacement therapy in patients with diabetes

Health and administrative databases are widely used in epidemiology, mostly in studies of chronic diseases, but biases may undermine the external validity of the results. In nephrology, the use of these data sources is in its early days and needs to be validated. The aim of this study was to compare the data on the incidence of ESRD and death in a cohort of patients with type 2 diabetes (DM2) obtained from administrative databases with the results of a traditional, well-performed cohort study.

The study was conducted in the Health District of Venice (Italy) on a cohort of 18,416 DM2 patients on hypoglycemic drug therapy enrolled from 1 January 1998 to 31 December 2002 from administrative databases. Comorbid conditions were recorded from hospital discharge records, the database of death certificates was used to identify patients who died within 31 December 2004, and the database of the Dialysis and Transplantation Registry of the Veneto Region served to identify patients who started renal replacement therapy within 31 December 2004. Record linkage was performed using the unique personal identification codes (fiscal number) of Italian citizens. The cumulative incidence of ESRD was estimated using Gray's method for competing risks.

The mortality rate was 50.95 per 1000 person-years, the ESRD incidence was 0.68 per 1000 person-years, with a relative risk of 2.62 with respect to all other causes of ESRD. The crude cumulative incidence of death was 22% and that of ESRD 0.33% at the end of follow-up. The results were similar to those obtained in traditional cohort studies.

The results of our study prove the external validity of the administrative database approach in epidemiological studies in nephrology. (G Ital Nefrol 2009; 26 (Suppl. S45): S7-11)

Conflict of interest: None

KEY WORDS:

Administrative databases,
Type 2 diabetes mellitus,
ESRD,
Record linkage,
Competitive risk,
External validity

PAROLE CHIAVE:

Database amministrativi,
Diabete mellito tipo 2,
ESRD,
Record linkage,
Rischi competitivi,
Validità esterna

✉ Indirizzo dell'Autore:

Dr. Maurizio Nordio
U.O.A. di Nefrologia e Dialisi
P.O. Camposampiero
Via P. Cosma, 1
35012 Camposampiero (PD)
e-mail: maurizio.nordio@gmail.com

INTRODUZIONE

Il Servizio Sanitario Nazionale raccoglie obbligatoriamente tutta una serie di informazioni anagrafiche, amministrative e sanitarie che costituiscono un flusso di dati che dalle singole Aziende Sanitarie viene inviato all'Amministrazione Regionale, all'ISTAT, ai Ministeri

della Salute e dell'Economia, andando a costituire delle banche dati.

Le banche dati disponibili riguardano dati demografici (età, sesso, stato civile, ecc.), amministrativi (mortalità, ricoveri, ecc.), di sorveglianza (malattie infettive, ADR, ecc.), di prescrizione (consumo di farmaci). Questo flusso informativo ha vari scopi che vanno dalla

sorveglianza, all'analisi dell'utilizzo delle strutture sanitarie, al governo della spesa sanitaria. Tutte queste banche dati hanno però in comune la caratteristica di essere personali, cioè costituite da dati riguardanti singoli individui, identificati con il codice fiscale.

È quindi possibile collegare tra loro le varie banche dati attraverso il codice fiscale del singolo soggetto in modo da condividere le informazioni contenute. Questo processo viene chiamato "record linkage" e viene definito come "l'integrazione di informazioni da due o più fonti indipendenti". I records riguardanti lo stesso individuo sono accoppiati tra loro e possono essere trattati come singolo record.

Il record linkage determina alcune conseguenze, in quanto facilita il controllo tra database indipendenti al fine di superare i problemi di qualità, e soprattutto permette l'utilizzo di questi dati per ricerche epidemiologiche.

La realizzazione di ricerche epidemiologiche utilizzando queste fonti di dati è apparentemente ottimale, in quanto teoricamente dovrebbe coprire l'intera popolazione in esame, ma ad un esame più approfondito si pongono parecchi problemi.

Il primo riguarda la correttezza del codice fiscale, infatti di solito non viene utilizzato il codice fiscale fornito direttamente dall'Autorità competente, ma dei codici calcolati da software che possono fornire risultati non esattamente coincidenti con quello ufficiale, ciò si verifica comunemente per soggetti nati all'estero o in comuni non più esistenti, senza contare la possibilità di errori di imputazione anche con l'uso del codice ufficiale.

Altri errori, o meglio informazioni deviate o incomplete, possono derivare da vincoli strutturali. Ad esempio l'uso della sola prima diagnosi e non delle complicanze nella compilazione delle schede di dimissione ospedaliera per disposizioni dell'Autorità Regionale, determina la perdita di importanti informazioni. La somministrazione diretta o per doppio canale di alcuni farmaci impedisce o, ancor peggio, riduce la loro catalogazione nel database delle prescrizioni farmaceutiche.

Questi e altri bias informativi pongono alla fine problemi di accuratezza e generalizzabilità dei risultati che, da un punto di vista epidemiologico, si traducono in problemi di validità esterna dello studio.

Un altro problema riguarda la possibilità di identificare tutti i soggetti con una determinata malattia (sensibilità) e di scartare quelli non affetti dalla malattia (specificità). In generale la sensibilità della diagnosi è alta nel caso delle malattie croniche (ad esempio il diabete), ma è più bassa nelle patologie a diagnosi più complessa. In Letteratura sono molto scarsi gli studi di valutazione di routine e in particolare quelli in campo nefrologico, dove l'utilizzo dei database sanitari e della tecnica del record linkage è ancora ai primi passi.

Questo studio ha lo scopo di valutare la generalizzabilità dei risultati di incidenza di insufficienza renale che richiede terapia sostitutiva e/o causa morte, ottenuti da una coorte di diabetici di tipo 2 reclutata da database sanitari, confrontandoli con quelli ottenuti da uno studio di coorte ben disegnato e ampiamente validato in Letteratura (1).

MATERIALI E METODI

Lo studio è stato condotto nell'Azienda Sanitaria ULSS 12 Veneziana, un'area urbana di circa 300000 abitanti. I dati sono stati ottenuti dai database sanitari delle prescrizioni farmaceutiche, delle schede di dimissione ospedaliera (SDO), dei certificati di morte e dal Registro di Dialisi Trapianto della Regione Veneto.

I soggetti affetti da diabete mellito di tipo 2 sono stati identificati dal database delle prescrizioni farmaceutiche, per cui solamente i diabetici in trattamento farmacologico costituiscono la popolazione esaminata. Sono stati esclusi i pazienti di età inferiore a 30 anni e trattati solo con insulina per almeno un anno, classificati come diabetici di tipo 1. Poiché le banche dati erano consistenti a partire dal 1998, sono stati arruolati i pazienti prevalenti nel 1998 e i nuovi pazienti diabetici dal 1/1/1999 al 31/12/2002. L'uso di insulina (cod. A10A) e di ipoglicemizzanti orali (A10B) è stato ricavato dal database delle prescrizioni farmaceutiche, da questo inoltre sono stati ottenuti dati sull'uso di ACE-inibitori e antagonisti recettoriali dell'angiotensina II (C09) e di altri farmaci antiipertensivi (C02, C07, C08).

La diagnosi alla dimissione è stata ottenuta dalla banca dati delle SDO per ciascun paziente ricoverato in ospedale tra l'1/1/1998 (inizio del periodo di osservazione) e il verificarsi dell'evento dialisi o morte o fine del periodo di osservazione (31/12/2004). Per stabilire le comorbidità sono stati considerati i codici ICD-9-CM, raggruppando le diagnosi in "infezioni" (cod. 001-139), "neoplasie" (cod. 130-239), cardiopatia ischemica (cod. 410-414), "insufficienza cardiaca" (cod. 428), "malattia vascolare cerebrale (cod. 430-438), "malattia vascolare periferica" (cod. 440-444), "ipertensione" (cod. 401-405), "insufficienza renale cronica" (cod. 585) e "insufficienza renale acuta" (cod. 584-584.9).

Le date di morte sono state recuperate dal database dei certificati di morte tra l'1/1/1998 e il 31/12/2004, mentre sono state ignorate le cause di morte.

Dal Registro Veneto di Dialisi e Trapianto è stata ottenuta la data di inizio del trattamento sostitutivo, inoltre sono state prese in considerazione anche la presenza di diabete mellito come causa di nefropatia o condi-

zione comorbida per controllare la concordanza con gli altri *database* sanitari.

Il *record linkage* con codice fiscale, nome e cognome e data di nascita è stato effettuato presso la struttura aziendale autorizzata al trattamento di tali dati, tabelle anonime sono state poi fornite per l'elaborazione.

Ogni *database* condivideva dati demografici dei pazienti, in particolar sesso ed età. L'età è stata calcolata alla data della prima prescrizione di un farmaco ipoglicemizzante.

I pazienti sono stati seguiti dalla data di reclutamento (dall'1/1/1998 al 31/12/2002) fino al verificarsi di morte o necessità di terapia renale sostitutiva (ESRD) o alla fine del *follow-up* (31/12/2004).

È stato determinato il tasso di incidenza di ESRD per i soggetti con diabete mellito di tipo 2 e per la popolazione generale dell'area geografica dell'ULSS 12 e calcolato il rischio relativo. Il tasso di incidenza di ESRD è stato calcolato come numero di casi di ESRD per pazienti-anno durante il periodo di *follow-up*.

La morte compete con l'ESRD, in quanto i soggetti che muoiono ovviamente non possono sviluppare ESRD, ma non possono essere considerati persi dal *follow-up*, perché hanno un esito diverso che è tra gli obiettivi dello studio. L'incidenza cumulativa di ESRD è stata stimata con un metodo che tiene conto della presenza di rischi competitivi (2).

Il pacchetto "cmprsk" versione 2.1-5 disponibile per il software statistico R è stato utilizzato per l'analisi dell'incidenza cumulativa in presenza di rischi competitivi (funzione "cuminc"), tutte le altre analisi sono state eseguite con STATA, release 9 (STATA Corporation, College Station, Texas, USA).

RISULTATI

Tra l'1/1/1998 e il 31/12/2002 sono stati identificati 18416 pazienti in trattamento con farmaci ipoglicemizzanti, la loro età mediana era di 67 anni (IQR = 59-75), il 51% erano maschi e il 49% femmine, il periodo medio di *follow-up* è stato di 4.85 anni (range: 1 giorno-6 anni).

Il 79.44% dei soggetti era trattato con ipoglicemizzanti orali, il 5.17% solo con insulina e il 15.39% con entrambi.

La distribuzione delle comorbidity ottenuta dalle SDO è riportata in Tabella I.

Il tasso di incidenza del diabete mellito di tipo 2 nell'area considerata è risultato essere 5.95/1000 persone-anno (95% IC: 5.81-6.09). La prevalenza è aumentata dal 3.83% della popolazione residente al 31/12/1998 al 5.19% al 31/12/2002 (Fig. 1).

Durante il *follow-up* sono morti 3901 pazienti on un

TABELLA I - DISTRIBUZIONE DELLE COMORBIDITÀ SECONDO LE DIAGNOSI DI DIMISSIONE OTTENUTE DAL DATABASE DELLE SDO

Comorbidity	Proporzione %	95% IC
Infezioni	3.7	3.4-4.0
Neoplasie	11.7	11.4-12.3
Cardiopatía ischemica	7.5	7.1-8.0
Insufficienza cardiaca	6.3	6.0-6.8
Malattia vascolare cerebrale	10.6	10.2-11.1
Malattia vascolare periferica	5.3	5.0-5.6
Ipertensione	21.3	20.9-22.1
Insufficienza renale cronica	3.8	3.5-4.1
Insufficienza renale acuta	0.3	0.25-0.41

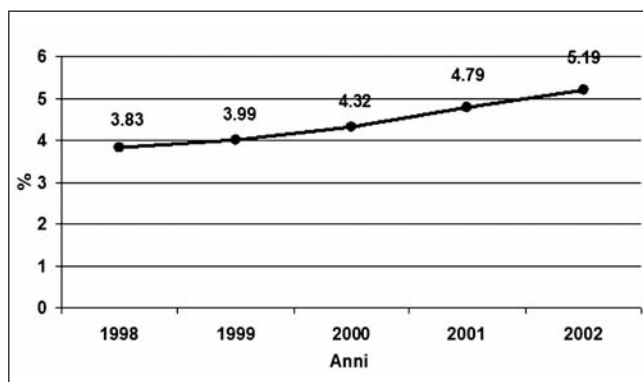


Fig. 1 - Prevalenza del diabete mellito di tipo 2 nell'ULSS 12 Veneziana tra il 1998 e il 2002.

tasso di mortalità di 50.95/1000 persone-anno (95% IC: 49.37-52.57) e un'incidenza cumulativa grezza del 22% a 5 anni.

Solo 59 pazienti hanno sviluppato ESRD durante il *follow-up* con un tasso di incidenza di 0.68/1000 persone-anno (95% IC: 0.53-0.88). Nello stesso periodo il tasso di incidenza per tutte le altre nefropatie tranne il diabete è stato di 0.26/1000 persone-anno (95% IC: 0.24-0.28) con un rischio relativo di ESRD per diabete di tipo 2 di 2.62 (95% IC: 1.97-3.44).

L'incidenza cumulativa di ESRD tenendo conto della morte come rischio competitivo è stata dello 0.33% (95% IC: 0.24-0.42) alla fine del *follow-up*.

DISCUSSIONE

Lo studio si focalizza sul rischio di ESRD e morte nei pazienti con diabete mellito di tipo 2 in trattamento

ipoglicemizzante. Questi pazienti sono facilmente identificabili tramite i *database* sanitari perché gli ipoglicemizzanti hanno come indicazione esclusiva il diabete mellito ed esistono criteri non ambigui di identificazione del diabete mellito di tipo 1.

I pazienti con diabete di tipo 2 non in trattamento farmacologico possono essere difficilmente identificati da *database* sanitari, se non da quelli dei laboratori, ammesso che più determinazioni della glicemia siano state eseguite, pertanto la loro prevalenza può essere stimata solo con studi *ad hoc* o con metodi indiretti come quello di cattura-ricattura (3). Gli studi *ad hoc* sono difficili da condurre e costosi e, anche in questo caso, non c'è la garanzia di coprire tutta la popolazione affetta.

Lo scopo principale di questo studio era di validare l'utilizzo dei *database* amministrativi e sanitari come fonti di indagini epidemiologiche in campo nefrologico.

La prevalenza del diabete di tipo 2 osservata nell'ULSS 12 Veneziana è simile a quella osservata nella vicina Verona (3), in Inghilterra (4, 5), più alta che in Olanda (6) e nel nord del Galles (7), più bassa di quella riportata in USA (8, 9), pur tenendo conto dei diversi anni di osservazione.

L'incidenza di diabete mellito di tipo 2 è comparabile con quella osservata in una regione Spagnola (10) e inferiore a quella osservata in Olanda (6). Questi confronti confermano che i risultati ottenuti attraverso i *database* sanitari e amministrativi sono coerenti con altri studi.

L'incidenza di ESRD non è significativamente diversa da quella riscontrata nella coorte prospettica di Casale Monferrato (1): 0.68/1000 persone-anno (95% IC: 0.53-0.88) vs 1.04 persone-anno (95% IC: 0.56-1.94). La mortalità cumulativa non è confrontabile tra i due studi perché il periodo di osservazione è diverso (5 anni per questo studio e 7 anni per (1)). Comunque l'ordine di grandezza è lo stesso: 22% in questo studio e 39.9% in (1).

Il confronto dei risultati di questo studio con gli altri conferma la fattibilità dell'utilizzo dei *database* sanitari e amministrativi anche in campo nefrologico, garantendo quindi la generalizzabilità dei risultati. In questo caso la scelta della popolazione diabetica offre una particolare facilità di trattamento dei dati, in quanto i pazienti sono ben identificabili e le tecniche di identificazione sono validate da molta letteratura. Inoltre il *linkage* con il Registro Veneto di Dialisi e Trapianto si è mostrato fattibile e ha portato a risultati coerenti.

È auspicabile che in futuro le imponenti fonti di informazione costituite dai *database* sanitari e amministrativi possano fornire un importante contributo alla ricerca epidemiologica in campo nefrologico.

RIASSUNTO

I database sanitari e amministrativi sono largamente usati in epidemiologia soprattutto per le malattie croniche, tuttavia vari bias possono porre dubbi sulla generalizzabilità dei risultati. In campo nefrologico l'applicazione di queste fonti di dati è ancora agli inizi, per cui sono necessarie delle validazioni. Lo scopo di questo studio è di valutare se l'incidenza di ESRD e di morte in una coorte di diabetici di tipo 2 (DM2) ottenuta da database amministrativi è confrontabile con i risultati di uno studio di coorte prospettico tradizionale.

Lo studio è stato condotto presso l'ULSS 12 Veneziana su una coorte di 18416 soggetti affetti da DM2 in trattamento ipoglicemizzante arruolati tra l'1/1/1998 e il 31/12/2002 da database amministrativi. Le comorbidità sono state identificate attraverso le schede di dimissione ospedaliera, il database dei certificati di morte è stato utilizzato per riconoscere i deceduti entro il 31/12/2004 e il database del Registro Veneto di Dialisi e Trapianto ha identificato i pazienti che hanno iniziato dialisi entro il 31/12/2004. Il record linkage è stato effettuato attraverso il codice fiscale. L'incidenza cumulativa di ESRD è stata stimata con il metodo di Gray per rischi competitivi.

Il tasso di mortalità è stato di 50.95/1000 persone-anno, il tasso di ESRD è stato di 0.68/1000 persone-anno con un RR=2.62 rispetto alle altre cause di ESRD. L'incidenza cumulativa di morte è stata del 22%, mentre quella di ESRD dello 0.33%. I risultati sono stati simili a quelli ottenuti in altri studi di coorte con impostazione tradizionale.

I risultati ottenuti mostrano una buona congruenza con differenti fonti di dati, suggerendo così una buona validità esterna delle informazioni ottenute.

DICHIARAZIONE DI CONFLITTO DI INTERESSI

Gli Autori dichiarano di non avere conflitto di interessi

BIBLIOGRAFIA

1. Bruno G, Biggeri A, Merletti F, et al. Low incidence of end-stage renal disease and chronic renal failure in type 2 diabetes: a 10-year prospective study. *Diabetes Care* 2003; 26: 2353-8.
2. Gray JR. A class of K-sample test for comparing cumulative incidence of a competing risk. *Ann Stat* 1988; 16: 1141-54.
3. Verlato G, Muggeo M. Capture-recapture method in the epidemiology of type 2 diabetes: a contribution from the Verona Diabetes Study. *Diabetes Care* 2000; 23: 759-64.
4. Forouhi NG, Merrick D, Goyder E, et al. Diabetes prevalence in England, 2001—estimates from an epidemiological model. *Diabet Med* 2006; 23: 189-97.
5. Congdon P. Estimating diabetes prevalence by small area in England. *J Public Health (Oxf)* 2006; 28: 71-81. Epub 2006 Jan 25.
6. Ubink-Veltmaat LJ, Bilo HJ, Groenier KH, Houweling ST, Rischen RO, Meyboom-de Jong B. Prevalence, incidence and mortality of type 2 diabetes mellitus revisited: a prospective population-based study in The Netherlands (ZODIAC-1). *Eur J Epidemiol* 2003; 18: 793-800.
7. Harvey JN, Craney L, Kelly D. Estimation of the prevalence of diagnosed diabetes from primary care and secondary care source data: comparison of record linkage with capture-recapture analysis. *J Epidemiol Community Health* 2002; 56: 18-23.
8. Engelgau MM, Geiss LS, Saaddine JB, et al. The evolving diabetes burden in the United States. *Ann Intern Med* 2004; 140: 945-50.
9. Harris MI, Flegal KM, Cowie CC, et al. Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. Adults. The Third National Health and Nutrition Examination Survey, 1988-1994. *Diabetes Care* 1998; 21: 518-24.
10. Mata-Cases M, Fernández-Bertolín E, Cos-Claramunt X, et al. Incidence of type 2 diabetes and its diagnosis process in the decade 1991-2000 in a primary health care centre. *Gac Sanit* 2006; 20: 124-31.