

Stima di prevalenza di diabete mellito in una provincia del Lazio attraverso i modelli cattura e ricattura

Estimating prevalence of diabetes mellitus in a Lazio province, Italy, by capture-recapture models

Maria Letizia Giarrizzo, Patrizio Pezzotti, Irene Silvestri, Domenico Di Lallo

Laziosanità - Agenzia di sanità pubblica della Regione Lazio

Corrispondenza: Maria Letizia Giarrizzo, Laziosanità - Agenzia di sanità pubblica della Regione Lazio, via di Santa Costanza 53, 00198 Roma; fax: 0683060463; e-mail: giarrizzo@asplazio.it

Cosa si sapeva già

- La stima della prevalenza del diabete ottenuta attraverso indagini campionarie risulta molto spesso complessa per i costi elevati dovuti all'esecuzione di appropriate indagini diagnostiche.
- Attraverso l'utilizzo di dati provenienti da sistemi informativi e l'applicazione del metodo cattura-ricattura è possibile stimare la prevalenza di diabete noto.

Cosa si aggiunge di nuovo

- Sono state ottenute stime di prevalenza in una provincia del Lazio attraverso l'utilizzo di 4 archivi sanitari e l'applicazione di due metodi cattura-ricattura.
- Le stime ottenute con tre archivi sanitari sono leggermente superiori a quelle ottenute recentemente in altre aree italiane e con altre metodologie di studio.
- Le stime non sembrano risentire del tipo di chiave di linkage utilizzata, né della metodologia statistica applicata.
- L'inclusione di un archivio che raccoglie le prescrizioni di emoglobina glicata ha un forte effetto sulle stime di prevalenza e resta incerta l'appropriatezza nel suo utilizzo per la stima della prevalenza di diabete.

Riassunto

Obiettivi: stimare la prevalenza del diabete mellito (DM) in una provincia del Lazio attraverso l'utilizzo di archivi sanitari e l'applicazione del metodo cattura-ricattura.

Disegno e setting: sono state utilizzate, selezionando solo i residenti della provincia di Viterbo e l'anno 2004, 4 liste provenienti da 3 archivi sanitari: dimissioni ospedaliere della regione Lazio (codice ICD9-CM 250, 2.488 soggetti, SIO-dimissioni); sistema informativo assistenza specialistica della regione Lazio (esenzione per DM, 3.582 soggetti, SIAS-esenzioni); prescrizioni del test per la determinazione dell'emoglobina glicata (7.169 soggetti, SIAS-emoglobina glicata); registro delle esenzioni per patologia della ASL di Viterbo (esenzione per DM, 8.529 soggetti, VT-esenzioni).

È stato effettuato un record-linkage deterministico con 5 diverse chiavi basate sui dati anagrafici della persona. La stima della prevalenza è stata ottenuta mediante l'applicazione di tecniche di cattura-ricattura: modelli log lineari e approccio *sample-coverage*.

Risultati: le procedure di abbinamento attraverso diverse chiavi di linkage hanno prodotto risultati simili. Al variare del tipo di chiave e di tecnica statistica le stime di prevalenza variavano tra il 6,3% e l'8,7% utilizzando tutte e quattro le fonti e tra il 5,7% e il 6,1% escludendo l'archivio SIAS-emoglobina glicata.

Conclusioni: le stime ottenute sono leggermente più elevate di quelle di altri studi pubblicati recentemente; l'inclusione nell'analisi dell'archivio SIAS-emoglobina glicata comporta una certa instabilità e un aumento non trascurabile della stima di prevalenza. Esse non risentono particolarmente della chiave di linkage, né del metodo statistico utilizzato. Gli archivi sanitari che raccolgono routinariamente informazioni amministrative possono essere utilizzati per ottenere una stima di prevalenza del DM attraverso la metodologia cattura-ricattura con una notevole riduzione di tempi e costi rispetto alle indagini tradizionali.

(*Epidemiol Prev* 2007; 31(6): 333-39)

Parole chiave: diabete mellito, record linkage, metodi di cattura-ricattura

Abstract

Objective: to estimate the prevalence of diabetes in a Lazio province, Italy, using a record linkage of four patient lists and applying capture-recapture techniques.

Design and setting: we selected records of individuals living in

Viterbo province (Lazio region, Italy) from four lists obtained from three different health administrative registries in 2004: the hospital discharge registry of Lazio region (ICD9-CM code 250, 2,488 subjects); the registry of outpatient care prescriptions (subjects exempted from medical charge for health care related to diabetes,

3,582 subjects); subjects with prescriptions of glicate haemoglobin (7,169 subjects); the registry of exemptions, from medical charge for health care of the local health unit of Viterbo province, related to diabetes (8,529 subjects). A deterministic record linkage using 5 different keys of linkage obtained from demographic information, was applied. The estimates were derived using two capture-recapture techniques: log-linear models and sample-coverage approach.

Results: record linkage results using the 5 different keys of linkage were similar. Prevalence estimates varied between 6.3% and 7.8% when all four lists were considered and they varied between 5.7% and 6.1% when the glicate haemoglobin list was excluded.

Conclusions: our estimates were slightly higher compared to those recently reported in the literature; estimates were particularly influenced by the inclusion/exclusion of the haemoglobin glicate prescriptions list. Estimates were not influenced either by the key of linkage, or by the capture-recapture technique. Health administrative registries can be used to estimate the prevalence of diabetes through record linkage and capture-recapture techniques. This approach leads to a reduction of time and costs compared with traditional health surveys.

(Epidemiol Prev 2007; 31(6): 333-39)

Key words: diabetes mellitus, record linkage, capture-recapture methods

Introduzione

La prevalenza del diabete mellito (DM) è in costante aumento. Stime e proiezioni effettuate dall'Organizzazione mondiale della sanità (OMS) per nazione prevedono un raddoppio della prevalenza a livello mondiale entro il 2030.¹⁻⁴ Il DM costituisce un importante problema sanitario non solo per l'elevata prevalenza, ma anche per la morbosità e mortalità a cui si associa, correlate prevalentemente alle complicanze croniche della malattia. È stato stimato che il diabete costituisce il 5,2% delle cause di mortalità a livello mondiale,⁵⁻⁷ inoltre, la sua rilevanza è anche correlata al notevole impatto che la malattia ha in termini di impiego di risorse economiche.^{8,9}

La diagnosi di DM si basa sul riscontro di iperglicemia: un soggetto è considerato diabetico se presenta valori di glicemia ≥ 126 mg/dL a digiuno o valori di glicemia ≥ 200 mg/dL due ore dopo un carico di 75g di glucosio. Questo criterio, in assenza di sintomi, può determinare, la mancata diagnosi per diversi anni dall'inizio della malattia.^{10,11} Di conseguenza, la stima della prevalenza del DM non risulta di facile e precisa valutazione.^{12,13}

In Italia vengono effettuate periodicamente dall'ISTAT stime sulla prevalenza di DM dichiarato per regione. Attraverso l'indagine multiscopo 1999-2000,¹⁴ è stata stimata una prevalenza di DM in Italia del 3,7%.

In alcune aree italiane sono stati effettuati studi su registri di popolazione con DM o su archivi sanitari amministrativi che riportano la diagnosi di DM; attraverso tecniche di record-linkage o l'applicazione di metodi cattura e ricattura sono state stimate prevalenze di DM tra 2,4 e 5,3%.¹⁵⁻²¹ Un'indagine condotta nel 1998-2000 ha stimato una prevalenza di DM noto e non noto, attraverso un prelievo capillare su sangue intero effettuato a digiuno e/o la valutazione di trattamento farmacologico per il DM, pari a 8,4% negli uomini e a 6% nelle donne; in particolare ha evidenziato che il 47% dei diabetici non sapeva di esserlo.²²

Tuttavia, ci sono diverse controindicazioni all'effettuazione di uno studio su un campione casuale della popolazione su cui effettuare un prelievo di sangue per la stima di prevalenza di DM quali la verosimile percentuale elevata di rifiuti, i costi, i tempi necessari per la realizzazione dello studio e la difficoltà nel garantire i diritti alla riservatezza.

Questo studio si propone di stimare la prevalenza di DM noto, in tutte le sue forme e per tutte le età, in una provincia del Lazio, attraverso l'utilizzo di quattro fonti amministrative di dati, di valutare la variabilità delle stime ottenute con diverse chiavi di linkage e con diversi modelli di cattura e ricattura.²³⁻²⁵

Materiali e metodi

Le fonti utilizzate

Le analisi sono state condotte con riferimento alla popolazione residente nella ASL di Viterbo, ovvero la popolazione residente nella provincia di Viterbo, per l'anno 2004.

Per identificare le persone con DM noto sono stati utilizzati i seguenti flussi informativi attivi nella regione Lazio:

- il Sistema informativo ospedaliero (SIO), che rileva e gestisce i dati analitici di tutti i ricoveri ospedalieri che ogni anno si registrano negli istituti di ricovero e cura della regione (http://www.asplazio.it/asp_online/att_ospedaliera/sio/sio_ind_ex.php);

- il Sistema informativo per l'assistenza specialistica ambulatoriale (SIAS), che raccoglie le prestazioni di assistenza specialistica erogate in ambito ambulatoriale nella regione (http://www.asplazio.it/asp_online/att_territoriale/sias/csias.php);

- l'archivio della ASL di Viterbo relativo ai soggetti esenti dal pagamento dei farmaci, dei presidi e delle prestazioni sanitarie per patologia diabetica.

Dal SIO sono stati selezionati tutti i ricoveri relativi a soggetti residenti nella provincia di Viterbo (SIO-dimissioni) in cui, per almeno una delle diagnosi, era riportata la categoria ICD9-CM 250 (diabete mellito), per un totale di 2.488 individui.

Dal SIAS sono state estratte due liste distinte di soggetti residenti nella provincia di Viterbo: la prima (SIAS-esenzioni) relativa a 3.582 soggetti che avevano fruito di prestazioni specialistiche ambulatoriali con esenzione dal pagamento del ticket per DM, la seconda (SIAS-emoglobina glicata) relativa a 7.169 soggetti che avevano prescrizioni per il dosaggio dell'emoglobina glicata.

L'archivio degli esenti per DM (VT-esenzioni) al 31.12.2003 era costituito da 8.529 soggetti; al momento dell'analisi non era disponibile un archivio più aggiornato.

Va precisato che, sebbene siano state considerate nell'analisi due liste di esenzioni per DM, esse non sono sovrapponibili in quanto VT-esenzioni è un archivio amministrativo della ASL di Viterbo relativo a soggetti che richiedono e ottengono un'esenzione per una delle patologie previste dalla normativa nazionale, mentre SIAS-esenzioni riporta l'indicazione relativa alla patologia di esenzione qualora il soggetto effettui delle prestazioni correlate alla patologia. Dato che lo stesso soggetto può avere contemporaneamente più esenzioni, non è detto che venga riportata nella prescrizione e quindi registrata nel SIAS l'esenzione per DM.

Analisi statistica

L'integrazione delle fonti è avvenuta attraverso una serie di record linkage di tipo deterministico, implementato attraverso il software SAS versione 8.2, utilizzando cinque diverse chiavi identificative: il codice fiscale e altri quattro codici realizzati dalla combinazione di cognome, nome, sesso e data di nascita. Questo al fine di valutare quanto eventuali errori presenti nelle variabili di linkage possano influenzare la stima della prevalenza. Nella sezione risultati, per semplificare la lettura delle tabelle e figure, si fa spesso riferimento solo a una di queste chiavi (cioè prime due lettere del cognome, prime due lettere del nome, sesso e data di nascita) perché ritenuta la più affidabile da controlli manuali effettuati a campione sui dati; va precisato, tuttavia, che l'utilizzo delle altre chiavi ha prodotto risultati molto simili.

La stima di prevalenza è stata poi ottenuta utilizzando la metodologia cattura e ricattura, che consente di determinare la numerosità di una popolazione, sulla base di fonti di dati già esistenti, e della diversa misura di sovrapposizione tra di esse.²³⁻²⁵

I record linkage sono stati effettuati tra due, tre e quattro archivi. Nel caso di due archivi, il numero totale di casi è stato stimato mediante lo stimatore di Chapman,²⁴ la cui validità è limitata dall'assunzione d'indipendenza ma che permette di osservare, nel caso di più fonti, l'eventuale grado di dipendenza tra di esse.²⁵ Nel caso di più di due archivi, tra i diversi metodi proposti dalla letteratura,^{24,26-28} sono stati utilizzati il modello log-lineare e l'approccio *sample-coverage*.²⁴ Il modello log-lineare può includere parametri per diversi livelli di dipendenza tra le fonti. La modalità di scelta si basa sui valori della devianza residua (più è bassa, migliore è l'adattamento) a parità di gradi di libertà.²⁴ L'approccio *sample-coverage*, attraverso la valutazio-

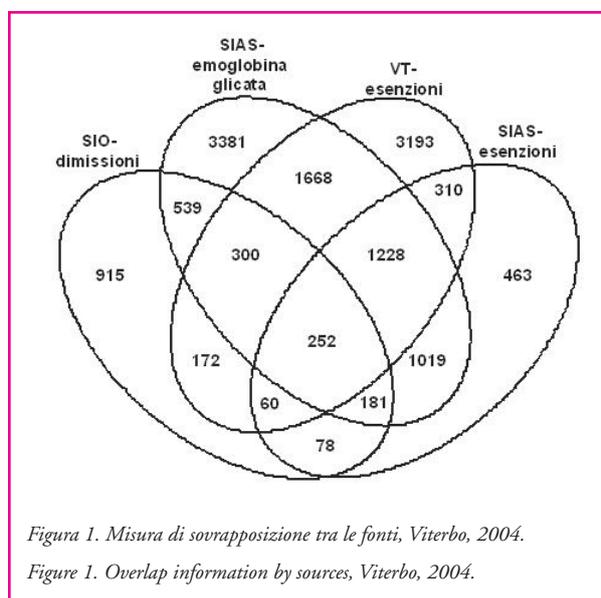


Figura 1. Misura di sovrapposizione tra le fonti, Viterbo, 2004.
Figure 1. Overlap information by sources, Viterbo, 2004.

ne della copertura campionaria data dalla media della frazione di sovrapposizione tra le liste, stima la numerosità della popolazione anche in presenza di dipendenza diretta (che viene quantificata dai coefficienti di covarianza CCV) e indiretta (dovuta all'eterogeneità tra le unità). Sono disponibili tre diversi stimatori: il primo nel caso in cui le fonti sono indipendenti, il secondo nel caso in cui l'informazione sovrapposta sia >55% e il terzo nel caso in cui la frazione sovrapposta sia <55%.²³⁻²⁵ Gli intervalli di confidenza al 95% (IC 95%) sono stati stimati attraverso una trasformazione esponenziale dei parametri e del loro errore standard nei modelli log-lineari e attraverso tecniche di *bootstrap* nell'approccio *sample-coverage*. I metodi sopra descritti sono stati applicati utilizzando il software gratuito CARE (<http://www.stat.nthu.edu.tw/~chao/>) sviluppato appositamente per le applicazioni cattura e ricattura.

In considerazione di una potenziale inapproprietezza dell'utilizzo dell'archivio SIAS-emoglobina glicata per l'identificazione di persone con DM, le analisi sono state effettuate prima includendo e poi escludendo i soggetti identificati solo attraverso questa lista.²⁹⁻³¹

Per il calcolo della prevalenza è stata considerata la numerosità della popolazione residente al 01.01.2004 nella provincia di Viterbo (295.702) disponibile all'indirizzo web: www.demo.istat.it

Tabella 1. Dimissioni, prescrizioni, esenzioni, soggetti, sesso, età nei quattro archivi utilizzati; Viterbo, Lazio 2004.

Table 1. Discharges, prescriptions, exemptions, subjects, sex, age in the four sources. Viterbo, Lazio 2004.

Archivio	n. record	Soggetti	Donne	Età mediana (Q1-Q3)
SIO-dimissioni (ICD9-CM 250)	3.486	2.488	1277 (51,3%)	72 (63-79)
SIAS-esenzioni	13.378	3.582	1829 (51,1%)	64 (56-72)
VT-esenzioni	8.529	8.529	4703 (55,1%)	69 (61-77)
SIAS-emoglobina glicata	13.529	7.169	3965 (55,3%)	66 (55-74)

Presenza nell'archivio				Chiave di linkage				
SIO dimissioni	SIAS esenzioni	VT esenzioni	SIAS-emoglobina glicata	CODF2 n.	CH1 n.	CH2 n.	CH3 n.	CH4 n.
si	no	no	no	946	930	928	915	912
no	si	no	no	500	476	471	463	457
no	no	si	no	3.470	3.428	3.422	3.381	3.375
no	no	no	si	3.304	3.245	3.235	3.193	3.180
si	si	no	no	76	77	76	78	77
si	no	si	no	542	537	538	539	540
si	no	no	si	164	172	172	172	172
no	si	si	no	1.015	1.021	1.024	1.019	1.021
no	si	no	si	320	307	306	310	309
no	no	si	si	1.612	1.638	1.643	1.668	1.672
si	si	no	si	59	60	61	60	61
si	no	si	si	278	290	290	300	301
si	si	si	no	180	182	182	181	181
no	si	si	si	1.189	1.220	1.225	1.228	1.234
si	si	si	si	243	246	248	252	255
totale soggetti in almeno un archivio				13.898	13.829	13.821	13.759	13.747

CODF2 codice fiscale;	CODF2 tax code;
CH1 combinazione di cognome, nome, sesso e data di nascita;	CH1 last name, first name, sex and date of birth;
CH2 combinazione di cognome, nome e data di nascita;	CH2 last name, first name and date of birth;
CH3 prime due lettere del cognome, prime due lettere del nome, sesso e data di nascita;	CH3 first and second letter of last name, first and second letter of first name, sex and date of birth;
CH4 prime due lettere del cognome, prime due lettere del nome e data di nascita.	CH4 first and second letter of last name, first and second letter of first name and date of birth.

Tabella 2. Risultati del record-linkage per quattro archivi attraverso l'utilizzo di 5 diverse chiavi di linkage, Viterbo, Lazio 2004.

Table 2. Results of record-linkage for four different registries using 5 different keys of linkage, Viterbo, Lazio, Italy 2004.

Risultati

La numerosità e le principali caratteristiche dei soggetti identificati dalle quattro liste sono riportati nella tabella 1. Le donne hanno una numerosità lievemente superiore a quella degli uomini e le distribuzioni per sesso differivano in maniera statisticamente significativa ($p < 0,08$ dopo correzione con il metodo di Bonferroni per confronti multipli) tra SIAS-esenzioni e VT-esenzioni e tra SIAS-esenzioni e SIAS-emoglobina glicata. Le distribuzioni per età erano significativamente diverse tra SIO-dimissioni e VT-esenzioni, tra SIO-dimissioni e SIAS-emoglobina glicata, tra SIAS-esenzioni e VT-esenzioni e tra SIAS-esenzioni e SIAS-emoglobina glicata; in particolare, questo archivio è caratterizzato da una percentuale molto elevata (10,8%) di soggetti di età inferiore ai 40 anni (dato non mostrato in tabella).

Le tabelle 2 e 3 mostrano i risultati del linkage tra 4 o 3 archivi avendo utilizzato 5 diverse chiavi. Il numero stimato di soggetti presenti in almeno un archivio varia di poco per le diverse chiavi di linkage considerate, con differenze massime pari all' 1,1% e allo 0,5%, rispettivamente nel linkage tra 4 e 3 archivi.

La figura 1 mostra il grado di sovrapposizione tra i 4 archivi nel caso dell'utilizzo della chiave di linkage CH3. Si nota che quasi il 50% dei soggetti è presente solo nell'archivio VT-esenzioni (25% del totale) o nella lista SIAS-emoglobina glicata

(23% del totale). Inoltre, il numero di casi presenti contemporaneamente in tutti e quattro gli archivi è pari al 2%. Va sottolineato che per i due archivi che riportano esenzioni per DM, solo il 28,3% dei soggetti è presente in entrambi.

Utilizzando lo stimatore di Chapman per tutte le possibili coppie di liste, le stime, nel caso dell'utilizzo della chiave di linkage CH3, vanno da 11.480 a 22.860 casi (dati non mostrati in tabella). La notevole variabilità tra esse suggerisce una dipendenza tra le fonti stesse e in particolare tra la lista SIAS-esenzioni e la lista VT-esenzioni (assumevano valori notevolmente più bassi rispetto a quelli per le altre coppie di liste). Utilizzando le altre quattro chiavi di linkage sono state ottenute stime molto simili.

In tabella 4 sono riportati i risultati di alcuni modelli log-lineari (selezionati tra tutti i modelli possibili scegliendo, a parità di gradi di libertà, quelli con la devianza più bassa) e dell'approccio *sample-coverage*.

Per quanto riguarda i modelli log-lineari, si ottengono stime decisamente poco stabili (da 18.623 a 25.920 e stime di prevalenza corrispondenti da 6,3% a 8,8%); sebbene il modello *f* mostri un adattamento significativamente migliore rispetto agli altri, nessuno dei modelli sembra mostrare un buon adattamento ai dati. Secondo l'approccio *sample-coverage*, la copertura campionaria stimata è del 67% per una stima di 23.042 casi (prevalenza 7,8%). Utilizzando le altre

Presenza nell'archivio			Chiave di linkage				
SIO dimissioni	SIAS esenzioni	VT esenzioni	CODF2 n.	CH1 n.	CH2 n.	CH3 n.	CH4 n.
si	no	no	1.110	1.102	1.100	1.087	1.084
no	si	no	820	783	776	773	765
no	no	si	5.082	5.061	5.059	5.042	5.039
si	si	no	135	136	136	137	137
si	no	si	820	825	826	836	837
no	si	si	2.204	2.238	2.243	2.243	2.248
si	si	si	423	426	428	431	434
totale soggetti in almeno un archivio			10.594	10.571	10.568	10.549	10.544
CODF2 codice fiscale;			CODF2 tax code;				
CH1 combinazione di cognome, nome, sesso e data di nascita;			CH1 last name, first name, sex and date of birth;				
CH2 combinazione di cognome, nome e data di nascita;			CH2 last name, first name and date of birth;				
CH3 prime due lettere del cognome, prime due lettere del nome, sesso e data di nascita;			CH3 first and second letter of last name, first and second letter of first name, sex and date of birth;				
CH4 prime due lettere del cognome, prime due lettere del nome e data di nascita.			CH4 first and second letter of last name, first and second letter of first name and date of birth.				

Tabella 3. Risultati del record-linkage per tre liste attraverso l'utilizzo di diverse chiavi di linkage, Viterbo, Lazio 2004.

Table 3. Results of record-linkage for three different registries with different identification keys of linkage, Viterbo, Lazio 2004.

chiavi di linkage i risultati sono molto simili senza particolari variazioni nella stima della prevalenza (da 7,8% con la chiave di linkage meno specifica, CH4, a 8,2% con quella più specifica, CODF2).

In tabella 5 vengono mostrate le stime di cattura-ricattura nel caso di record linkage di tre archivi, avendo escluso l'archivio SIAS-emoglobina glicata. Tra i modelli log-lineari selezionati, sulla base della più bassa devianza residua a parità di gradi di libertà, i modelli *b* e *c* avevano un adattamento soddisfacente ai dati con stime del numero complessivo di casi tra 17.105 e 18.009 (prevalenza tra 5,8% e 6,1%). Nell'approccio *sample-coverage* risultava una copertura campionaria del 59% con una stima di 17.345 (prevalenza del 5,9%). Utilizzando le altre chiavi di linkage i risultati sono molto simili nella stima della prevalenza (da 5,8%, CH4, a 6,1%, CODF2). Sia nel caso di utilizzo di quattro sia di tre archivi, i coefficienti di covarianza (CCV), che quantificano la dipendenza tra due o più liste con l'approccio *sample-coverage*, evidenziavano una dipendenza positiva tra le liste SIAS-esenzioni e VT-esenzioni (rispettivamente $CCV_{SIAS-esenzioni,VT-esenzioni}=0,65$ con quattro archivi e $CCV_{SIAS-esenzioni,VT-esenzioni}=0,51$ con tre archivi).

Discussione e conclusioni

Le stime di prevalenza di DM noto nell'area in studio, ottenute con diversi modelli cattura e ricattura utilizzando quattro archivi sanitari con informazioni relative al DM, sono comprese tra il 6,3% e l'8,8% ed escludendo dall'analisi l'archivio SIAS-emoglobina glicata tra il 5,7% e il 6,1%. In particolare, si è evidenziato che il solo record linkage identifica circa il 60% dei casi stimati dal metodo di cattura e ricattura quando sono stati utilizzati tutti e quattro gli archivi e circa il 61% quando è stato escluso l'archivio SIAS-emoglobina glicata.

Già altri studi avevano utilizzato la metodologia cattura e ricattura per la stima di prevalenza di DM¹⁵⁻²⁰ e alcuni avevano già evidenziato che le sole procedure di record linkage sottostimano la prevalenza della patologia, indipendentemente dal numero di fonti utilizzate.³²

Le nostre stime di prevalenza di DM noto sono più elevate di quelle riportate da altri studi italiani. Nell'indagine multi-scopo ISTAT del 1999-2000,¹⁴ si riportava una prevalenza nel Lazio del 3,4%. Per quanto riguarda studi che hanno stimato la prevalenza mediante la combinazione di 3 o più archivi, le stime nel periodo 1988-1994 variavano dal 2,6% al 3,3%.¹⁵⁻¹⁸ Due studi più recenti hanno stimato una prevalenza del 5,3% nel 2003, e del 4,1% nel 2004, quest'ultima ottenuta esclusivamente attraverso record linkage.²⁰⁻²¹ Le stime dell'OMS e dello studio di Wild et al. fornivano entrambe, per l'Italia, una prevalenza di DM noto e non noto del 7,4% nel 2000,^{4,33} simili a quelle ottenute in uno studio che aveva previsto la misurazione della glicemia a digiuno o la valutazione di un trattamento farmacologico per il DM.²² Sebbene l'accuratezza delle procedure d'abbinamento dovrebbe essere valutata attraverso un controllo esterno e non semplicemente confrontando abbinamenti diversi delle stesse variabili utilizzate, è comunque confortante che le analisi effettuate abbiano prodotto risultati simili per cinque diverse chiavi di linkage.

I due metodi statistici di cattura e ricattura utilizzati hanno prodotto risultati coerenti. Tuttavia, quando abbiamo applicato i modelli log-lineari a tutti e quattro gli archivi, la bontà di adattamento ai dati non era soddisfacente. La bassa sovrapposizione osservata fra gli archivi non ha limitato invece l'applicabilità dell'approccio *sample-coverage*. Altri vantaggi di questo tipo di approccio erano nel calcolo della copertura campionaria, nella quantificazione della dipendenza tra le fonti (attraverso i

Quattro occasioni di cattura	Devianza	Gradi di libertà	Stima casi di diabete (IC 95%)	Prevalenza (IC 95%)
record linkage			13.759	4,7
modelli_log lineari				
• a	691,87*	9	18.623 (18.352 - 18.909)	6,3 (6,2 - 6,4)
• b	346,31*	8	24.426 (23.441 - 25.512)	8,3 (7,9 - 8,6)
• c	328,37*	8	19.859 (19.500 - 20.240)	6,7 (6,6 - 6,8)
• d	135,01*	7	25.371 (24.271 - 26.587)	8,6 (8,2 - 9,0)
• e	101,30*	6	25.272 (24.092 - 26.587)	8,5 (8,1 - 9,0)
• f	63,04*	5	25.920 (24.660 - 27.325)	8,8 (8,3 - 9,2)
modello sample coverage			23.042 (22.236 - 23924)	7,8 (7,5 - 8,1)
<p>a. Con tutti gli effetti semplici e l'effetto interattivo tra le liste SIAS-esenzioni e SIAS-emoglobina glicata;</p> <p>b. con tutti gli effetti semplici, l'effetto interattivo tra le liste SIAS-esenzioni e SIAS-emoglobina glicata e il termine di eterogeneità di primo ordine che considera tutte le interazioni doppie identiche e positive;</p> <p>c. con tutti gli effetti semplici e gli effetti interattivi tra le liste SIAS-esenzioni e VT-esenzioni, SIAS-esenzioni e SIAS-emoglobina glicata;</p> <p>d. con tutti gli effetti semplici, gli effetti interattivi tra le liste SIO-dimissioni e VT-esenzioni e SIAS-esenzioni e SIAS-emoglobina glicata e il termine di eterogeneità di primo ordine che considera tutte le interazioni doppie identiche e positive;</p> <p>e. con tutti gli effetti semplici e le interazioni doppie tra le liste SIO-dimissioni e SIAS-emoglobina glicata, SIAS-esenzioni e VT-esenzioni, SIAS-esenzioni e SIAS-emoglobina glicata, VT-esenzioni e SIAS-emoglobina glicata;</p> <p>f. con tutti gli effetti semplici e le interazioni doppie tra le liste SIO-dimissioni e SIAS-esenzioni, SIO-dimissioni e SIAS-emoglobina glicata, SIAS-esenzioni e VT-esenzioni, VT-esenzioni e SIAS-emoglobina glicata.</p> <p>*La devianza mostra che il modello non ha un buon adattamento ai dati.</p>				

Tabella 4. Risultati dei modelli cattura e ricattura per quattro liste.

Table 4. Results of capture-recapture models for four lists.

CCV) e nel fatto che l'aumento del numero delle liste utilizzate non ha comportato alcuna difficoltà di interpretazione. È risultata evidente una differenza non trascurabile tra le stime ottenute con tutti e quattro gli archivi utilizzati e quelle ottenute escludendo la lista SIAS-emoglobina glicata; per esempio, mediante il metodo *sample-coverage*, la stima di prevalenza del DM era pari al 7,8%, con quattro fonti e 5,9% quando è stato escluso questo archivio. L'introduzione nell'analisi della lista delle prescrizioni SIAS-emoglobina glicata, se da una parte ha aggiunto informazione perché ha permesso di individuare soggetti con DM non rintracciati dalle altre fonti, dall'altra potrebbe aver introdotto una distorsione se questo test fosse stato utilizzato inappropriatamente come strumento di conferma diagnostico,²⁹⁻³⁰ determinando una sovrastima dei casi di DM. Tuttavia, la misurazione dell'emoglobina glicata non rientra tra le raccomandazioni per gli standard internazionali per la diagnosi del DM.³¹ Da una specifica analisi si evidenzia un'associazione con età e sesso per i soggetti presenti nell'archivio SIAS-emoglobina glicata non identificati negli altri archivi: tra i soggetti <45 anni i non linkati sono il 76%, rispetto al 60% tra quelli ≥45 anni. Inoltre, tra i soggetti <45 anni, le donne non identificate dalle altre tre liste sono oltre l'84% rispetto al 61% dei maschi. Quest'ultimo dato sembrerebbe suggerire un utilizzo diffuso di questa prescrizione in sottogruppi di popolazione, quali per esempio le donne in gravidanza. Sfortunatamente l'archivio delle prestazioni specialistiche, con riferimento all'anno 2004, non permetteva di identificare, e quindi escludere dall'analisi, quei casi la cui prescrizione verosimilmente non era dovuta al monitoraggio di soggetti con diagnosi di DM.

La dipendenza tra SIAS-esenzioni e VT-esenzioni può essere spiegata dal fatto che si tratta di due sistemi che operano nello stesso ambito; queste ultime due fonti avevano invece una bassa dipendenza con SIO-dimissioni. Si può ipotizzare in que-

sto caso che i pazienti ospedalizzati abbiano presumibilmente un quadro clinico più complesso, esenzioni per altre patologie o un'età più avanzata per cui l'esenzione non viene più richiesta. La dipendenza, comunque bassa, tra SIAS-esenzioni e SIAS-emoglobina glicata è attribuibile invece al fatto che si tratta di fonti che provengono dallo stesso sistema informativo.

Potenziati sviluppi migliorativi di questo approccio per la stima della prevalenza di DM potrebbero considerare l'utilizzo di altri archivi sanitari (ad esempio di prescrizioni farmaceutiche), di procedure di record linkage di tipo probabilistico³⁴ e lo sviluppo di metodologie che prevedano l'inclusione dell'errore di *matching* all'interno dei modelli cattura e ricattura.³⁵ Un limite dello studio è la sua riproducibilità in un ambito nazionale in quanto è improbabile poter disporre di archivi simili analoghi per tutto il territorio. Tuttavia, la metodologia è potenzialmente applicabile per tutte quelle aree dove più archivi sono disponibili.

In conclusione, il nostro studio evidenzia che gli archivi sanitari che raccolgono routinariamente informazioni amministrative possano essere utilizzati per ottenere una stima di prevalenza del DM noto, attraverso la metodologia cattura e ricattura con una notevole riduzione di tempi e costi dell'analisi rispetto alle indagini tradizionali.

Conflitti di interesse: nessuno

Ringraziamenti: si ringrazia per la collaborazione Luca Tardella (Università degli studi di Roma «La Sapienza») e le seguenti persone dell'Agenzia di sanità pubblica della Regione Lazio: Carmelina Guerrera per la revisione della terminologia relativa al diabete mellito, Amina Pasquarella (responsabile per il Lazio del progetto «Analisi dei sistemi e metodi per la valutazione dell'appropriatezza delle prestazioni diagnostiche utilizzate nell'ambito del SSN»), Cinzia Torri (responsabile del Sistema informativo dell'assistenza specialistica ambulatoriale), Paolo Papini (responsabile del Sistema informativo ospedaliero), Riccardo Salvatori per aver fornito i dati e Giulia Viola per i suggerimenti dati.

Tre occasioni di cattura	Devianza	Gradi di libertà	Stima casi di diabete (IC 95%)	Prevalenza (IC 95%)
record linkage			10.549	3,6
modelli_log lineari				
• a	5,21	2	16.788 (16.263-17.361)	5,7 (5,5-5,9)
• b	0,58	1	17.105 (16.486-17.778)	5,8 (5,6-6,0)
• c	0,44	1	18.009 (16.764-19.503)	6,1 (5,7-6,6)
sample coverage			17.345 (16.481-18.344)	5,9 (5,6-6,2)
a. con tutti gli effetti semplici e l'effetto interattivo tra le liste SIAS-esenzioni e VT-esenzioni			c. quasi simmetrico parziale, include tutti gli effetti semplici e le interazioni, che vengono poste uguali, tra le liste SIO-dimissioni e SIAS-esenzioni, SIO-dimissioni e VT-esenzioni	
b. con tutti gli effetti semplici e le interazioni tra le liste SIO-dimissioni e SIAS-esenzioni, SIAS-esenzioni e VT-esenzioni				

Tabella 5. Risultati dei modelli cattura e ricattura utilizzando tre archivi (escluso archivio prescrizioni emoglobina glicata).

Table 5. Results of capture-recapture models when using three archives (glycated haemoglobin archive excluded).

Bibliografia

- Passa P. Diabetes trends in Europe. *Diabetes Metab Res Rev* 2002; 18(Suppl 3): S3-8.
- King H, Rewers M. WHO Ad Hoc Diabetes Reporting Group: global estimates for prevalence of diabetes and impaired glucose tolerance in adults. *Diabetes Care* 1993; 16(1): 157-77.
- King H, Aubert RE, Herman WH. Global Burden of Diabetes, 1995-2025-Prevalence, numerical estimates and projections. *Diabetes Care* 1998; 21(9): 1414-31.
- Wild S, Roglic G, Green A, Sicree R, King H. Global Prevalence of Diabetes. *Diabetes Care* 2004; 27: 1047-53.
- Roglic G, Unwin N, Bennet PH et al. The Burden of Mortality Attributable to Diabetes. *Diabetes Care* 2005; 28: 2130-35.
- Morrish NJ, Wang SL, Stevens LK, Fuller JH, Keen H. WHO Multinational Study Group. Mortality and causes of death in the WHO multinational study of vascular disease in diabetes. *Diabetologia* 2001; 44: S14-S21.
- Scherthaner G. Cardiovascular mortality and morbidity in type 2 diabetes mellitus. *Diabetes Res Clin Pract* 1996; 31(Suppl): S3-S13.
- Massi-Benedetti M. The Cost of Diabetes Type II in Europe The CODE-2 Study. *Diabetologia* 2002; 45: S1-S4.
- Lucioni C, Garancini MP, Massi Benedetti M, Mazzi M, Serra G: Il costo sociale del diabete di tipo 2 in Italia: lo studio Code-2. *Pharmacoeconomics, Italian Research Articles* 2000; 1: 1-21.
- World Health Organization. Definition, diagnosis and classification of diabetes mellitus and its complications. Report of a WHO consultation. Part 1: Diagnosis and classification of diabetes mellitus. WHO/NCD/NCS/99.2 Geneva: WHO Department of Noncommunicable Disease Surveillance, 1999. (http://whqlibdoc.who.int/hq/1999/WHO_NCD_NCS_99.2.pdf)
- World Health Organization. International Diabetes Federation. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia: report of a WHO/IDF consultation. Geneva, 2006 (http://www.who.int/diabetes/publications/Definition%20and%20diagnosis%20of%20diabetes_new.pdf).
- IDF Clinical Guidelines Task Force. Global guideline for Type 2 diabetes. Brussels: International Diabetes Federation, 2005:26. <http://www.idf.org/home/index.cfm?unode=B7462CCB-3A4C-472C-80E4-710074D74AD3>
- Gnavi G, Karaghiosoff L, Dalmaso M, Bruno G. Validazione dell'archivio di esenzione per diabete della regione Piemonte: vantaggi e limiti per uno studio epidemiologico. *Epidemiol Prev* 2006; 30: 59-64.
- ISTAT. Indagine Statistica Multiscopo sulle Famiglie "Condizioni di salute e ricorso ai servizi sanitari" Anni 1999-2000.
- Bruno G, Biggeri A, Merletti F, La Porte R, McCarthy D and Pagano G. Applications of capture-recapture to "count" diabetes? *Diabetes Care* 1994; 17: 548-56.
- Di Cianni G, Benzi L, Giannarelli R et al. A prevalence study of known diabetes mellitus in Tuscany assessed from pharmaceutical prescriptions and other independent sources. *Acta Diabetol* 1994; 31: 87-90.
- Garancini MP, Sergi A, Lazzari P, Gallus G. Epidemiology of known diabetes in Lombardy, north Italy. Clinical characteristics and methodological aspects. *Acta diabetol* 1995; 32: 268-72.
- 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.
- Ciardullo AV, Menna A, Brunetti M et al. Esenzioni per diabete e distribuzione dei ricoveri nei diabetici tipo 1 e tipo 2 in Emilia-Romagna. BEN 2002, Notiziario ISS, Vol.15 - n.12.
- Bruno G, Carta Q, Runzo C, Prina Cerai S, Pagano G, Incidenza e prevalenza del diabete di tipo 2. *Il Diabete* 2004; 16: 295-99.
- Brocco S, Visentin C, Fedeli U, Schievano E, et al. Monitoring the occurrence of diabetes mellitus and its major complications: the combined use of different administrative databases. *Cardiovasc Diabetol*, 2007; 6: 5 (in press).
- Pilotto L, Gaggioli A, Lo Noce C et al. Il diabete in Italia: un problema di sanità pubblica. *Ital Heart J Suppl* 2004; 5: 480-86.
- International Working Group for Disease Monitoring and Forecasting (IWGDMF). Capture-recapture and multiple-record system estimation I: History and Theoretical Development. *Am J Epidemiol* 1995; 142: 1047-58.
- Chao A, Tsay PK, Lin SH, Shau WI, Chao DY. The application of capture-recapture models to epidemiological data. *Stat Med* 2001; 20: 3123-57.
- Chao A, Tsay PK, Shau WY, Chao DY. Population size estimation for capture-recapture models with applications to epidemiological data. *Proceedings of Biometrics Section, American Statistical Association* 1996; 108-17.
- Wittes JT, Colton T and Sidel VW. Capture-recapture models for assessing the completeness of case ascertainment using multiple information sources. *J Chronic diseases* 1974; 27: 25-36.
- Bartolucci F, Forcina A. Analysis of capture-recapture data with a Rasch-type model allowing for conditional dependence and multidimensionality. *Biometrics* 2001; 57: 714-19.
- Coull BA, Agresti A. The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics* 1999; 55: 294-301.
- Curt LR, Randie RL, Hsiao-Mei W et al. Use of GHb (HbA1c) in Screening for Undiagnosed Diabetes in the U.S. Population. *Diabetes Care* 2000; 23: 187-91.
- http://www.progettodiabete.org/indice_ie1000.html?expert/e1_84.html
- Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* 1997; 20: 1183-97.
- 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 Comm Health* 2002; 56: 18-23.
- http://www.who.int/diabetes/facts/world_figures/en/index4.html
- Fortini M, Liseo B, Nuccitelli A e Scanu M. On Bayesian Record Linkage. *Research in Official Statistics*, 2001; 185-98.
- Fortini M. Modello cattura-ricattura in presenza di errori di abbinamento: un approccio bayesiano. In: Liseo B, Montanari GE, Torelli N. *Metodi statistici per l'integrazione di dati da fonti diverse*. Milano. Edizioni Franco Angeli, 2006; 211-27.