

# Stima della prevalenza di diabete basata su dati sanitari correnti mediante un algoritmo comune in differenti aree italiane

Diabetes prevalence estimated using a standard algorithm based on electronic health data in various areas of Italy

Roberto Gnani,<sup>1</sup> Ludmila Karaghiosoff,<sup>1</sup> Daniela Balzi,<sup>2</sup> Alessandro Barchielli,<sup>2</sup> Cristina Canova,<sup>3</sup> Moreno Demaria,<sup>4</sup> Michele Pellizzari,<sup>5</sup> Stefano Rigon,<sup>5</sup> Roberta Tessari,<sup>3,6</sup> Lorenzo Simonato<sup>3</sup>

<sup>1</sup>Servizio di epidemiologia ASL TO3, Regione Piemonte

<sup>2</sup>Unità di epidemiologia, Azienda sanitaria 10, Firenze

<sup>3</sup>Dipartimento di medicina ambientale e sanità pubblica, Università di Padova

<sup>4</sup>Epidemiologia ambientale, ARPA Piemonte

<sup>5</sup>Servizio epidemiologico ULSS 4 Alto Vicentino, Thiene (VI)

<sup>6</sup>Unità di epidemiologia, Dipartimento di prevenzione, Azienda ULSS 12 Veneziana

Corrispondenza: Roberto Gnani, Servizio regionale di Epidemiologia ASL 5, Regione Piemonte, via Sabaudia 164, 10095 Grugliasco (To), Italia; tel +39 011.40188208; fax +39 011.40188201; e-mail: roberto.gnani@epi.piemonte.it

## Riassunto

**Obiettivo:** scopo del lavoro è stimare la prevalenza di diabete attraverso l'uso integrato di fonti informative correnti.

**Disegno:** è stato definito un algoritmo per l'identificazione dei casi prevalenti attraverso record linkage tra gli archivi dei ricoveri ospedalieri (SDO), delle prescrizioni farmaceutiche (PF) e delle esenzioni ticket per patologia (ET). Dalle SDO sono stati estratti i dimessi con diagnosi ICD9-CM 250\* (principale o secondaria) nell'anno di stima o nei 4 anni precedenti; dall'archivio PF sono stati estratti i soggetti con almeno due prescrizioni di farmaci antidiabetici (ATC = A10\*) nell'anno di stima; dall'ET sono stati selezionati i soggetti con esenzione per diabete (013.250) rilasciata nell'anno di stima o nei 3 anni precedenti.

Le stime di prevalenza grezza, per età e per sesso, e la prevalenza standardizzata per età sono state confrontate in quattro aree geografiche italiane. E' stato anche valutato il contributo delle diverse fonti informative al calcolo della prevalenza.

**Risultati:** nel 2003, la prevalenza di diabete varia tra i centri da 3,93% a 5,55% negli uomini e da 3,55% a 4,52% nelle donne. Dopo standardizzazione per età le differenze si riducono tra gli uomini e scompaiono nelle donne. La prevalenza aumenta con l'età ed è costantemente più elevata negli uomini rispetto alle donne.

**Conclusioni:** l'algoritmo utilizzato ha prodotto stime di prevalenza di diabete omogenee tra i quattro centri partecipanti e simili a quelle prodotte da altre indagini condotte in Italia. La principale fonte di casi è costituita dalle prescrizioni farmaceutiche, ma il contributo delle altre due fonti è comunque di grande rilievo e permette di migliorare la completezza della casistica, compensando le differenze geografiche nelle modalità di ricovero o di ricorso all'esenzione ticket. E' quindi possibile stimare la prevalenza di malattia e individuare popolazioni di soggetti con diabete da seguire nel tempo.

(*Epidemiol Prev* 2008; 32(3) suppl 1: 15-21)

**Parole chiave:** diabete, archivi elettronici, record linkage, prevalenza.

## Abstract

**Aims:** the goal of this study was to estimate the prevalence of diabetes through record linkage of various data sources in four Italian areas.

**Setting:** Aulss 12 Veneziana, Aulss 4 Alto vicentino, Torino, ASL10 of Firenze.

**Participants:** all 2002 to 2004 residents in the four areas (n = 2,123,913 on 30th June 2003).

**Main outcome:** crude prevalence by age and gender and standardized prevalence by gender.

**Methods:** we used three different data sources. The first was the set of files of all persons discharged from hospitals with a primary or secondary diagnosis of diabetes (ICD-9-CM code 250\*)

in the year of interest or in the four previous years. The second data source was the set of files of all prescriptions of antidiabetic drugs (ATC code: A10A\* and A10B\*) prescribed in the year of interest; we considered as persons with diabetes only those who had at least two prescriptions of antidiabetic drugs at two different times. The third source was the set of files of all subjects who obtained exemption from payment of drugs or laboratory testing due to a diagnosis of diabetes mellitus in the year of interest or in the 3 previous years. All data sources were matched by a deterministic linkage procedure. We defined as «prevalent case» those persons who were present in at least one of the three data sources. We compared the estimated prevalence in the four different areas.

**Results:** in 2003, the prevalence of diabetes in the four areas ranged from 3.93% to 5.55% among men, and from 3.55% to 4.52% among women. After adjustment for age, differences among men were reduced and were no longer present among women. Prevalence is higher among the elderly and among men.

**Conclusions:** using routinely collected data we were able to

identify large cohorts of persons with known diabetes and to estimate the prevalence of the disease, which was shown to be highly homogeneous among participating centres, and similar to that reported in other studies conducted in Italy with more costly and time consuming methods.

(Epidemiol Prev 2008; 32(3) suppl 1: 15-21)

**Keywords:** diabetes, electronic archives, prevalence, record linkage.

## Introduzione

Il diabete mellito è una malattia cronica in progressivo aumento a livello internazionale.<sup>1</sup> Le cause principali di questo incremento sono attribuibili all'allungamento della vita media e alle modificate abitudini di vita della popolazione generale.<sup>2</sup> Nel 2003, il diabete era l'ottava causa di morte nei paesi sviluppati.<sup>3</sup>

In Italia la sopravvivenza dei diabetici è di circa il 30-40% inferiore rispetto a quella della popolazione generale, prevalentemente a causa dell'aumentato rischio cardiovascolare.<sup>4,5</sup> Il diabete, inoltre, è la principale causa di ricorso alla terapia dialitica<sup>6</sup> e una delle cause più rilevanti di cecità,<sup>7</sup> ed è inoltre responsabile del 50% delle amputazioni degli arti inferiori.<sup>8</sup>

L'elevato impatto della malattia sulle risorse sanitarie fa sì che gli studi epidemiologici volti a monitorare l'andamento della malattia e degli indicatori della qualità della cura rivestano un grande interesse sia per i clinici sia per gli amministratori della sanità. Anche il Piano prevenzione attiva 2004-2006 del Ministero della salute ha recepito questa esigenza, individuando come obiettivi prioritari «l'attivazione di un sistema di monitoraggio della malattia su base informatizzata» e «identificare i pazienti diabetici attraverso l'iscrizione in appositi registri», non ancora realizzati, tuttavia, nella maggioranza delle regioni italiane.<sup>9</sup>

La possibilità di disporre di sistemi di sorveglianza in grado di monitorare la malattia nelle diverse aree geografiche del Paese riveste quindi una grande importanza. Archivi di soggetti con diabete raccolti per fini amministrativi (ricoveri ospedalieri, prescrizioni farmaceutiche, esenzioni ticket, mortalità) sono ormai facilmente disponibili in molte realtà italiane e possono essere associati tra loro per la ricostruzione di popolazioni di diabetici.<sup>10,11</sup>

Scopo di questo lavoro è mettere a punto, attraverso l'uso integrato di fonti informative correnti, un algoritmo per stimare la prevalenza di malattia e per definire coorti di soggetti con diabete da seguire al fine di valutare la qualità dell'assistenza.

## Materiali e metodi

Lo studio ha interessato i residenti di quattro centri del Nord Italia: l'AULSS 12 Veneziana (di seguito denominata Venezia), l'AULSS 4 Alto vicentino (di seguito denominata Thiene), la città di Torino (di seguito denominata Torino) e l'A-

SL 10 di Firenze (di seguito denominata Firenze), vivi all'1 gennaio dell'anno di stima, per un totale di 2.123.913 soggetti (anno 2003). La prevalenza è stata stimata per gli anni 2002, 2003 e 2004. A causa dell'indisponibilità di alcune fonti, nel caso di Firenze la stima ha riguardato solo l'anno 2003; nel caso di Thiene solo il 2003 e il 2004.

Per l'identificazione dei casi sono stati utilizzati gli archivi dei ricoveri ospedalieri (SDO), delle prescrizioni farmaceutiche (PF) e delle esenzioni ticket (ET).

Dall'archivio SDO sono state selezionate tutte le dimissioni (comprehensive della mobilità extraregionale) avvenute in residenti con diagnosi (principale o secondarie) di diabete (codice IDC9-CM = 250\*) per ogni anno di stima e nei 4 anni precedenti. Per ogni quinquennio sono stati quindi individuati i singoli soggetti (escludendo quindi i ricoveri ripetuti) e la residenza è stata definita sulla base di quanto riportato sulla SDO. Solo nel caso di Torino e Thiene è stato possibile escludere i soggetti emigrati prima dell'1 gennaio dell'anno di stima, mentre per gli altri due centri sono stati inclusi come residenti anche i soggetti che risultavano residenti al momento della dimissione ed emigrati entro l'1 gennaio dell'anno di stima.

Dall'archivio delle prescrizioni farmaceutiche, per ogni anno di stima, sono state estratte tutte le prescrizioni di farmaci antidiabetici (codice ATC = A10\*) effettuate a residenti. Al fine di ridurre il numero di falsi positivi legati a errori di trascrizione del codice fiscale/tessera sanitaria si è ritenuto di considerare come soggetti con diabete solo i casi con almeno due prescrizioni di farmaci antidiabetici nel corso dell'anno di stima.

Dall'archivio delle esenzioni ticket sono stati selezionati tutti i residenti con codice esenzione per diabete (013.250) rilasciato nell'anno di stima o nei tre anni precedenti. Nel solo caso di Torino è stato utilizzato l'archivio regionale dei soggetti con diabete che, di fatto, ha le stesse caratteristiche di un archivio di esenzione.<sup>12</sup>

Queste tre fonti sono state appaiate tra loro con tecniche di record linkage descritte nel capitolo 1 di questa monografia allo scopo di escludere i soggetti presenti in più di una fonte (capitolo 1). Per escludere i soggetti deceduti entro l'1 gennaio di ogni anno gli archivi sono stati appaiati a fonti localmente disponibili in grado di identificare l'esistenza in vita dei soggetti (registri delle cause di morte, anagrafe comunale, anagrafe degli assistiti). Sono stati quindi definiti

| Fonte                                                                                                                                                                                                                                                                                                | Criteri di selezione casistica                                                                                                                                             | Criteri per la definizione di prevalenza |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------|
| schede di dimissione ospedaliera                                                                                                                                                                                                                                                                     | dimissione nell'anno di stima o nei 4 anni precedenti con diagnosi di diabete (ICD9-CM = 250*) in qualunque campo diagnosi, e non deceduto al 1 gennaio dell'anno di stima | presenza in almeno una delle fonti       |
| prescrizioni farmaceutiche                                                                                                                                                                                                                                                                           | almeno due prescrizioni nell'anno di stima per farmaci antidiabetici (ATC = A10*)                                                                                          |                                          |
| esenzioni ticket                                                                                                                                                                                                                                                                                     | esenzione per diabete (codice esenzione = 013.250) nell'anno di stima o nei tre anni precedenti e non deceduto all'1 gennaio dell'anno di stima                            |                                          |
| ICD9-CM: Classificazione internazionale delle malattie – 9ª revisione – con modifiche cliniche; <i>International classification of diseases – 9th revision – clinical modifications</i><br>ATC: Classificazione anatomica terapeutica chimica; <i>Chemical therapeutic anatomic classification</i> . |                                                                                                                                                                            |                                          |

Tabella 1. Fonti dei dati e criteri di definizione di caso prevalente di diabete.

Table 1. Data sources and criteria defining prevalent case of diabetes.

come casi di diabete prevalente tutti i soggetti che, per ogni singolo anno, comparivano almeno una volta in almeno uno degli archivi utilizzati e risultavano essere vivi e residenti all'1 gennaio dell'anno di stima. La [tabella 1](#) riassume l'algoritmo utilizzato.

Per il calcolo della prevalenza il numero di casi ottenuto per ogni singolo anno è stato rapportato alle relative popolazione Istat per sesso e fascia di età. Per il confronto tra centri le prevalenze sono state standardizzate per età con metodo diretto (standard: popolazione italiana 2001) e sono stati calcolati gli intervalli di confidenza al 95%. Vengono inoltre riportati, disaggregati per area, età, sesso e anno in studio, alcuni indicatori di qualità e di confrontabilità dei dati raccolti (numero di record selezionati dagli archivi suddivisi per anno di rilevazione, contributo delle singole fonti informative eccetera).

## Risultati

Nella [tabella 2](#) sono riportati il numero di casi e la prevalenza, grezza e standardizzata per età, distribuiti per anno,

sesso e centro. La prevalenza è più alta negli uomini rispetto alle donne e incrementa progressivamente negli anni. In entrambi i sessi i valori più elevati si riscontrano a Venezia e i più bassi a Thiene. Dopo standardizzazione per età ([tabella 2](#)) si accentuano le differenze a sfavore degli uomini, si conferma un lieve trend temporale in aumento (più evidente a Torino), mentre si attenuano le differenze geografiche che, nel caso delle donne, risultano pressoché assenti. Solo per il sesso maschile, Venezia continua a presentare il valore più elevato.

La prevalenza cresce con il crescere dell'età con modalità sostanzialmente omogenee in tutti i centri e in entrambi i sessi ([figura 1](#)). Venezia si conferma la realtà territoriale con i valori più consistenti in tutte le fasce di età (fa eccezione, nelle donne, la classe di età più anziana). Nella classe di età «oltre 74 anni» la prevalenza di diabete varia da un minimo di 13,3% a Torino a un massimo di 16,3% a Venezia nel caso degli uomini, e da un minimo di 11,1%, sempre a Torino, a un massimo di 15,2% a Thiene nelle donne.

Le modalità con cui si giunge alla definizione dei casi di dia-

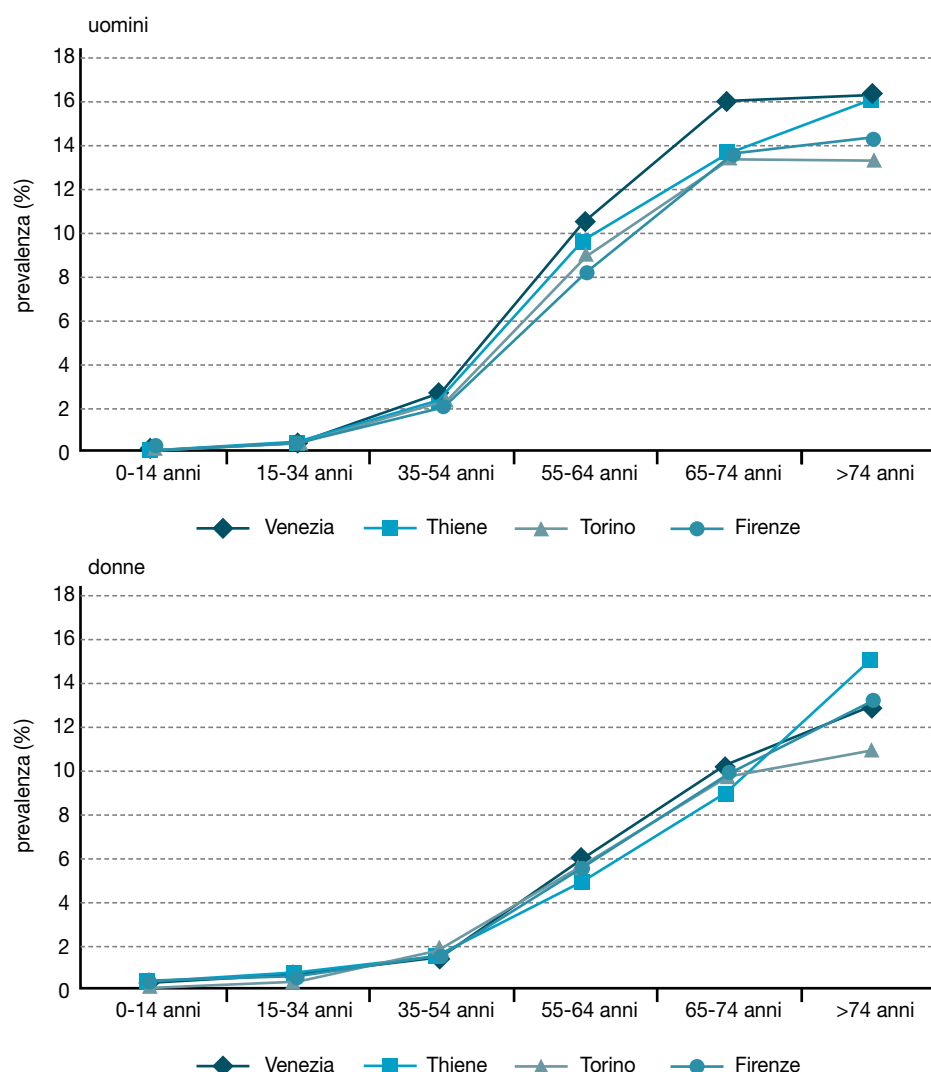
| Area    | Anno | Uomini  |      |      |             | Donne   |      |      |             | Totale  |      |      |             | Rapporto U/D (TS) |
|---------|------|---------|------|------|-------------|---------|------|------|-------------|---------|------|------|-------------|-------------------|
|         |      | n. casi | TG   | TS   | IC 95%      | n. casi | TG   | TS   | IC 95%      | n. casi | TG   | TS   | IC 95%      |                   |
| Venezia | 2002 | 7.819   | 5,45 | 5,09 | 4,99 – 5,20 | 7.043   | 4,45 | 3,31 | 3,23 – 3,39 | 14.862  | 4,93 | 4,11 | 4,05 – 4,18 | 1,54              |
|         | 2003 | 7.990   | 5,55 | 5,15 | 5,04 – 5,25 | 7.164   | 4,52 | 3,33 | 3,26 – 3,41 | 15.154  | 5,01 | 4,15 | 4,09 – 4,21 | 1,55              |
|         | 2004 | 8.108   | 5,60 | 5,16 | 5,06 – 5,27 | 7.303   | 4,58 | 3,38 | 3,30 – 3,45 | 15.411  | 5,07 | 4,18 | 4,12 – 4,24 | 1,53              |
| Thiene  | 2003 | 3.473   | 3,93 | 4,66 | 4,51 – 4,81 | 3.230   | 3,55 | 3,31 | 3,20 – 3,42 | 6.703   | 3,74 | 3,96 | 3,87 – 4,05 | 1,41              |
|         | 2004 | 3.578   | 3,99 | 4,65 | 4,51 – 4,80 | 3.375   | 3,67 | 3,40 | 3,29 – 3,51 | 6.953   | 3,83 | 4,01 | 3,92 – 4,10 | 1,37              |
| Torino  | 2002 | 18.468  | 4,51 | 4,27 | 4,21 – 4,33 | 18.561  | 4,08 | 3,23 | 3,18 – 3,28 | 37.029  | 4,28 | 3,69 | 3,65 – 3,73 | 1,32              |
|         | 2003 | 19.184  | 4,70 | 4,39 | 4,33 – 4,46 | 19.138  | 4,22 | 3,30 | 3,25 – 3,35 | 38.322  | 4,45 | 3,78 | 3,74 – 3,82 | 1,33              |
|         | 2004 | 19.985  | 4,86 | 4,55 | 4,48 – 4,61 | 19.874  | 4,35 | 3,41 | 3,36 – 3,45 | 39.859  | 4,59 | 3,91 | 3,87 – 3,94 | 1,33              |
| Firenze | 2003 | 17.319  | 4,69 | 4,35 | 4,28 – 4,41 | 17.993  | 4,42 | 3,38 | 3,33 – 3,43 | 35.312  | 4,55 | 3,83 | 3,79 – 3,86 | 1,29              |

Tabella 2. Numero totale di casi prevalenti di diabete, tassi di prevalenza grezzi (TG per 100), standardizzati (TS per 100) e intervalli di confidenza (IC 95%) per area, anno e sesso. Rapporto uomini/donne (U/D) dei tassi di prevalenza standardizzati.

Table 2. Total number of diabetic prevalent cases, crude prevalence rate (TG per 100), standardized prevalence rate (TS per 100) and confidence intervals (CI 95%) by area, year and sex. Male/female (U/D) ratio of standardized prevalence rates.

Figura 1. Prevalenza grezza di diabete per area, sesso ed età. Anno 2003.

Figure 1. Crude prevalence of diabetes by area, sex and age. Year 2003.



| Centro  | Anno | SDO   | %   | PF      | %   | ET    | %    |
|---------|------|-------|-----|---------|-----|-------|------|
| Venezia | 2002 | 3.379 | 5,3 | 87.061  | 3,9 | 692   | 6,7  |
|         | 2003 | 3.698 | 5,9 | 86.131  | 4,1 | 1.232 | 4,3  |
|         | 2004 | 3.514 | 6,0 | 89.133  | 4,1 | 1.061 | 2,7  |
| Thiene  | 2003 | 1.037 | 3,2 | 37.169  | 3,6 | 791   | 5,8  |
|         | 2004 | 1.103 | 3,5 | 38.648  | 3,5 | 718   | 6,8  |
| Torino  | 2002 | 9.607 | 5,5 | 214.966 | 3,4 | 2.915 |      |
|         | 2003 | 9.523 | 5,6 | 234.666 | 3,7 | 2.728 |      |
|         | 2004 | 9.676 | 5,6 | 250.481 | 3,7 | 2.851 |      |
| Firenze | 2003 | 8.448 | 5,4 | 242.507 | 3,9 | 1.626 | 12,2 |

SDO: schede di dimissione ospedaliera; hospital discharges  
 PF: prescrizioni farmaceutiche; drugs prescriptions  
 ET: esenzione ticket; health-tax exemption

Tabella 3. Numero di record selezionati con i codici di diabete e percentuale sul totale dei record per area e anno.

Table 3. Number of records selected by diabetes codes and proportion (%) on the total number of records by area and year.

bete differiscono tra i vari centri. La tabella 3 riporta la percentuale di codici di interesse all'interno dei singoli archivi. Per quanto riguarda l'archivio SDO, Venezia, Torino e Firenze hanno comportamenti molto simili con una frequenza di diagnosi di diabete intorno al 5,5%; Thiene, invece, ha una frequenza decisamente inferiore (circa 3,5%). Più omogenea è la prescrizione di farmaci antidiabetici, che rappresentano il 3,5- 4% del totale delle prescrizioni nei quattro centri. La percentuale di esenzioni ticket per diabete è invece più variabile: dal 4% Venezia al 12% di Firenze. La figura 2 riporta il contributo assoluto delle singole fonti informative (ovverosia, la percentuale di casi individuati da ogni archivio, indipendentemente dal fatto che siano presenti anche in altri archivi). In tutti i centri il contributo maggiore è fornito dall'archivio PF, con differenze minime: a Venezia e Thiene circa l'81% dei soggetti con diabete è rintracciabile dall'archivio PF, mentre a Torino e Firenze il contributo è del 78%. La seconda fonte in ordine di importan-

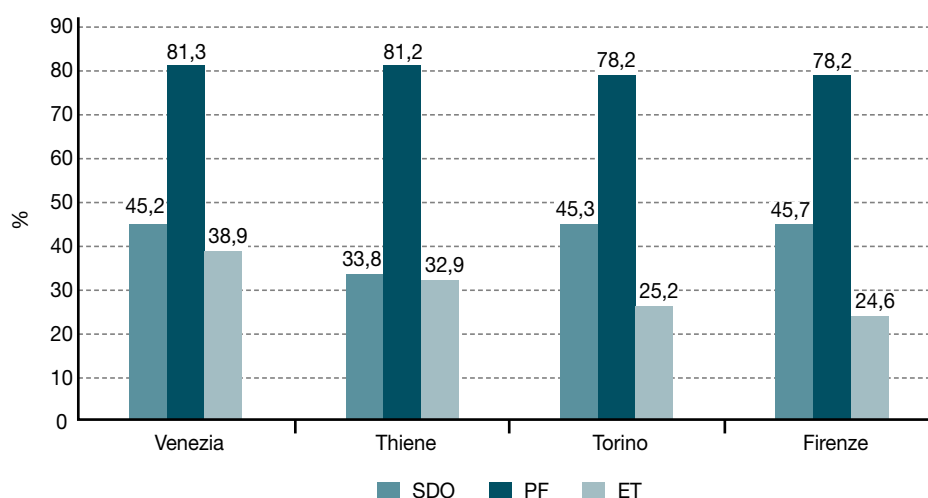


Figura 2. Contributo assoluto delle fonti dei dati per area. Anno 2003.

Figure 2. Absolute contribution of data sources by area. Year 2003.

|               |        | Area    |      |      |           |        |      |      |           |        |      |      |           |         |      |      |           |
|---------------|--------|---------|------|------|-----------|--------|------|------|-----------|--------|------|------|-----------|---------|------|------|-----------|
|               |        | Venezia |      |      |           | Thiene |      |      |           | Torino |      |      |           | Firenze |      |      |           |
|               |        | SDO     | PF   | ET   | più fonti | SDO    | PF   | ET   | più fonti | SDO    | PF   | ET   | più fonti | SDO     | PF   | ET   | più fonti |
| sesso         | uomini | 9,8     | 31,2 | 6,5  | 52,5      | 6,4    | 40,7 | 10,0 | 42,8      | 11,1   | 35,6 | 8,9  | 44,3      | 14,8    | 37,0 | 5,4  | 42,9      |
|               | donne  | 10,7    | 30,7 | 5,8  | 52,7      | 7,7    | 43,3 | 10,1 | 38,8      | 10,3   | 38,2 | 9,1  | 42,4      | 15,8    | 39,5 | 4,1  | 40,7      |
| classi di età | 0-14   | 0       | 4,3  | 0    | 95,7      | 0      | 4,2  | 0    | 95,8      | 11,5   | 5,0  | 2,2  | 81,3      | 24,1    | 19,9 | 1,4  | 54,6      |
|               | 15-34  | 7,7     | 26,8 | 11,5 | 54,0      | 1,6    | 24,7 | 38,7 | 34,9      | 5,8    | 25,7 | 33,5 | 35,0      | 16,9    | 33,3 | 7,3  | 42,5      |
|               | 35-54  | 7,5     | 32,7 | 12,3 | 47,5      | 3,1    | 36,4 | 19,5 | 41,1      | 7,3    | 32,9 | 17,1 | 42,7      | 12,0    | 36,0 | 10,2 | 41,8      |
|               | 55-64  | 6,1     | 33,0 | 8,7  | 52,2      | 3,4    | 44,6 | 12,0 | 40,0      | 7,6    | 38,7 | 10,9 | 42,8      | 10,3    | 40,7 | 7,2  | 41,7      |
|               | 65-74  | 8,0     | 32,3 | 4,7  | 54,9      | 5,8    | 48,2 | 7,8  | 38,1      | 9,2    | 39,5 | 7,3  | 44,0      | 11,7    | 42,9 | 3,9  | 41,5      |
|               | 75+    | 16,8    | 27,7 | 3,3  | 52,3      | 13,3   | 39,0 | 3,9  | 43,8      | 17,3   | 35,5 | 3,6  | 43,6      | 22,3    | 34,0 | 1,9  | 41,8      |
| totale        |        | 10,4    | 31,5 | 6,3  | 53,7      | 7,1    | 42,0 | 10,1 | 40,9      | 10,7   | 36,9 | 9,0  | 43,4      | 15,3    | 38,3 | 4,7  | 41,8      |

SDO: schede di dimissione ospedaliera; *hospital discharges*  
 PF: prescrizioni farmaceutiche; *drug prescriptions*  
 ET: esenzione ticket; *health-tax exemption*

Tabella 4. Contributo esclusivo delle fonti dei dati per area, sesso, classe di età. Anno 2003.

Table 4. Relative contribution of data sources by area, sex, age class. Year 2003.

za è rappresentata dall'archivio SDO: il 45% dei casi è individuato da questa fonte, tranne che a Thiene, dove il contributo è solo del 33,8%. Più modesto, ma non trascurabile, il contributo dell'esenzione ticket: circa il 25% a Torino e Firenze, il 33% a Thiene e il 39% a Venezia.

La [tabella 4](#) riporta il contributo esclusivo di ogni singola fonte (ovverosia il numero di casi individuati in un unico archivio) e il contributo derivante dall'incrocio di più fonti. Anche in questo caso l'archivio PF risulta la fonte informativa più rilevante nell'identificare soggetti con diabete: il 31,5% dei casi a Venezia (valore minimo) e il 42% dei casi a Thiene (valore massimo) sono identificati unicamente da questa fonte. Le SDO rappresentano l'unica fonte per il 15% dei casi di Firenze contro il 7% di Thiene, mentre il contributo esclusivo dell'archivio ET è minore, ma non trascurabile, variando da un minimo di 4,7% a Firenze a un massimo del 10% a Thiene. Questi valori sono sostanzialmente omogenei per sesso, mentre presentano alcune differenze per classe di età: il contributo esclusivo della fonte SDO cresce con l'età, mentre quello dell'ET ha un comportamento opposto: diventa quasi irrilevante nelle fasce di età più anziane per l'ovvia competizione dell'esenzione per età rispetto a quella per patologia. Il contributo esclusivo dell'archivio PF per età è omogeneo a Venezia, Torino e Firenze, mentre a Thiene è più alto nelle classi di età intermedie. Un caso particolare è rappresentato dalla fascia di età infantile, in cui i comportamenti appaiono molto differenziati tra centri; questa differenza potrebbe essere imputabile sia alla bassa numerosità di casi, sia a diverse modalità di diagnosi sulla SDO, ma potrebbe anche rappresentare la spia

rabile, variando da un minimo di 4,7% a Firenze a un massimo del 10% a Thiene. Questi valori sono sostanzialmente omogenei per sesso, mentre presentano alcune differenze per classe di età: il contributo esclusivo della fonte SDO cresce con l'età, mentre quello dell'ET ha un comportamento opposto: diventa quasi irrilevante nelle fasce di età più anziane per l'ovvia competizione dell'esenzione per età rispetto a quella per patologia. Il contributo esclusivo dell'archivio PF per età è omogeneo a Venezia, Torino e Firenze, mentre a Thiene è più alto nelle classi di età intermedie. Un caso particolare è rappresentato dalla fascia di età infantile, in cui i comportamenti appaiono molto differenziati tra centri; questa differenza potrebbe essere imputabile sia alla bassa numerosità di casi, sia a diverse modalità di diagnosi sulla SDO, ma potrebbe anche rappresentare la spia



di errori nella codifica di diabete o nelle procedure di linkage con gli altri archivi, essendo poco probabile una dimissione ospedaliera con diagnosi di diabete in età pediatrica (come è il caso di Torino e Firenze) non accompagnata da terapia insulinica.

### Discussione

Questo lavoro ha sperimentato la fattibilità di un algoritmo per l'identificazione di casi prevalenti di diabete utilizzando alcune delle fonti informative amministrative comunemente disponibili presso le aziende sanitarie della maggioranza delle regioni italiane e rappresenta l'unico lavoro italiano pubblicato che confronta in aree geografiche differenti risultati derivati dall'applicazione di un medesimo protocollo.

La prevalenza grezza stimata nei diversi centri partecipanti varia tra il 4 e il 5%, valori in linea con quelli stimati dalle poche altre esperienze italiane disponibili. La recente indagine campionaria Istat *Salute, stili di vita e ricorso ai servizi* stima nel 4,5% la prevalenza grezza di diabete (diagnosi autoriferita) in Italia, con valori di 4,2% in Piemonte e Veneto e 4,6% in Toscana; dopo standardizzazione per età (standard: Italia 2001) questi valori diventano, rispettivamente, 4,6% (Italia), 3,8% (Piemonte), 4,2% (Veneto) e 4% (Toscana), sovrapponibili a quelli rilevati in questo lavoro.<sup>13</sup> In una popolazione italiana di oltre 10 milioni di persone coperta da una sistema di rilevazione delle prescrizioni farmaceutiche omogeneo, e utilizzando come indicatore una sola prescrizione/anno, è stata stimata una prevalenza, nel 2006, del 4,5%.<sup>14</sup>

Vi sono poi alcune esperienze locali che, nuovamente, misurano valori di prevalenza simili a quelli riportati. Nel 2004, utilizzando un algoritmo solo in parte simile a quello utilizzato in questo studio,<sup>11</sup> in tre ASL del Veneto è stata stimata una prevalenza grezza di 4,3% negli uomini e 3,8% nelle donne (simili ai valori stimati a Thiene); dati che, dopo standardizzazione per età (standard: Italia 2000) sono diventati, rispettivamente, 3,9% e 3,5%.

In Piemonte, nell'area di Casale Monferrato, nel 2000 si è stimata una prevalenza grezza di 3,7% negli uomini e 3,8% nelle donne; dopo standardizzazione per età (standard: Italia 2001) si è ottenuto il 3,1% in entrambi i sessi.<sup>15</sup> A Torino, nel 2003, utilizzando un algoritmo simile a quello utilizzato in questo studio e aggiustando per la sottostima con un modello cattura-ricattura è stata stimata una prevalenza grezza di 4,9% negli uomini e 4,7% nelle donne.<sup>16</sup>

Si tratta, nel complesso, di valori simili a quelli ottenuti in questo lavoro, anche se le differenze nelle fonti e negli algoritmi utilizzati non consentono una comparabilità adeguata e l'assenza di un gold standard per misurare la validità delle stime prodotte è un limite di questo studio.

Un altro limite è rappresentato dall'utilizzo di fonti informative nate per fini amministrativi, senza alcuna valida-

zione formale dei codici utilizzati per selezionare i casi di diabete.

In Italia non sono disponibili studi di validazione di questo genere, ma una recente revisione internazionale (che non includeva studi italiani) sui metodi utilizzati per identificare casi di diabete ha confermato l'elevato valore d'uso di queste fonti che presentano buoni livelli di sensibilità e un'elevata specificità, rafforzando quindi l'indicazione a utilizzare queste fonti per la ricostruzione di coorti e, soprattutto, a utilizzare queste popolazioni ai fini del monitoraggio dell'assistenza.<sup>17</sup> Infatti, le potenzialità di questo approccio vanno oltre la misura puntuale della prevalenza (per la quale sarebbe importante garantire un alto livello di completezza), ma risiedono nella possibilità di ricostruire coorti di soggetti sicuramente affetti da diabete (da qui l'importanza di un alto livello di specificità) da sottoporre successivamente a follow up per la valutazione degli esiti dell'assistenza (mediante successivi record linkage con gli archivi di mortalità, dei ricoveri ospedalieri, delle prestazioni ambulatoriali, delle prescrizioni farmaceutiche eccetera). Per quanto riguarda la completezza, invece, attraverso tecniche di cattura-ricattura si è stimato che la quota di soggetti con diabete rilevata attraverso le fonti correnti corrisponde a circa l'80% della popolazione con diabete.<sup>15,16</sup> E' probabile che il restante 20% sia costituito da soggetti con forme di diabete meno grave e trattate con la sola terapia dietetica.

Tra le fonti informative proposte per l'algoritmo si è ritenuto di non includere la mortalità. Le motivazioni sono legate al bassissimo contributo esclusivo di questa fonte (0,15% a Firenze, 0,1% a Venezia, 0,07% a Torino) e alla possibilità che si tratti di falsi positivi, essendo assai poco probabile un decesso per diabete in assenza di ricovero o di prescrizioni farmaceutiche.

Tra i limiti di questa procedura vanno riportati:

- la difficoltà di distinguere il diabete di tipo 1 dal diabete tipo 2, essendo ormai poco applicabile una differenziazione basata solo sull'età;<sup>18</sup>

- la difficoltà di definire il livello di gravità della patologia e la presenza di comorbidità, se non ricavandolo attraverso l'identificazione di altre terapie farmacologiche in atto o la presenza di altre diagnosi ospedaliere nel corso del ricovero indice o di ricoveri precedenti;

- la difficoltà di definire la durata della malattia non essendo possibile, a causa dell'indisponibilità di archivi storici di SDO e PF (se non per pochi anni), individuare la data di insorgenza della malattia. Queste ultime condizioni, che riducono le possibilità di risk adjustment, possono costituire un'importante limitazione all'utilizzo di queste coorti per studi di esito.

Per quanto riguarda la valutazione del contributo informativo delle differenti fonti all'individuazione dei casi di diabete, va innanzitutto osservato come esistano differenze tra i diversi centri nella diagnosi di diabete all'interno della SDO

e nel ricorso all'esenzione ticket. Se ciò dipenda da una diversa attenzione alla diagnosi di diabete, da un minore ricorso al ricovero in ospedale nei soggetti con diabete o, ancora, da differenze nella validità dei codici utilizzati, non è noto. Le differenze nel contributo assoluto ed esclusivo delle singole fonti rivelano l'importanza della disponibilità di tutte le tre fonti: le prescrizioni farmaceutiche rappresentano certamente la fonte informativa più rilevante per l'individuazione dei casi di diabete, ma il contributo delle altre due è irrinunciabile.

In conclusione, pur con alcuni limiti propri delle fonti informative nate per fini amministrativi, l'algoritmo di stima proposto produce valori di prevalenza di diabete confrontabili con quelli documentati da altri studi condotti in Italia. Si tratta di un algoritmo di calcolo semplice e alla portata della maggior parte delle realtà italiane, che permette sia di stimare il peso della malattia diabetica all'interno di una popolazione, sia di disporre di popolazioni di malati da seguire nel tempo per fini di monitoraggio della qualità dell'assistenza.

**Conflitti di interesse:** nessuno.

## Bibliografia

1. Wild S, Roglic G, Green A et al. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 2004; 27:1047-53.
2. Stumvoll M, Goldstein BJ, van Haeften TW. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 2005; 365: 1333-46.
3. Lopez AD, Mathers CD, Ezzati J et al. Global and regional burden of disease and risk factors, 2001; systematic analysis of population health data. *Lancet* 2006; 367: 1747-57.
4. De Marco R, Locatelli F, Zoppini G et al. Cause specific mortality in type 2 diabetes. The Verona diabetes study. *Diabetes Care* 1999; 22: 756-761.
5. Gnani R, Petrelli A, Demaria M et al. Mortality and educational level among diabetic and non-diabetic people in the Turin Longitudinal Study: a 9-year follow-up. *Int J Epidemiol* 2004; 33: 864-71.
6. Reboldi GP, Quintaliani G, Bueti A. Nefropatia e ipertensione nel diabete tipo 2. *Il Diabete* 2004; Suppl. 1: 348-57.
7. Porta M. Retinopatia diabetica. *Il Diabete* 2004; Suppl. 1: 343-47.
8. Vaccaro O, De Feo ME. Amputazioni associate al diabete. *Il Diabete* 2004; Suppl. 1: 340-42.
9. [http://www.epicentro.iss.it/focus/piano\\_prevenzione/PianoPrevenzione04-06.pdf](http://www.epicentro.iss.it/focus/piano_prevenzione/PianoPrevenzione04-06.pdf) (ultimo accesso 23 gennaio 2006).
10. Bruno G, La Porte R, Merletti F et al. National diabetes programmes: application of capture-recapture to "count" diabetes? *Diabetes Care* 1994; 17: 548-56.
11. Brocco S, Visentin C, Fedeli U et al. Monitoring the occurrence of diabetes mellitus and its major complications: the combined use of different administrative databases. *Cardiovasc Diabetol* 2007; 6: 5 doi:10.1186/1475-2840-6-5.
12. Gnani R, Karaghiosoff L, Dalmaso M, Bruno G. Validazione dell'archivio di esenzione per diabete della Regione Piemonte: vantaggi e limiti per un uso epidemiologico. *Epidemiol Prev* 2006; 30: 59-64.
13. [http://www.istat.it/salastampa/comunicati/non\\_calendario/20070302\\_00/testointegrale.pdf](http://www.istat.it/salastampa/comunicati/non_calendario/20070302_00/testointegrale.pdf).
14. De Rosa M, Marchesini G (a cura di). Osservatorio ARNO diabete, analisi di dieci anni di prescrizioni. Collana Rapporti ARNO. Rapporto 2007- volume XI. CINECA, Bologna 2007.
15. Bruno G, Merletti F, Barbero G et al. Changes over time in the prevalence and quality of care of type 2 diabetes in Italy: the Casale Monferrato surveys, 1988 and 2000. *Nutr Metab Cardiovasc Dis* 2006. doi:10.1016/j.numecd.2006.08.007.
16. Gnani R, Karaghiosoff L, Costa G et al. Socio-economic differences in the prevalence of diabetes in Italy: the population-based Turin study. *Nutr Metab Cardiovasc Dis* 2008; doi:10.1016/J.numecd.2007.11.004.
17. Saydah SH, Geiss L, Tierney ED et al. Review of the performance of methods to identify diabetes cases among vital statistics, administrative, and survey data. *Ann Epidemiol* 2004; 14: 507-16.
18. Rosenbloom AL, Joe JR, Young RS, Winter WE. Emerging epidemic of type 2 diabetes in youth. *Diabetes Care* 1999; 22: 345-54.