

I Corso introduttivo

Uso epidemiologico delle fonti di dati sanitari correnti

Cristina Canova cristina.canova@unipd.it

Paolo Rigobello paolo.rigobello@unipd.it

Alessandro Barchielli alessandro.barchielli@asf.toscana.it

Algoritmi

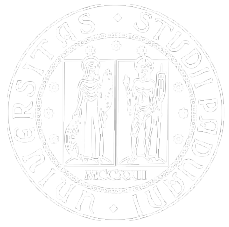
per gli studi epidemiologici



Indice

Introduzione, lavoro E&P

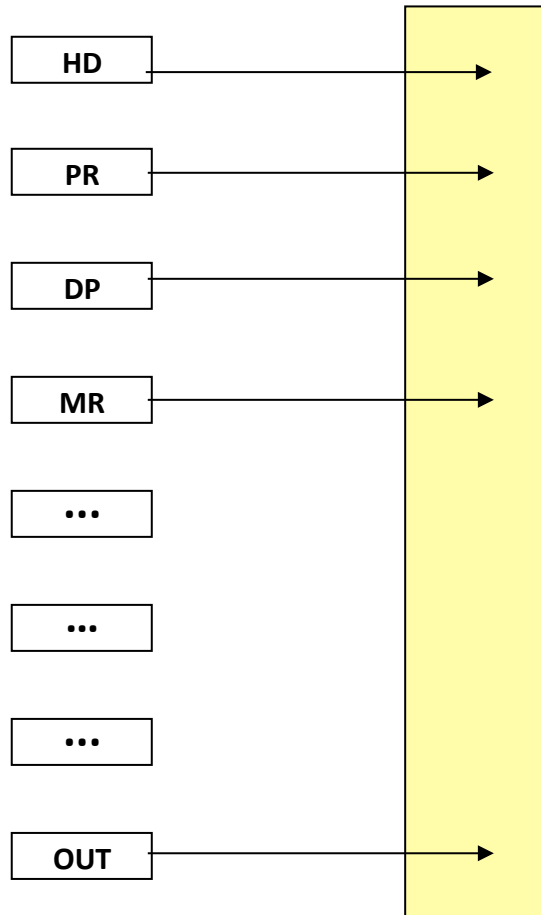
- 1) Esempio stima prevalenza diabete utilizzando Mysql.
- 2) Esempio stima BPCO utilizzando diverse definizioni di caso.
- 3) Esempio utilizzo algoritmi identificazione casi per studi osservazionali.
- 4) Validazione della casistica.



Algoritmi identificazione casi

(DATA SOURCES)

VALIDATION POPULATION REGISTRY



SUMMARIZATION

ID	HD1	HD2	DP	MR
ID	HD1	HD2		
ID	HD	PR		
ID	PR1	PR2	DP1	DP2
ID	DP1	DP2	DP3	
ID	PR1	PR2	HD	DP
ID	MR			
ID	PR	DP	MR	

CONSOLIDATION

ID	ICD1
ID	ICD2
ID	ICD3
ID	ICD4
ID	ICD5
ID	ICD6
ID	ICD...

RECORD KEEPING

To estimate incidence/prevalence



UNIVERSITÀ

EPIDEMIOLOGIA & PREVENZIONE

Rivista dell'Associazione Italiana di epidemiologia

ANNO 32 (3) MAGGIO-GIUGNO 2008 SUPPLEMENTO

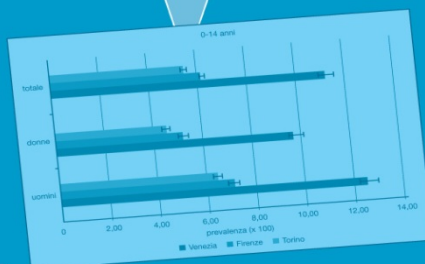


Utilizzo epidemiologico di archivi sanitari elettronici correnti

Un'esperienza di standardizzazione
delle procedure per la stima
di frequenza di alcune malattie
in diverse aree italiane

Exploiting electronic health archives for epidemiological purposes

An experience using a standardized
approach to estimate diseases
in various areas of Italy



A cura di:
Lorenzo Simonato
Cristina Canova
Gianni Corrao
Giuseppe Costa
Roberta Tessari

i inferenze

AIE ASSOCIAZIONE
ITALIANA DI
EPIDEMIOLOGIA

sismec
Società Italiana
di Statistica Medica
ed Epidemiologia Clinica

PATOLOGIE NON NEOPLASTICHE

Infarto miocardico acuto

Cardiopatía ischemica

Ictus

Diabete

Asma

BPCO

Patologie polmonari ostruttive
(MPCO)

* *Simonato et al. Epidemiol Prev. 2008*

FONTI	
APF	Archivio Prescrizioni farmaceutiche
SDO	Schede di dimissione ospedaliera ordinarie/Day Hospital + Mobilità Passiva
CM	Certificati di Morte
ET	Esenzioni ticket

Stime di prevalenza annuali (cardiopatìa, diabete, asma, BPCO, MPCO)

Stime di incidenza annuali con intervallo di 60 mesi dall'ultimo evento per controllo prevalenza (IMA, ictus)

PERIODO 2002-2004

* *Simonato et al. Epidemiol Prev. 2008*

STANDARDIZZAZIONE

- Per età (classi quinquennali)
- Censimento 2001

OUTPUT

- Tassi grezzi e standardizzati per anno e sesso
- Tassi specifici per età (0-14, 15-34, 35-54, 55-64, 65-74, 75+)
- Contributo di ciascuna fonte totale* ed esclusivo** per anno, sesso, e classi di età

****il contributo totale si calcola per ciascuna fonte come numero di soggetti presenti in quella fonte nell'anno considerato. La somma dei contributi totali per fonte è $\geq 100\%$ dei casi***

*****il contributo esclusivo si calcola per ciascuna fonte considerando i soggetti presenti solamente con quella fonte nell'anno considerato. La somma dei contributi esclusivi per fonte + contributo soggetti identificati da più di una fonte è $= 100\%$ dei casi***

**** Simonato et al. Epidemiol Prev. 2008***

Esempio output CARDIOPATIA ISCHEMICA

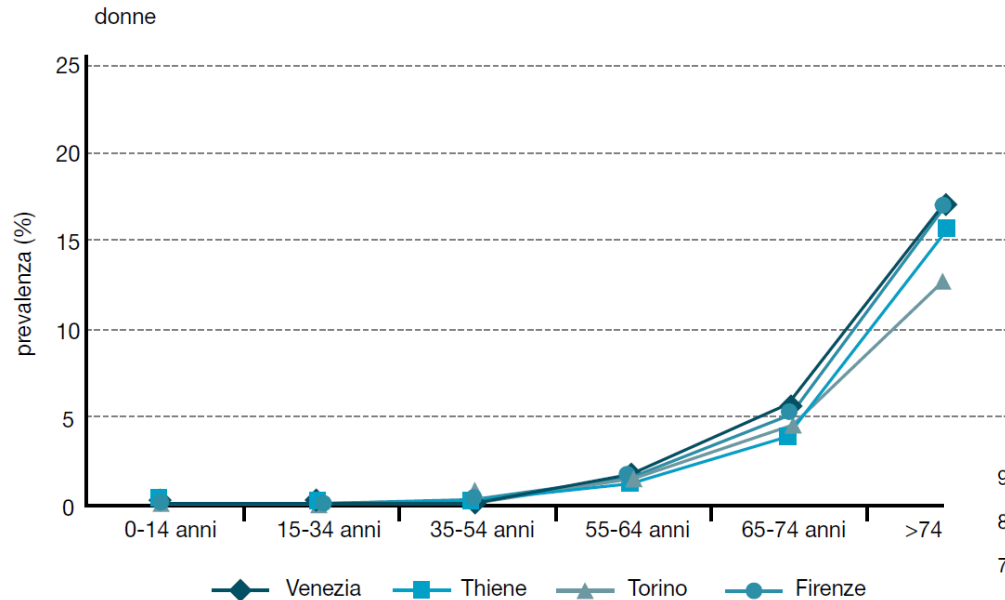
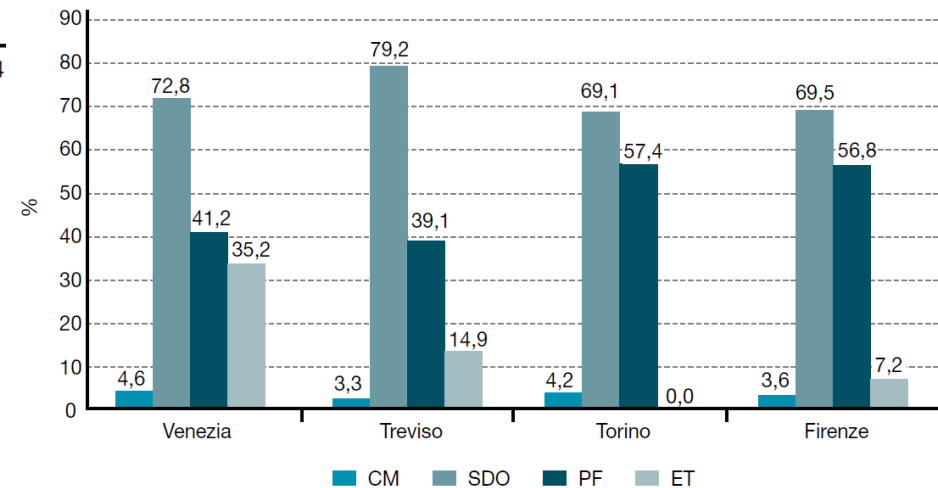


Figura 1. Prevalenza grezza di cardiopatia ischemica per area, sesso ed età. Anno 2003.



CM: certificati di morte; *causes of death*

SDO: schede di dimissione ospedaliera; *hospital discharges*

PF: prescrizioni farmaceutiche; *drug prescriptions*

ET: esenzione ticket; *health-tax exemption*

*** Balzi et al. Epidemiol Prev. 2008**

1) Stima prevalenza diabete

-utilizzo Mysql-

Sistema integrato con Open Source Software

- Mysql server 5.1 (server DBMS open source, acquisito ora da oracle)
- Mysql workbench 5.2 (GUI di accesso)
- Linguaggio principalmente utilizzato sql (standard a parte qualche eccezione)
- Possibilità di definire utenti con permessi diversi
- Repository centralizzato

Setup del sistema

- Varie fonti disponibili in dataset sas o db access
- Normalizzazione delle tabelle e convenzione sui nomi (documentazione)
- Import dati usando librerie sas o strumenti di export per access e strumenti ETL
- Se le tabelle non fossero linkate è necessario eseguire le procedure di record linkage (nel nostro caso già eseguito)
- Definizione di indici, viste, e funzioni

Algoritmo stima prevalenza diabete*

Fonte	Criteri selezione casistica	Criteri definizione prevalenza
schede di dimissione ospedaliera	dimissione nell'anno di stima o nei 4 anni precedenti con diagnosi di diabete (ICD9-CM ¹ = 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 prescrizione nell'anno di stima per farmaci antidiabetici (ATC ² = A10*)	
esenzioni ticket	esenzione per diabete (codice esenzione = 013.250) nell'anno di stima o nei tre anni precedenti e non deceduto al 1 gennaio dell'anno di stima	

¹ International classification of diseases IX revision

² Anatomical Therapeutic Chemical classification

* *Gnavi et al. Epidemiol Prev. 2008*

AS = anno di stima

SDO

dia* = 250*
(AS - 4) <= anno <= AS
[anno di morte] >= AS

APF

[COD ATC] = A10*
anno = AS
(num_prescr) >= 2

ET

Cod_esenzione = 013.250
(AS - 3) <= anno <= AS
[anno di morte] >= AS

soggetti_diabete

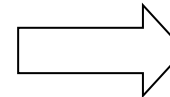
soggetti_diabete	
tessera_sanitaria	VARCHAR(10)
data_nascita	DATE
Sesso	VARCHAR(1)
g_eta	VARCHAR(10)
anno	BIGINT(20)
fonte_apf	DECIMAL(41,0)
fonte_sdo	DECIMAL(41,0)
fonte_et	DECIMAL(41,0)
apf	INT(1)
sdo	INT(1)
et	INT(1)
contr_escl	INT(1)
Indexes	

STIME

•pop_anno_eta
•pop_anno_sesso

REPORT

•contr_assoluto
•contr_escl_anno
•contr_escl_sesso



Esempio MySQL

2) Stima prevalenza BPCO
-utilizzo diverse definizioni di caso-

Algoritmo stima prevalenza BPCO soggetti 35+*

Fonte	Criteri selezione casistica	Criteri definizione prevalenza
schede di dimissione ospedaliera	essere stati dimessi nell'anno di stima o nei quattro anni precedenti se vivi al 1 gennaio dell'anno in stima, con diagnosi principale o secondaria di BPCO, (codici ICD 9-CM ¹ = 490*-492*, 494*, 496*)	presenza in almeno una delle fonti in soggetti età al 1 gennaio anno stima 35+
mortalità	essere deceduti nell'anno di stima per BPCO (codici ICD9-CM ¹ = 490*- 492*, 494*, 496*, ICD10 ² = J40-J44)	

¹ International classification of diseases IX revision

² International classification of diseases

Risultati prevalenza BPCO 2 fonti

Area	Anno	Uomini					Donne					Totale				
		n. casi	TG	TS	IC 95%		n. casi	TG	TS	IC 95%		n. casi	TG	TS	IC 95%	
Torino	2002	7.831	3,08	3,33	3,26	3,40	5.378	1,75	1,46	1,42	1,49	13.209	2,35	2,19	2,15	2,23
	2003	7.699	3,00	3,23	3,16	3,31	5.384	1,74	1,43	1,39	1,47	13.083	2,31	2,13	2,09	2,17
	2004	7.428	2,86	3,08	3,01	3,15	5.248	1,68	1,38	1,34	1,42	12.676	2,22	2,04	2,01	2,08
Venezia	2002	1.925	2,11	2,31	2,21	2,41	1.637	1,49	1,19	1,13	1,24	3.562	1,77	1,62	1,57	1,67
	2003	1.934	2,10	2,28	2,18	2,38	1.666	1,51	1,18	1,13	1,24	3.600	1,78	1,60	1,55	1,66
	2004	1.838	1,97	2,13	2,03	2,22	1.614	1,44	1,13	1,07	1,18	3.452	1,68	1,51	1,46	1,56
Pisa^	2001	778	3,01	3,09	3,07	3,11	371	1,16	0,91	0,90	0,92	1.149	1,99	1,76	1,75	1,77
	2002	826	3,18	3,25	3,23	3,28	401	1,25	0,98	0,97	0,99	1.227	2,12	1,87	1,86	1,88
	2003	859	3,29	3,33	3,31	3,36	406	1,26	0,99	0,98	1,00	1.265	2,17	1,90	1,89	1,91
Firenze*	2004	7.152	2,94	2,98	2,91	3,04	5.157	1,81	1,39	1,35	1,43	12.309	2,33	2,03	2,00	2,07
Roma	2002	34.689	4,90	5,53	5,48	5,58	27.194	3,15	2,83	2,80	2,87	61.883	3,94	3,92	3,89	3,94
	2003	35.597	4,98	5,65	5,60	5,70	28.799	3,30	2,97	2,94	3,01	64.396	4,05	4,04	4,01	4,07
	2004	35.737	4,99	5,65	5,60	5,71	29.399	3,34	3,01	2,98	3,04	65.136	4,08	4,06	4,03	4,09
Taranto	2002	3.065	5,91	6,95	6,92	6,98	1.787	2,90	2,91	2,90	2,93	4.852	4,28	4,61	4,59	4,62
	2003	3.329	6,36	7,39	7,37	7,42	2.065	3,32	3,28	3,27	3,30	5.394	4,71	5,01	5,00	5,02
	2004	3.448	6,55	7,49	7,46	7,52	2.219	3,54	3,47	3,46	3,49	5.667	4,91	5,17	5,16	5,18

CONCLUSIONI: dati sottostimati, algoritmo identifica casi medio gravi di malattia

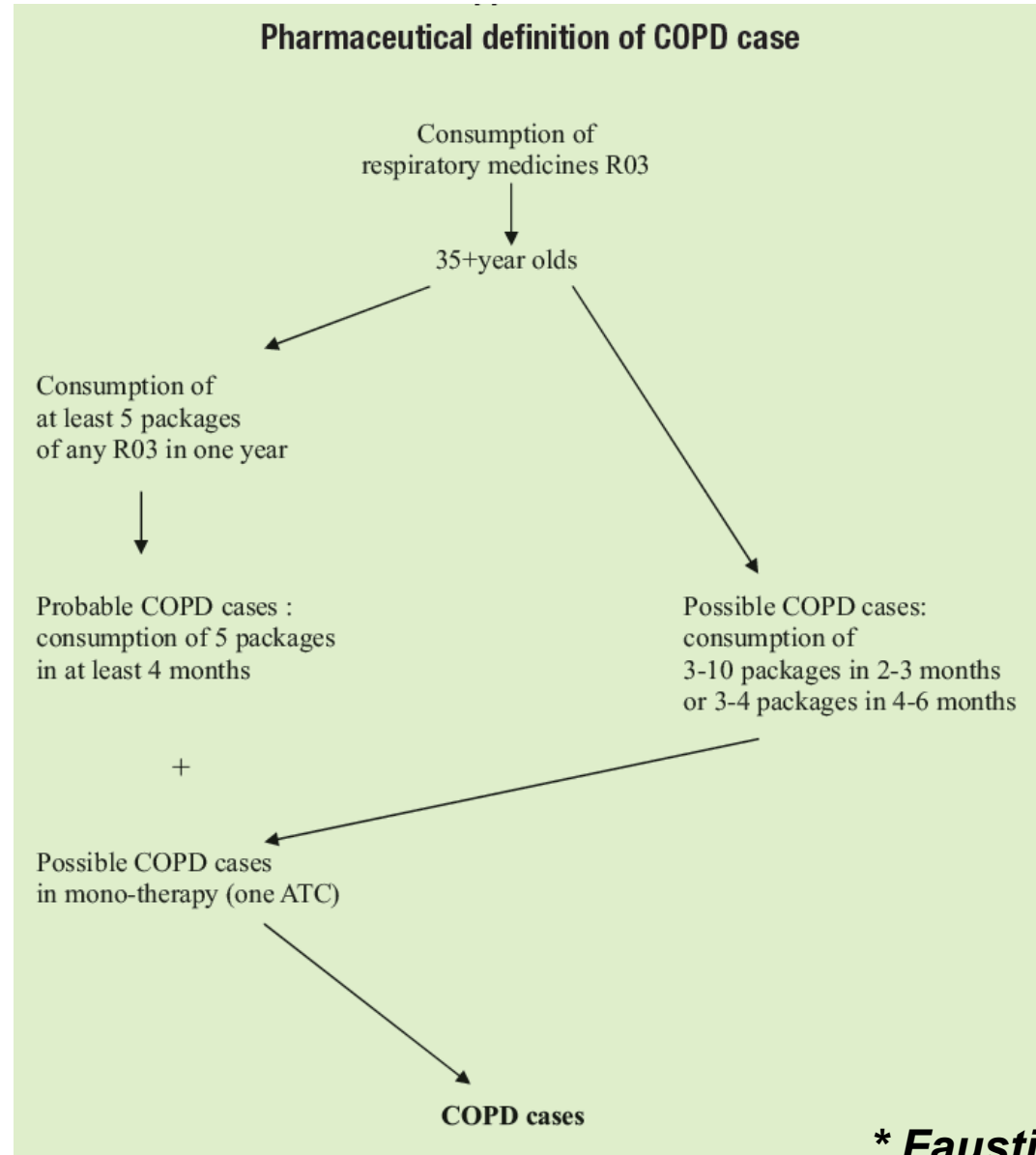
** Faustini et al. Epidemiol Prev. 2008*

Estensione algoritmo prevalenza BPCO 3 fonti

OBIETTIVI

- Introduzione fonte farmaci con algoritmo 'gold';
- Stimare l'incremento della prevalenza di BPCO introducendo la fonte farmaci;
- Analizzare la possibile selezione e misclassificazione dei casi di BPCO da fonti amministrative correnti.

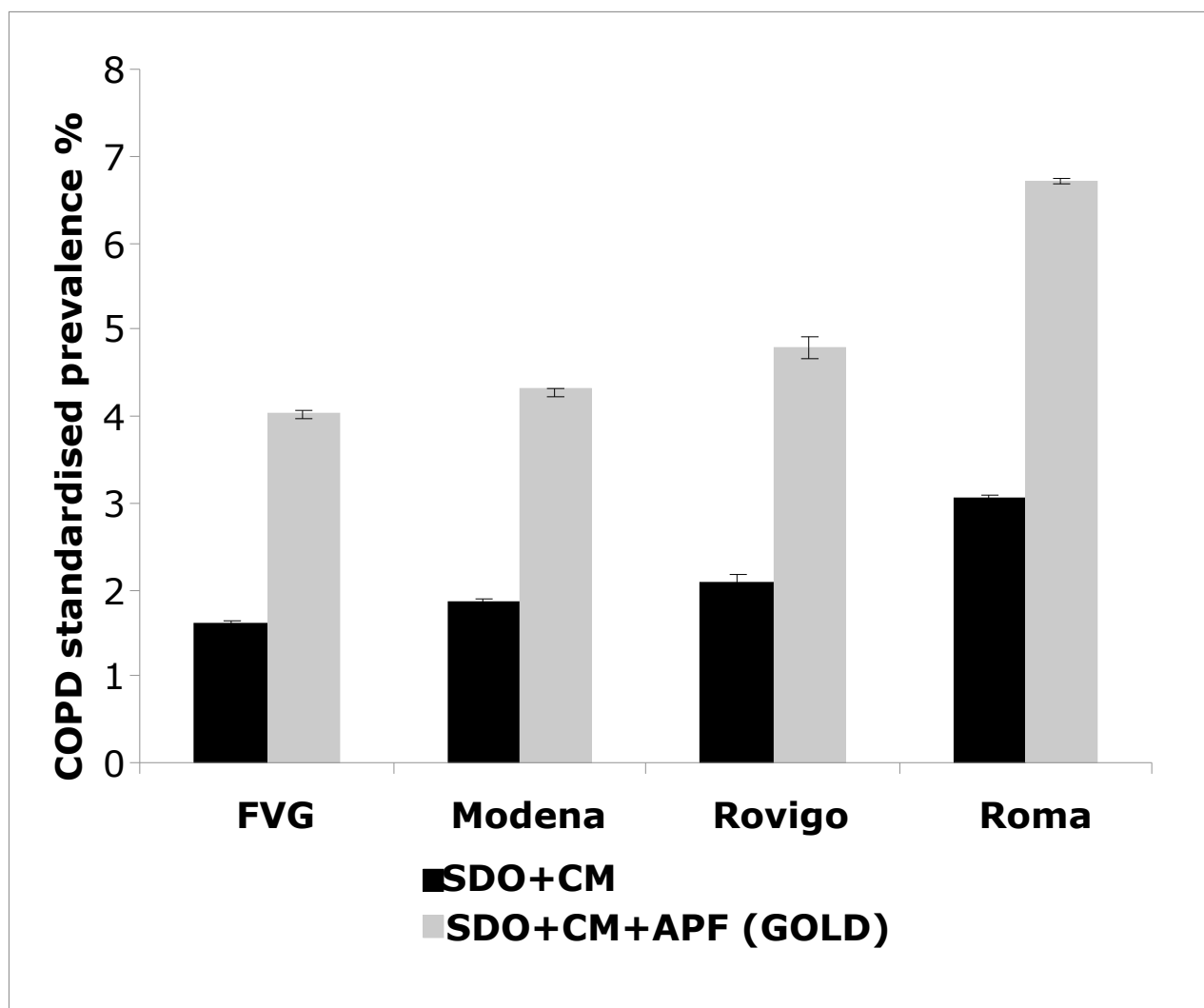
Algoritmo 'gold' APF BPCO



* *Faustini et al. COPD. 2012*

Risultati (1)

prevalenza BPCO 3 fonti



* *Faustini et al. COPD. 2012*

Metodi-Valutazione misclassificazione

Box 1		COPD pharmaceutical cases	
		YES	NO
COPD hospital or dead cases	YES	a	c
	NO*	b	??
		a+b	

a = subject who were discharged or dead with COPD and took medicines following discharge or preceding death, according to pharmaceutical COPD case definition
 c = COPD hospital or dead cases who did not take medicines
 b = COPD pharmaceutical cases who had no COPD hospital diagnosis nor died from COPD

* included those not hospitalised in 2002-2006 or discharged with diagnosis other than COPD, respiratory failure or congestive heart failure.

- Per stimare l'entità della misclassificazione dei casi di BPCO da fonti amministrative, si è assunto che chi avesse una diagnosi di dimissione di BPCO (o diagnosi morte) e prescrizione farmaceutica avesse una probabilità maggiore di essere un vero caso di BPCO.

Sono state stimate:

- Proporzione di casi BPCO da fonte APF con precedente dimissione ospedaliera nei 5 anni precedenti o diagnosi morte: $a/a+b$ SENSIBILITA'
- Proporzione di soggetti con diagnosi ospedaliera o da fonte mortalità che avessero preso farmaci respiratori: $a/a+c$ VALORE PREDITTIVO POSITIVO

* **Faustini et al. COPD. 2012**

Risultati (2)

Table 3. Pooled results: % of pharmaceutical COPD cases who were discharged with a COPD or died from COPD, by hospital case definition and age, 2006.

COPD hospital case definition		COPD cases		sensitivity			p-value heterogen ^b
		all pharm	hosp/dead using drugs				
		N	N	% ^a	95% CI		
Any COPD diagnosis							
age (yrs)	35+	120492	25233	22.7	19.9	25.9	0.832
	35-64	47663	3358	7.2	6.4	8.0	
	65-84	63386	18315	31.0	28.0	34.3	
	85+	9443	3560	40.0	34.5	46.4	
Principal diagnosis of COPD							
age (yrs)	35+	120492	8691	7.3	6.9	7.3	0.220
	35-64	47663	1365	2.6	2.2	3.0	
	65-84	63386	6145	9.7	8.8	10.8	
	85+	9443	1181	13.0	11.0	15.2	

* *Faustini et al. COPD. 2012*

Risultati (3)

Table 4. Pooled results: % of hospitalised or dead COPD cases who consumed drugs after discharge or before death, by hospital case definition and age, 2006.

COPD hospital case definition		COPD cases		PPV			p-value heterogen
		all hosp/ dead	hosp/dead, using drugs	%	95% CI		
		N	N				
Any COPD diagnosis							
age (yrs)	35+	77098	25233	37.2	29.4	47.1	0.016
	35-64	13104	3358	29.3	23.7	36.3	
	65-84	50414	18315	41.7	34.8	49.9	
	85+	13580	3560	28.5	22.0	36.9	
Principal diagnosis of COPD							
age (yrs)	35+	20465	8691	44.1	38.2	51.0	0.021
	35-64	4371	1365	32.9	25.6	42.2	
	65-84	12708	6145	50.9	45.7	56.7	
	85+	3386	1181	35.1	30.9	39.9	

* *Faustini et al. COPD. 2012*

Conclusioni algoritmo BPCO

- 1) L'uso dei dati di ospedalizzazione e mortalità sottostima la prevalenza della BPCO, secondo l'atteso.
- 2) La sensibilità dei casi SDO-CM è molto bassa, aumenta all'aumentare dell'età e non eccede mai il 40%. Questo indica che i dati da fonte SDO-CM sottostimano la prevalenza di BPCO. La probabilità che i casi di BPCO che consumano farmaci respiratori siano stati ospedalizzati nei 5 anni in studio era bassa (23%).
- 2) I casi di BPCO ospedalizzati mostrano un profilo meno grave di quanto atteso. La probabilità che i casi ospedalizzati o deceduti per BPCO consumino farmaci respiratori 'gold' è del 37% (VVP).
- 3) Studi differenti richiedono differenti caratteristiche di validità dei dati: stimare la prevalenza richiede maggiore sensibilità, arruolare coorti per studi di follow-up richiede maggiore affidabilità.

*3) Utilizzo fonti correnti per studi
osservazionali*

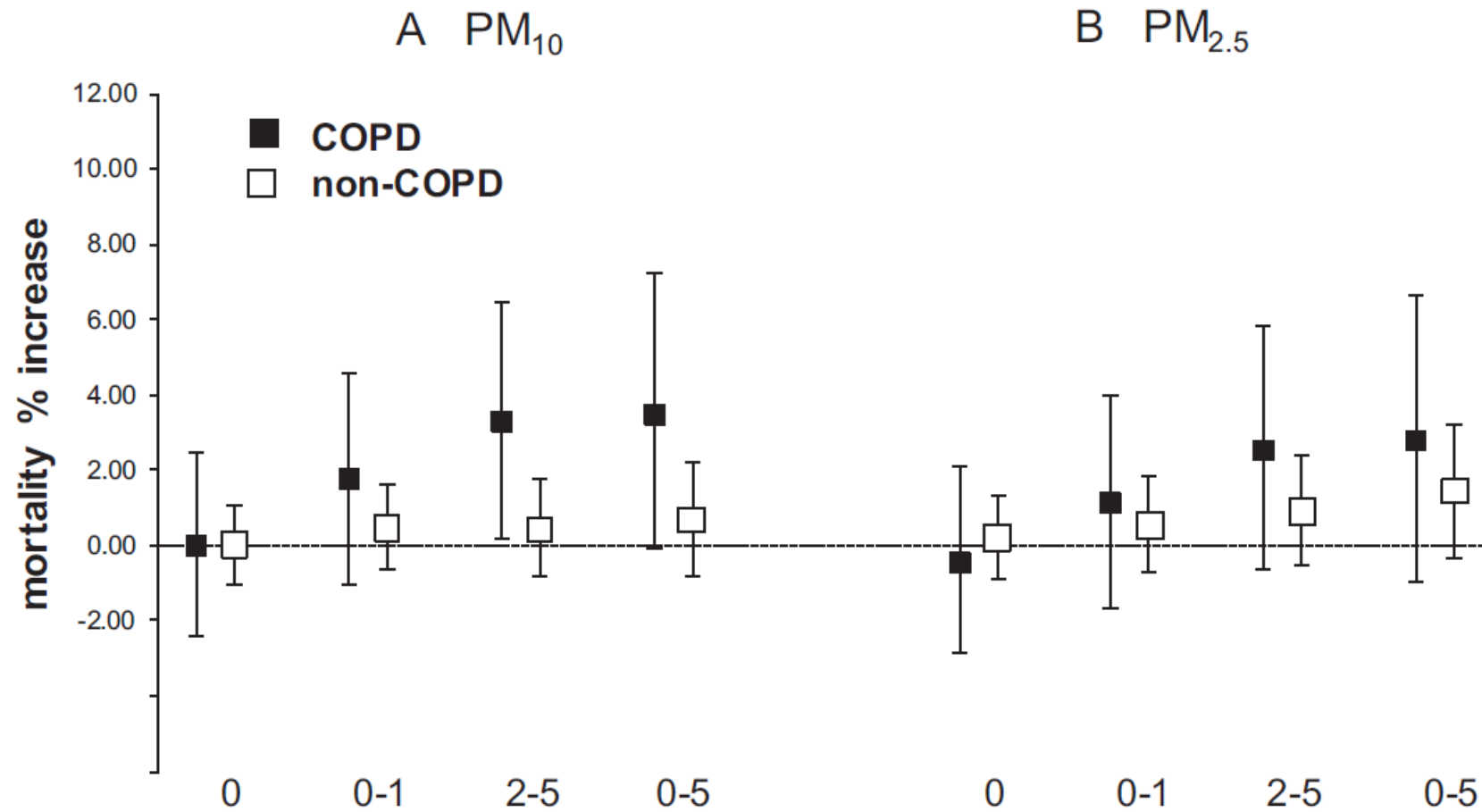
Short-Term Effects of Air Pollution in a Cohort of Patients With Chronic Obstructive Pulmonary Disease

Annunziata Faustini, Massimo Stafoggia, Giovanna Cappai, and Francesco Forastiere

- Studio case-crossover sull'effetto breve termine inquinamento atmosferico
- Fattore suscettibilità: avere o meno una precedente diagnosi di BPCO da fonte farmaceutica

** Faustini et al. Epidemiology 2012*

Risultati principali



* ***Faustini et al. Epidemiology 2012***

Mortality, incidence of cardiovascular diseases, and educational level among the diabetic and non-diabetic populations in two large Italian cities

R. Gnavi^{a,*}, C. Canova^b, R. Picariello^a, R. Tessari^c, C. Giorda^d, L. Simonato^b, G. Costa^{a,e}

- Studio coorte su popolazione diabetici (identificati da algoritmo E&P) e non diabetici.
- Effetto determinante socio-economico (livello istruzione) sulla sopravvivenza nelle due popolazioni.
- Follow-up per mortalità e incidenza IMA e Ictus 2002-2006

* *Gnavi et al. Diab Res and Clin Prac 2011*

Risultati principali

Table 3 – Mortality and incidence among diabetic and non-diabetic populations by educational level; hazard ratios (HR) and 95% CI (95% CI); Turin and Venice, 2002–2005.

	Diabetic population			Non-diabetic population		
	Medium ^a HR (95%CI)	Low ^a HR (95% CI)	p for trend	Medium ^a HR (95% CI)	Low ^a HR (95% CI)	p for trend
Men						
Mortality						
All	1.19 (1.08–1.32)	1.26 (1.15–1.38)	<0.0001	1.27 (1.22–1.37)	1.46 (1.40–1.51)	<0.0001
Circulatory disease	1.30 (1.10–1.53)	1.28 (1.10–1.48)	0.017	1.15 (1.07–1.23)	1.26 (1.19–1.34)	<0.0001
CHD	1.29 (1.01–1.63)	1.14 (0.92–1.43)	0.661	1.16 (1.04–1.30)	1.30 (1.17–1.44)	<0.0001
CVD	1.23 (0.86–1.74)	1.50 (1.10–2.04)	0.007	1.13 (0.99–1.29)	1.10 (0.98–1.24)	0.326
Cancer	1.30 (1.08–1.55)	1.29 (1.09–1.52)	0.004	1.31 (1.22–1.40)	1.53 (1.44–1.63)	<0.0001
Incidence						
Ami	1.05 (0.87–1.28)	1.00 (0.83–1.20)	0.712	1.26 (1.17–1.36)	1.32 (1.22–1.42)	<0.0001
Stroke	1.02 (0.82–1.28)	1.14 (0.93–1.39)	0.153	1.30 (1.18–1.43)	1.43 (1.31–1.56)	<0.0001
Women						
Mortality						
All	1.03 (0.88–1.20)	1.05 (0.91–1.21)	0.635	1.07 (1.02–1.12)	1.16 (1.11–1.22)	<0.0001
Circulatory disease	1.08 (0.84–1.38)	1.10 (0.88–1.37)	0.522	1.05 (0.97–1.14)	1.17 (1.09–1.26)	<0.0001
CHD	1.23 (0.79–1.89)	1.25 (0.84–1.86)	0.405	1.07 (0.91–1.26)	1.24 (1.07–1.44)	0.0002
CVD	0.98 (0.61–1.55)	1.01 (0.67–1.53)	0.823	1.07 (0.94–1.22)	1.17 (1.03–1.32)	0.012
Cancer	1.15 (0.82–1.61)	1.02 (0.75–1.39)	0.610	1.00 (0.92–1.08)	1.04 (0.96–1.13)	0.049
Incidence						
Ami	0.98 (0.70–1.36)	1.00 (0.74–1.35)	0.985	1.24 (1.08–1.42)	1.29 (1.13–1.46)	0.0002
Stroke	0.98 (0.69–1.40)	1.14 (0.83–1.57)	0.160	1.23 (1.10–1.37)	1.37 (1.24–1.52)	<0.0001

^a High educational level as reference. HR adjusted for age and centre.

*** Gnani et al. Diab Res and Clin Prac 2011**

La validazione degli algoritmi

L'uso dei dati sanitari correnti per la identificazioni dei casi di malattia e per le stime di incidenza/prevalenza attraverso specifici algoritmi richiede una validazione dei risultati ottenuti (es. per errori di codifica di diagnosi di dimissione/causa di morte, farmaci utilizzati per più patologie ecc.).

La validazione può essere fatta solo con diagnosi con conferma clinicamente.

		Casi con conferma clinica		
		SI	NO	Totale
Casi identificati da algoritmo	SI	Veri Positivi	Falsi Positivi	Totale Casi + algoritmo
	No	Falsi Negativi	Veri Negativi	Totale Casi - algoritmo
	Totale	Totale Casi + clinici	Totale Casi - clinici	Totale generale

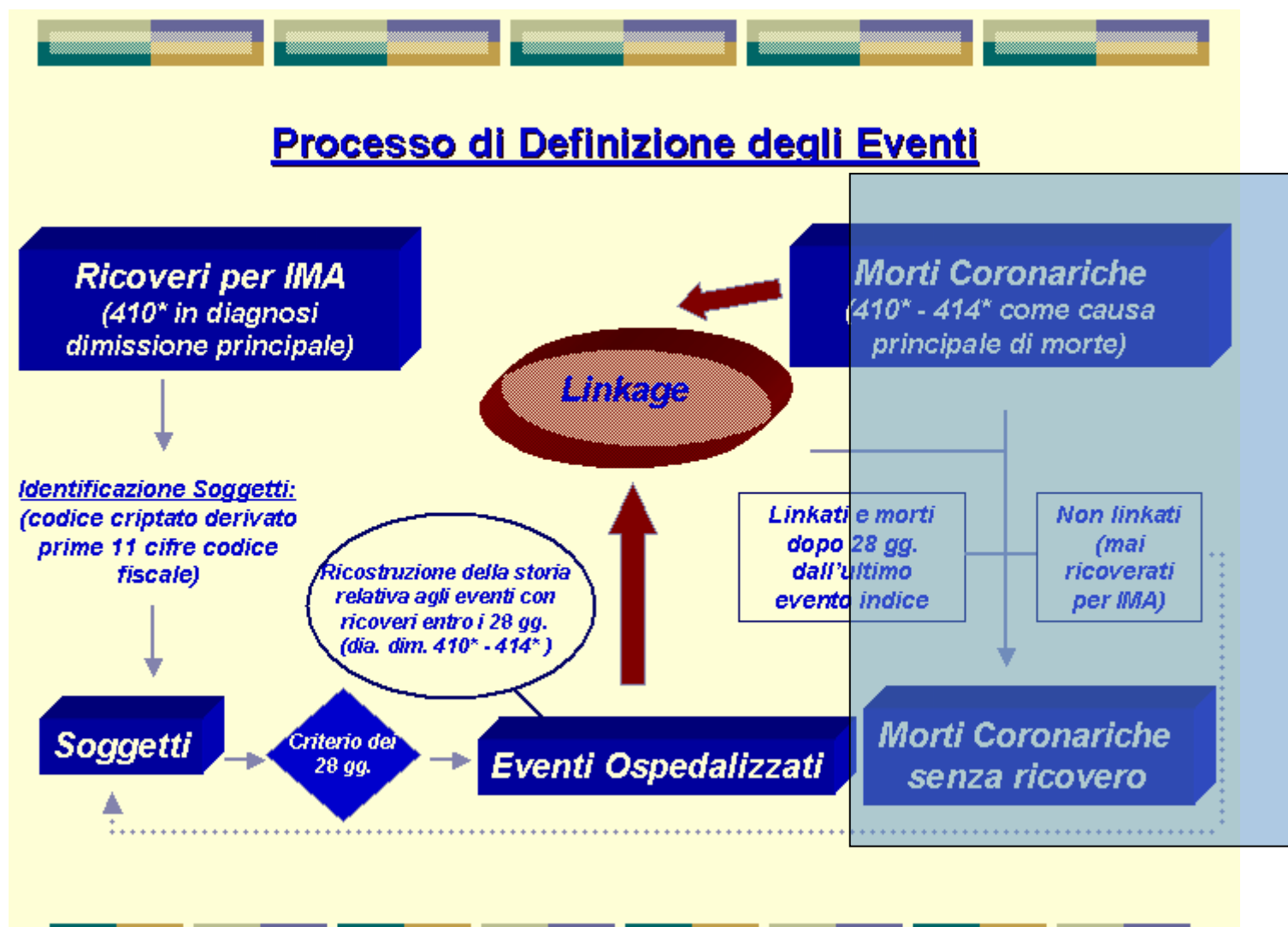
**Tasso: casi/
popolazione**

Schematicamente questo può essere fatto con due approcci:

Confronto singoli casi: **Veri positivi (VPP: % Veri positivi/ Totale Casi + algoritmo)**

Confronto frequenze/tassi: **Totale Casi + algoritmo = Totale Casi + clinici**

Un esempio dal Registro Toscano dell'Infarto miocardico acuto: la validazione degli eventi ospedalizzati



Hospital discharge data for assessing myocardial infarction events and trends, and effects of diagnosis validation according to MONICA and AHA criteria

J Epidemiol Community Health 2012;**66**:462–467. doi:10.1136/jech.2010.110908

Alessandro Barchielli,^{1,2} Daniela Balzi,^{1,2} Paola Naldoni,^{1,2} Anna Teresa Roberts,³ Francesco Profili,² Francesco Dima,⁴ Luigi Palmieri⁴

- **Revisione cartelle cliniche di un campione casuale di:**
 - **380 eventi ospedalizzati (DDP: 410*)**
 - **100 “eventi” ospedalizzati di Altre forme CI (DDP: 411, 413, 414)**

Criteri di diagnosi di IMA per studi clinici ed epidemiologici

Table 1 Some details of American Heart Association (AHA)⁶ definitions and diagnostic criteria for acute myocardial infarction (AMI)

AHA definitions	Definite AMI	Evolving diagnostic ECG or Diagnostic enzymes
	Probable AMI	Positive ECG and cardiac sign or symptoms, or Positive ECG and equivocal enzymes
	Possible AMI	Positive ECG or Equivocal enzymes
	Diagnostic enzymes: one positive test in an adequate set of biomarkers (at least 2 measurement of the same marker taken at least 6 h apart) in absence of non-cardiac >causes of elevation. Equivocal enzymes: only one positive measurement is available or in the presence of non-cardiac causes of elevation. Enzymes: troponin, CK-MB mass or CK-MB.	

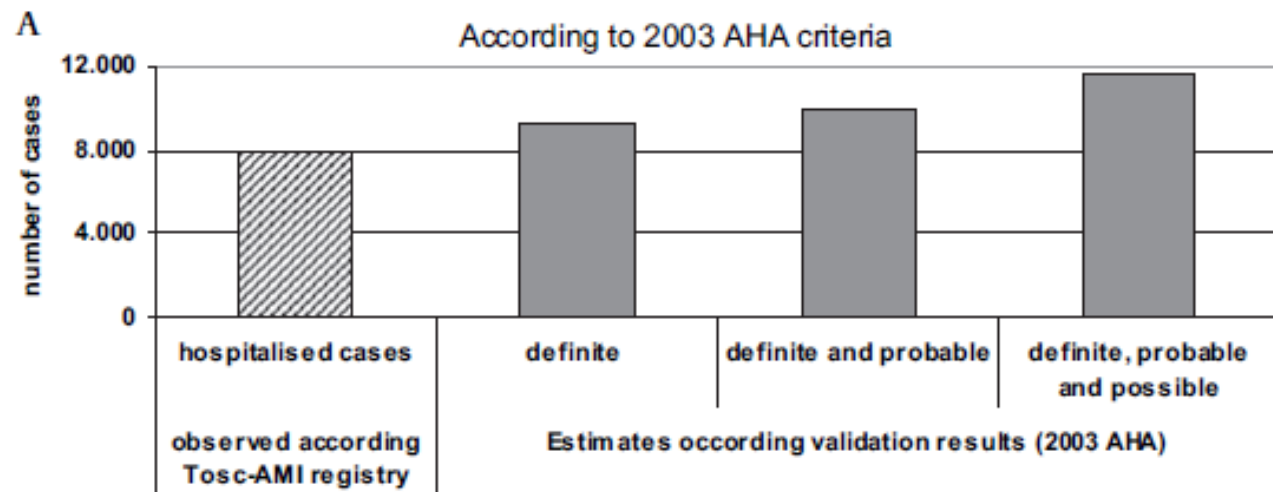
ECG
Enzimi
Sintomi

Risultati validazione

Table 2 Distribution of validation samples according to hospital discharge diagnoses and to diagnostic validation results (2003 American Heart Association (AHA))

Validation criteria	Acute myocardial infarction (AMI) diagnoses validated	Hospital discharge diagnoses (ICD9-CM)					
		Code 410 (sample 1)			Codes 411, 413, 414 (sample 2)		
		N	%	(95% CI)	N	%	(95% CI)
AHA criteria	Definite	320	86.0	(82.4 to 89.6)	16	16.5	(9.0 to 24.0)
	Definite or probable	325	87.3	(84.0 to 90.8)	21	21.6	(13.3 to 30.0)
	Definite or probable or possible	352	94.6	(92.3 to 96.9)	29	29.9	(20.6 to 32.9)
	No AMI	20	5.4	(3.1 to 7.7)	68	70.1	(60.8 to 79.4)
	Total	372	100.0	—	97	100.0	—

Figure 3 Comparison of the number of hospitalised acute myocardial infarction (AMI) events occurring in 2003, based on the Tosc-AMI Registry, and estimated according to the proportions of AMI-validated diagnosis of hospital discharge diagnosis provided from validation samples.



B

According to MONICA criteria

Conclusioni

Possibile utilizzo fonti amministrative correnti

- Stima occorrenza (confronto spaziale e temporale)
- Identificazione di soggetti per studi osservazionali

Necessità validazione

Bibliografia 1

- Balzi D, Barchielli A, Battistella G, Gnani R, Inio A, Tessari R, Picariello R, Canova C, Simonato L. *Ischemic heart disease prevalence estimated using a standard algorithm based on electronic health data in various areas of Italy*. Epidemiol Prev. 2008 May-Jun;32(3 Suppl): 22-9. Italian.
- Barchielli A, Balzi D, Naldoni P, Roberts AT, Profili F, Dima F, Palmieri L. *Hospital discharge data for assessing myocardial infarction events and trends, and effects of diagnosis validation according to MONICA and AHA criteria*. J Epidemiol Community Health. 2012 May;66(5):462-7. doi: 10.1136/jech.2010.110908. Epub 2010 Oct 19.
- Faustini A, Cascini S, Arcà M, Balzi D, Barchielli A, Canova C, Galassi C, Migliore E, Minerba S, Protti MA, Romanelli A, Tessari R, Vigotti MA, Simonato L. *Chronic obstructive pulmonary disease prevalence estimated using a standard algorithm based on electronic health data in various areas of Italy*. Epidemiol Prev. 2008 May-Jun;32(3 Suppl):46-55. Italian.
- Faustini A, Canova C, Cascini S, Baldo V, Bonora K, De Girolamo G, Romor P, Zanier L, Simonato L. *The reliability of hospital and pharmaceutical data to assess prevalent cases of Chronic Obstructive Pulmonary Disease*. COPD. 2012 Apr;9(2):184-96.
- Faustini A, Stafoggia M, Cappai G, Forastiere F. *Short-term effects of air pollution in a cohort of patients with chronic obstructive pulmonary disease*. Epidemiology. 2012 Nov;23(6): 861-79.
- Fusco D, Barone AP, Sorge C, D'Ovidio M, Stafoggia M, Lallo A, Davoli M, Perucci CA. *P.Re.Val.E.: outcome research program for the evaluation of health care quality in Lazio, Italy*. BMC Health Serv Res. 2012 Jan 27;12:25.

Bibliografia 2

- Gini R, Francesconi P, Mazzaglia G, Cricelli I, Pasqua A, Gallina P, Brugaletta S, Donato D, Donatini A, Marini A, Zocchetti C, Cricelli C, Damiani G, Bellentani M, Sturkenboom MC, Schuemie MJ. *Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey*. BMC Public Health. 2013 Jan 9;13:15.
- Gnani R, Karaghiosoff L, Balzi D, Barchielli A, Canova C, Demaria M, Pellizzari M, Rigon S, Tessari R, Simonato L. *Diabetes prevalence estimated using a standard algorithm based on electronic health data in various areas of Italy*. Epidemiol Prev. 2008 May-Jun;32(3 Suppl):15-21. Italian.
- Gnani R, Canova C, Picariello R, Tessari R, Giorda C, Simonato L, Costa G. Mortality, incidence of cardiovascular disease, and educational level among diabetic and non diabetic population in Turin and Venice. Diabetes Res Clin Pract. 2011 May;92(2):205-12. Epub 2011 Mar 5.
- Suissa S, Garbe E. *Primer: administrative health databases in observational studies of drug effects--advantages and disadvantages*. Nat Clin Pract Rheumatol. 2007 Dec;3(12):725-32.
- Simonato L, Baldi I, Balzi D, Barchielli A, Battistella G, Canova C, Cesaroni G, Corrao G, Collini F, Conti S, Costa G, Demaria M, Fornari C, Faustini A, Galassi C, Gnani R, Inio A, Madotto F, Migliore E, Minelli G, Pellizzari M, Protti M, Romanelli A, Russo A, Saugo M, Tancioni V, Tessari R, Vianello A, Vigotti MA. *Objectives, tools and methods for an epidemiological use of electronic health archives in various areas of Italy*. Epidemiol Prev. 2008 May-Jun;32(3 Suppl):5-14. Italian.
- Mysql server 5.1 : <http://dev.mysql.com/doc/refman/5.1/en/index.html> [consultato il 26/03/2013]
- Mysql workbench 5.2 : <http://www.mysql.com/products/workbench/> [consultato il 26/03/2013]