

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Il sistema SEI e la sua replicazione

Firenze, 24 Maggio 2012

Cristina Canova



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Outline presentazione

1. Storia nascita 'Sistema Epidemiologico Integrato' (SEI).
2. Progetto CCM 2010.
3. Esempio di utilizzo.



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Automated cancer registration

Case consolidation process

Based on three independent sources: hospital discharges, pathology records, death certificates.

Each source contributes one or more computerized record with coded diagnosis in ICD (for pathology the SNOMED is transformed in ICD).

The case resolution software assigns date of diagnosis and ICD code (three digits) if:

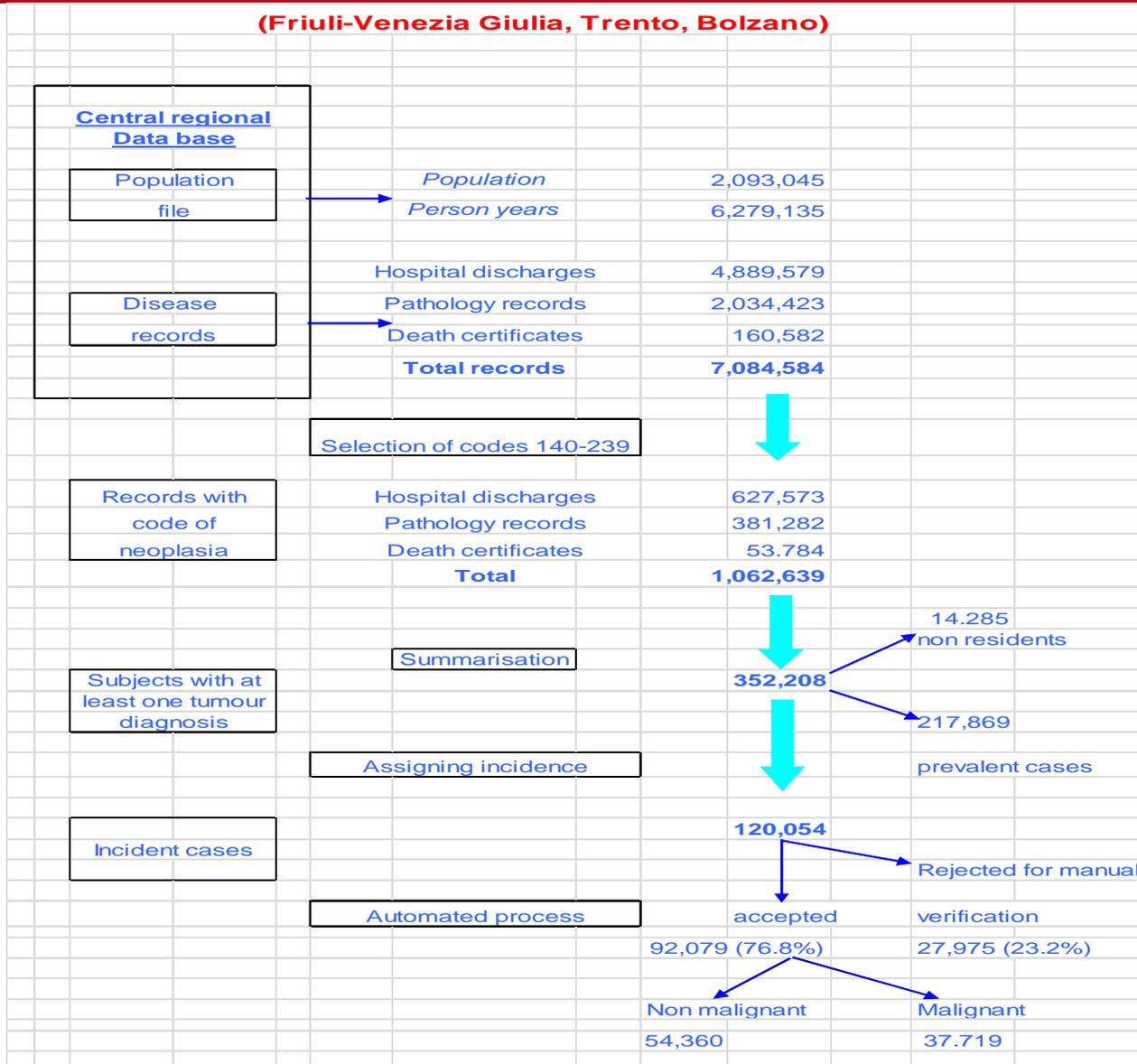
- A. At least two independent sources are concordant at three digits level;
- B. there is a ICD primary tumor diagnosis MV in presence of diagnosis of metastases or ill defined codes;
- C. the diagnosis is 153 MV in presence of 154;
- D. the diagnosis is 172 MV;
- E. the diagnosis is 173 MV;
- F. only one or more pathology records are available.

MV= Microscopically Verified



Esempio flusso ACR 2003-2005

(Friuli-Venezia Giulia, Trento, Bolzano)



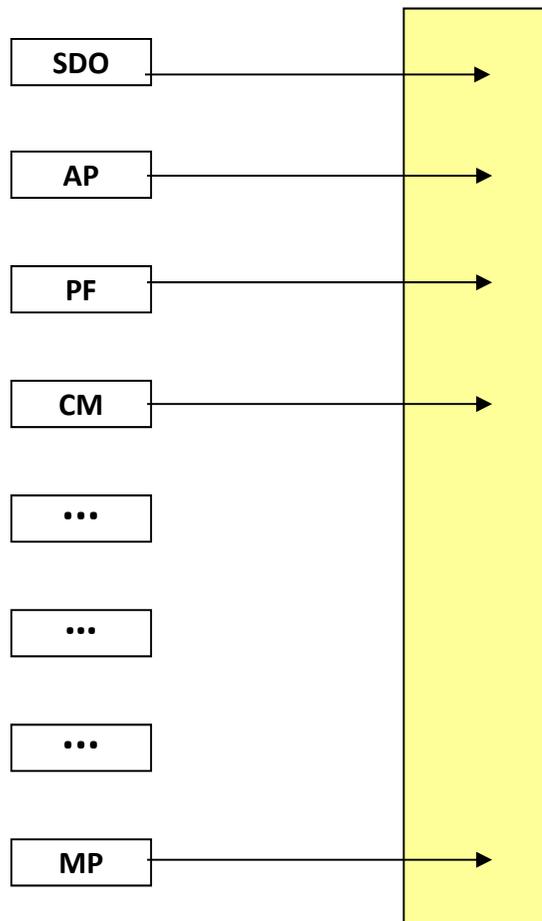


Applicabilità ad altre patologie

VALIDAZIONE

(FONTI)

(ANAGRAFE)



ALLINEAMENTO

ID	SDO1	SDO2	PF	CM
ID	SDO1	SDO2		
ID	SDO	AP		
ID	AP1	AP2	PF1	PF2
ID	PF1	PF2	PF3	
ID	AP1	AP2	SDO	PF
ID	CM			
ID	AP	PF	CM	

RICONOSCIMENTO

determina il caso

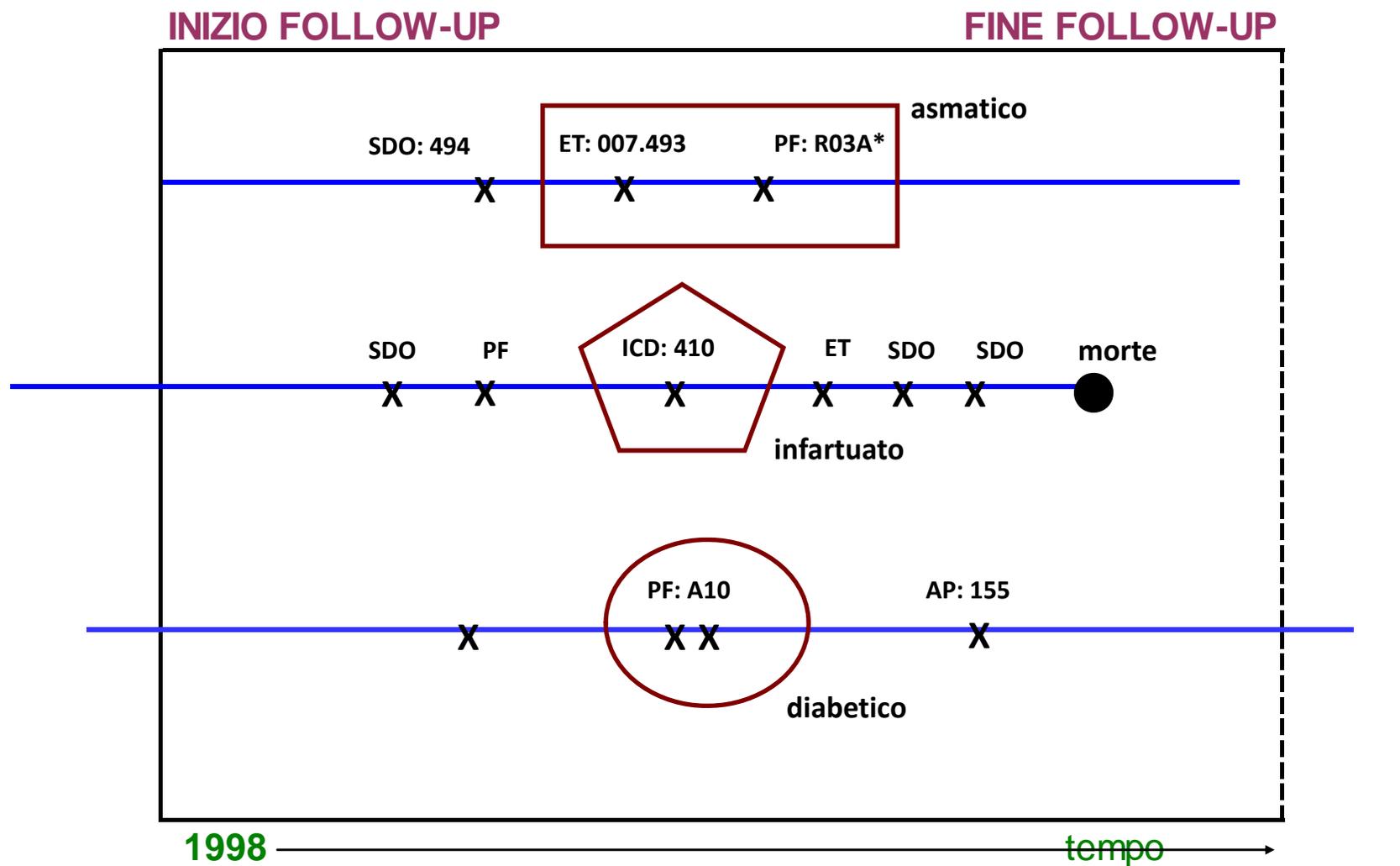
ID	ICD1
ID	ICD2
ID	ICD3
ID	ICD4
ID	ICD5
ID	ICD6
ID	ICD...

ARCHIVIAZIONE

produce misure di frequenza



Storia naturale della malattia



X evento sanitario



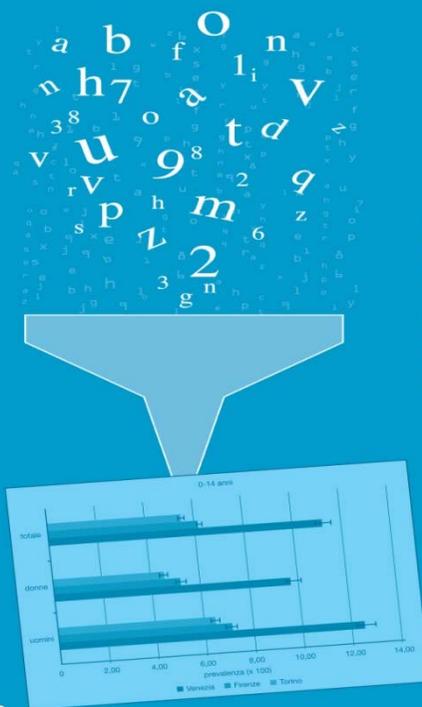
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Esperienza gruppo AIE&SISMEC Supplemento 2005-2008

EPIDEMIOLOGIA & PREVENZIONE

Rivista dell'Associazione italiana di epidemiologia ANNO 32 (3) MAGGIO-GIUGNO 2008 SUPPLEMENTO

Inferenza Edizioni - Via Riccardelli 29, 20148 Milano, Pagine Italiane spa - Sped. in abb. post. DL 353/2003 convertita in legge 27/02/04 n. 46 art. 1, comma 1, D.38 Milano - data copia 25.01.08 euro ISSN 1120-9753 maggio-giugno 2008



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

 inferenze

AIE ASSOCIAZIONE
ITALIANA DI
EPIDEMIOLOGIA

sismec
Società Italiana
di Statistica Medica
ed Epidemiologia Clinica

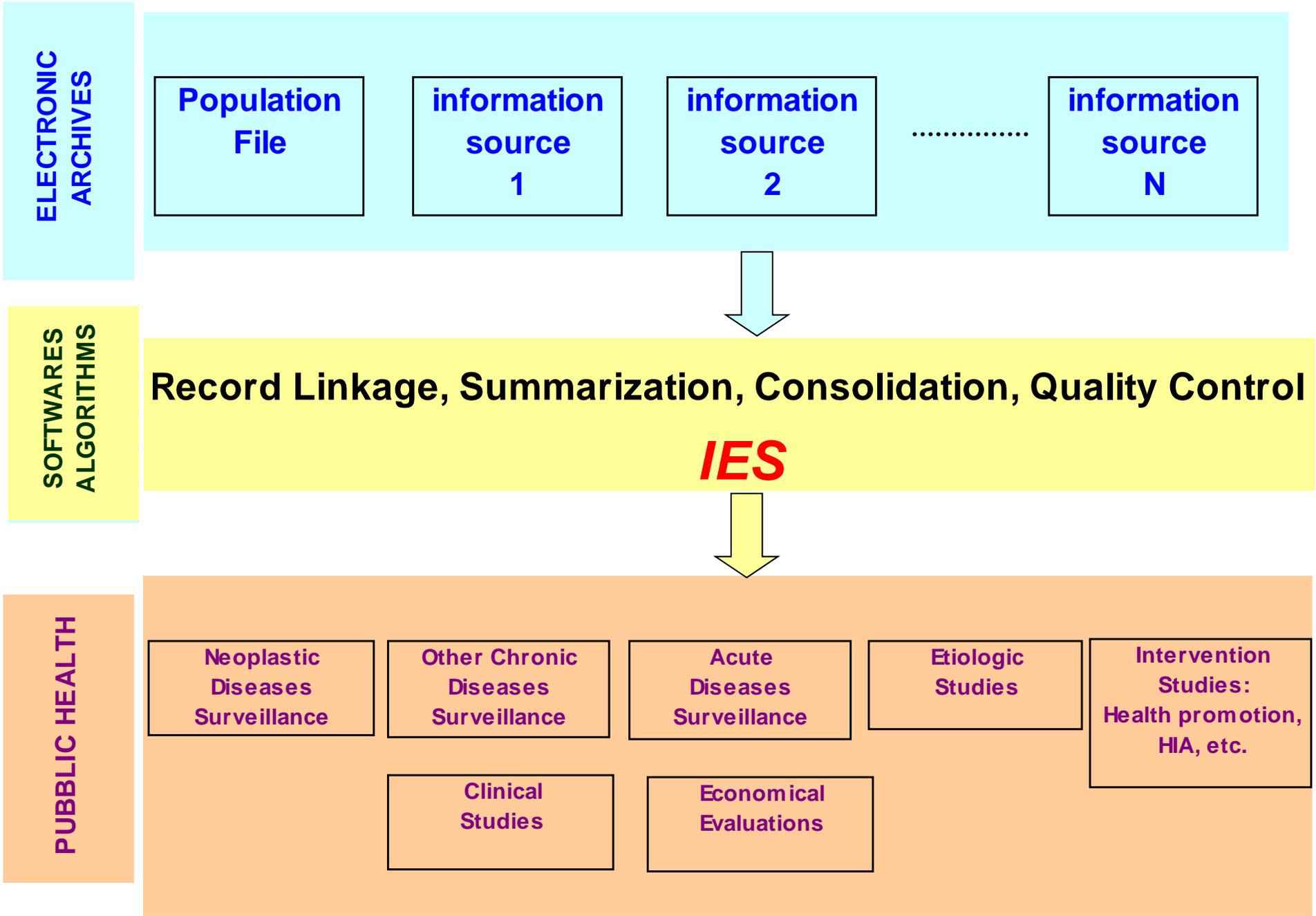


UNIVERSITÀ
DEGLI STUDI
DI PADOVA

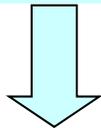
Esperienza gruppo AIE&SISMEC

Patologie Indagate

PATOLOGIA (n°centri)	REFERENTE
IMA (7)	Barchielli
Cardipatia ischemica (4)	Balzi
Ictus acuto (4)	Tancioni
Diabete (6)	Gnavi
Ca. mammella (6)	Simonato, Baldi
Ca. colon retto (4)	Simonato, Baldi
Asma (3)	Tessari
BPCO (5)	Faustini
MPCO (3)	Migliore



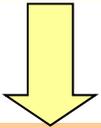
**ELECTRONIC
ARCHIVES**



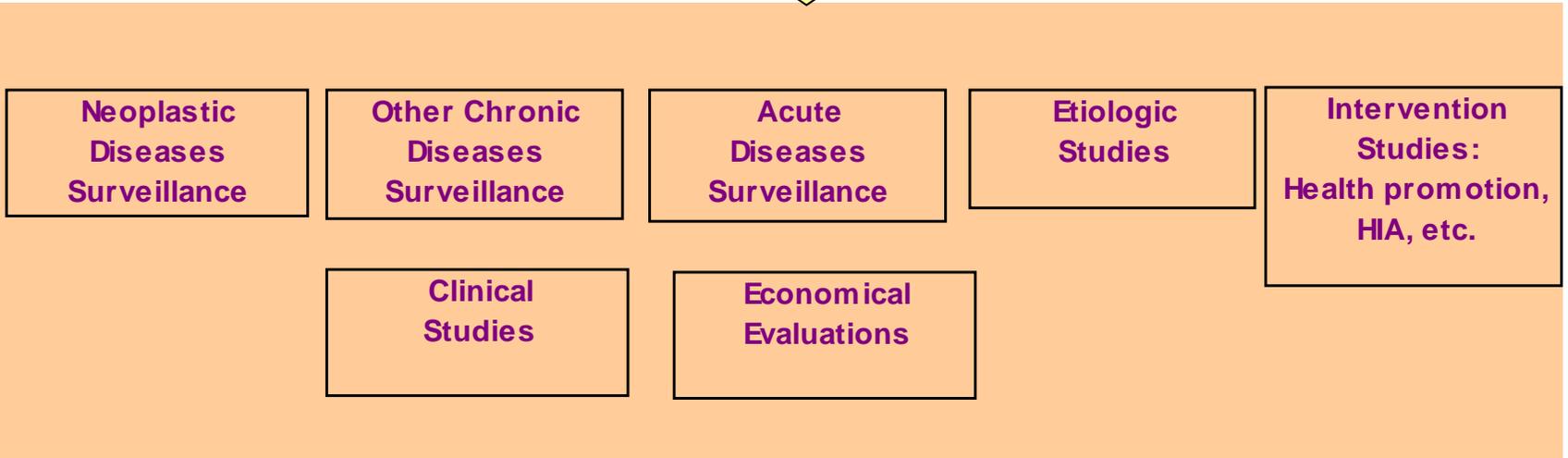
**SOFTWARES
ALGORITHMS**

Record Linkage, Summarization, Consolidation, Quality Control

IES



PUBLIC HEALTH





UNIVERSITÀ
DEGLI STUDI
DI PADOVA

IL SISTEMA EPIDEMIOLOGICO INTEGRATO (S.E.I.) DELL'ASL12 VENEZIANA 2006-2012

- Sintesi di diverse fonti informative elettroniche di dati sanitari incrociati con l'anagrafe sanitaria mediante RL deterministico a cascata
- dati epidemiologici georeferenziati rispetto alla residenza dell'assistito
- sistema interfacciabile con altri archivi (ad es.: dati Censimento, sistemi di rilevazione di inquinamento ambientale o fonti espositive puntuali, etc.)



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

LE FONTI INFORMATIVE DEL S.E.I. Veneziano

- Certificati di Morte (1987-)
- Referti Anatomico Patologici (1981-)
- Schede di Dimissione Ospedaliera (1982-)
- Archivio mobilità passiva (1982-)
- Ricoveri in regime di Day-Hospital (1998-)
- Archivio di consumo Farmaci (1998-)
- Esenzioni Ticket (1998-)
-
- **Anagrafe degli Assistiti**



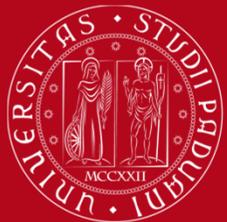
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

PROGETTO CCM 2008-2010

***‘SISTEMI DI SORVEGLIANZA EPIDEMIOLOGICA
BASATI SULL’UTILIZZO DEGLI ARCHIVI ELETTRONICI
SANITARI IN AMBITO DI SANITA’ PUBBLICA: UNO
STUDIO PILOTA IN ALCUNE ASL SELEZIONATE’***



- **OBIETTIVO GENERALE:** Attivare un sistema di sorveglianza epidemiologica basato su meccanismi di integrazione di dati correnti sanitari
- **OBIETTIVO SPECIFICO 1** Valutare la fattibilità di attivazione del sistema di sorveglianza proposto in alcune ASL selezionate esaminando gli archivi sanitari informatizzati necessari allo scopo sotto il profilo della disponibilità temporale, la completezza e le qualità
- **OBIETTIVO SPECIFICO 2** Trasferire al personale delle ASL coinvolte gli strumenti metodologici ed informatici necessari all'integrazione delle fonti correnti
- **OBIETTIVO SPECIFICO 3** Attivazione del sistema di sorveglianza e produzione dei risultati epidemiologici per le maggiori patologie presenti nella popolazione studiata
- **OBIETTIVO SPECIFICO 4** Coordinamento di un gruppo di esperti con compiti di ricerca e sviluppo dei metodi



ASL selezionate, popolazioni - CCM

- Il territorio dell'Azienda Sanitaria Locale 18 di Rovigo comprende 41 Comuni, tra i quali Rovigo, capoluogo di provincia.
- Il territorio dell'Azienda Unità Sanitaria Locale di Modena che coincide con l'intera provincia di Modena comprende 47 Comuni.
- Il territorio dell'Azienda Sanitaria Locale 2 di Perugia comprende 24 Comuni, tra i quali Perugia, capoluogo di provincia.

anno	Rovigo	Modena	Perugia
2001	170,106		
2002	170,106	634,713	330,564
2003	170,323	643,043	334,319
2004	171,428	651,996	340,976
2005	172,151	659,925	347,870
2006	172,420	665,367	355,815
2007	172,737	670,098	357,008
2008	174,045	677,672	362,915



1. sistemazione archivi
 - acquisizione degli archivi aziendali richiesti per il progetto
 - organizzazione dei dati in un unico dataset per archivio
 - valutazione della qualità dei dati
 - distribuzioni di qualità per l'Anagrafe Sanitaria
 - distribuzioni di qualità per gli altri archivi
2. record linkage degli archivi con l'Anagrafe Sanitaria
 - valutazione storicizzazione dell'archivio anagrafico
 - individuazione e valutazione delle chiavi di linkage utilizzabili per singolo archivio
 - esecuzione di procedure di record linkage deterministico a cascata
 - valutazione delle percentuali di linkage
3. produzione di stime di frequenza
 - applicazione agli archivi di algoritmi forniti
 - confronto con stime ottenute applicando gli stessi algoritmi in altre realtà italiane



1 sistemazione archivi - CCM

1. Ottenuta copia degli archivi per tutti gli anni a disposizione senza alcuna selezione delle variabili presenti negli archivi originali. Specificare da chi sono ottenuti tali archivi (CED, controllo gestione,...).
2. Specificare il formato dei dati ricevuti e se gli archivi sono stati ottenuti in file separato per anno.
3. Per ogni archivio mantenere una copia originale. Riportare il tracciato record di ciascun database ricevuto per ciascun anno/periodo con il nome e il significato di ciascuna variabile.
4. Per ogni fonte valutata la qualità dei dati ricevuti mediante analisi descrittive, prima di accodare l'archivio di aggiornamento al database esistente.



Fonti a disposizione

FONTE	Modena		Perugia		Rovigo	
	DA	A	DA	A	DA	A
ANAGRAFE SANITARIA		2009	2006	2010	1996	2008
MORTALITA'	1996	2007	1994	2007	1987	2008
SCHEDE DIMISSIONI OSPEDALIERE (SDO)	2002	2008	1998	2008	1995	2008
MOBILITA' PASSIVA DELLE SDO (MP)	2002	2008	1998	2008	2000	2008
PRESCRIZIONI FARMACEUTICHE (PF)	2003	2008	2005	2008	1997	2008
ESENZIONI TICKET (ET)		2008		2010	1991	2008



1. Prima di procedere al record linkage fra gli archivi è necessario individuare le variabili da utilizzare nel linkage, per impostare l'algoritmo specifico per ogni archivio; inoltre, è necessario valutare la completezza dei campi per tessera sanitaria (TS) e codice fiscale (CF).
- Le chiavi di linkage che utilizzate sono state le seguenti:
 - **Tessera sanitaria (TS)**
 - **Codice fiscale (CF)**
 - **Nome**
 - **Cognome**
 - **Data di nascita**
- Non è da escludere possibili valutazioni di sensibilità del linkage utilizzando altre chiavi di linkage come ad esempio il comune di nascita o di residenza o sesso.
- Valutate per ogni fonte le possibili chiavi di linkage a disposizione, la loro completezza (per anno e nel totale) ed il loro formato.



2 Record Linkage - CCM

Passi di chiave-esempio Rovigo

Legenda:

T tessera sanitaria

C cognome

N nome

DN data di nascita

CF codice fiscale

nuovo_cf sequenza
di chiavi comprendenti il
CF ricalcolato

Passi di chiave	SDO	mort	MP	AP	ET
T C N DN					
T C DN					
T C N					
T N DN					
CF C N DN					
CF C DN					
CF C N					
CF N DN					
C N DN					
nuovo_cf C N DN					
nuovo_cf C DN					
nuovo_cf C N					
nuovo_cf N DN					



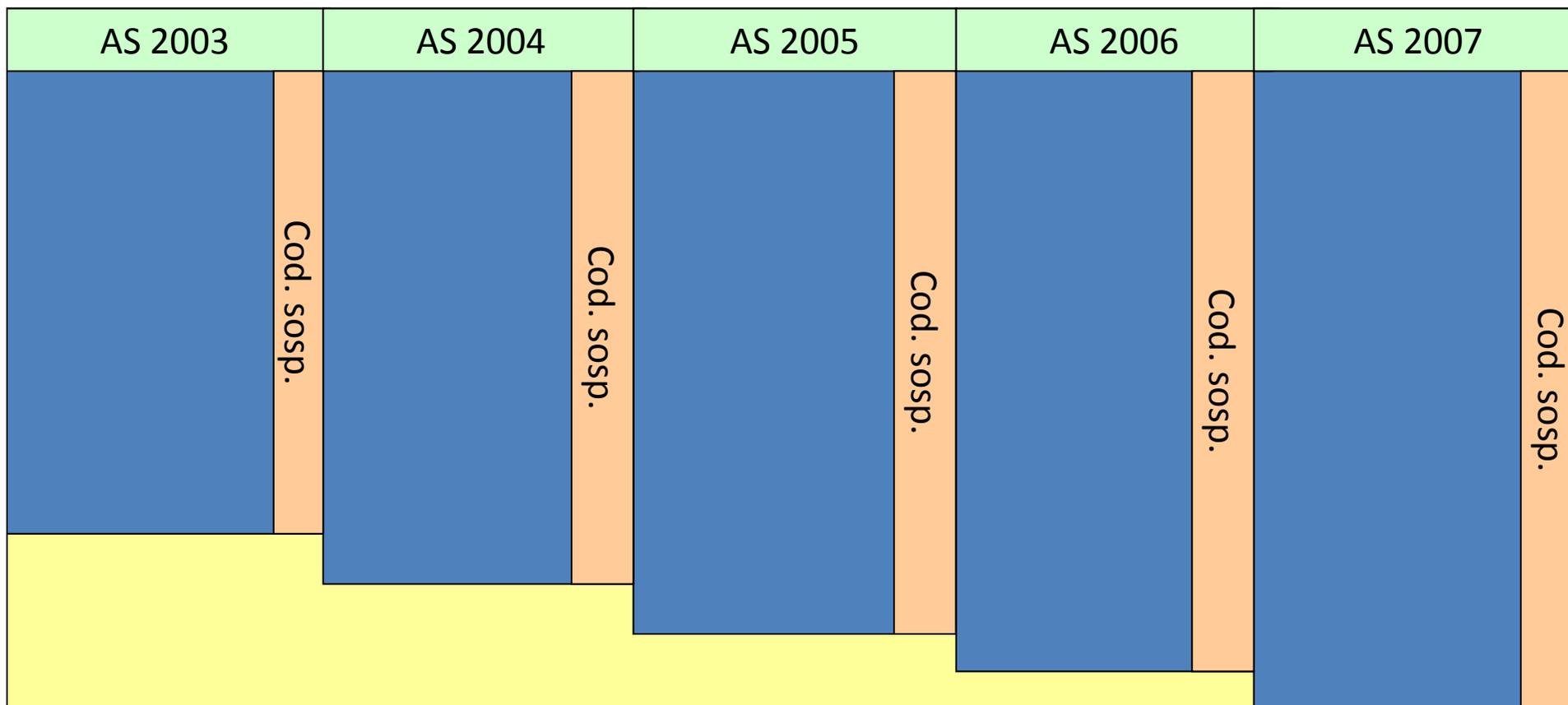
2. SISTEMAZIONE ANAGRAFE SANITARIA

- Completezza dei campi: valutare la possibilità di ricalcolare il Codice Fiscale attraverso un algoritmo di calcolo sulla base dei dati anagrafici disponibili per ogni soggetto, aggiungendo il CF ricalcolato come nuova variabile. Valutare concordanze e discordanza del CF presente nell'anagrafe e di quello ricalcolato.
- Duplicati: verificare che non ci siano soggetti con diversa Tessera Sanitaria (generalmente chiave primaria dell'archivio AS), ma stesso CF o nome, cognome, data di nascita. Dopo aver identificato eventuali soggetti duplicati si manterrà, fra due duplicati, il record con il progressivo di TS più alto (più recente) mantenendo una copia dell'Anagrafe integrale. Nel nuovo dataset verrà stata aggiunta una variabile per tracciare le modifiche effettuate.



2 Record Linkage - CCM

'Storicizzazione AS'





Criticità utilizzo AS

- Non sempre disponibilità 'fotografie' annuali dell'AS
- Poca affidabilità codici di sospensione assistenza sui quali identificare gli 'assistibili'
- Mancanza informazione storicizzata completa data entrata-uscita dall'anagrafe sanitaria



2 Record Linkage - CCM

Completezza dei campi utilizzati per il record linkage – valori percentuali. Rovigo

ANNO	CF	Cognome	Nome	Data di Nascita
1996	0.0	100.0	100.0	100.0
1997	0.0	100.0	100.0	100.0
1998	4.3	100.0	100.0	100.0
1999	95.9	100.0	100.0	100.0
2000	99.4	100.0	100.0	100.0
2001	99.6	100.0	100.0	100.0
2002	99.8	100.0	100.0	100.0
2003	98.7	100.0	100.0	100.0
2004	99.7	100.0	100.0	100.0
2005	100.0	100.0	100.0	100.0
2006	98.8	100.0	100.0	100.0
2007	100.0	100.0	100.0	100.0



2 Record Linkage - CCM

Mortalità

record linkage - percentuali incroci

Fonte	Periodo	Passo di chiave	Appaiati	Appaiati cumulati	% appaiati	% appaiati cumulati
mortalità	1996-2008	K0	28663			
		CFCNDN	14118	14118	49,26	49,3
		CFCDN	203	14321	0,71	50,0
		CFCN	45	14366	0,16	50,1
		CFNDN	207	14573	0,72	50,8
		CNDN	9162	23735	31,96	82,8
		nuovo_cf*	102	23837	0,36	83,2

* incrocio non effettuato per 2008

non appaiati 4826 16,8

Anno	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
% appaiati cumulati	75.3	73	76.2	80.6	79.2	82.4	84	84.2	85.9	86	90.6	91.4	90.8
% appaiati cumul. ASL					92.3	95.6	95	93.7	94.7	95	99.6	99.1	99.4



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

3 produzione di stime di frequenza - CCM

Esempio algoritmo diabete

fonte	codice	modalità
SDO	ICD9CM 250**	in tutti i campi di diagnosi
Farmaceutica	ATC A10**	≥ 2 prescrizioni / anno
Esenzione ticket	013.250	

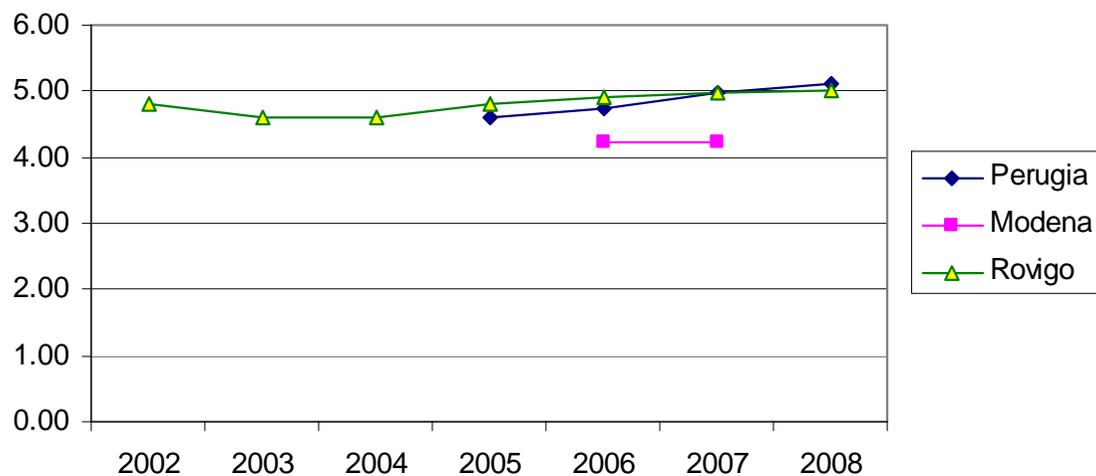
Stima di prevalenza annuale



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

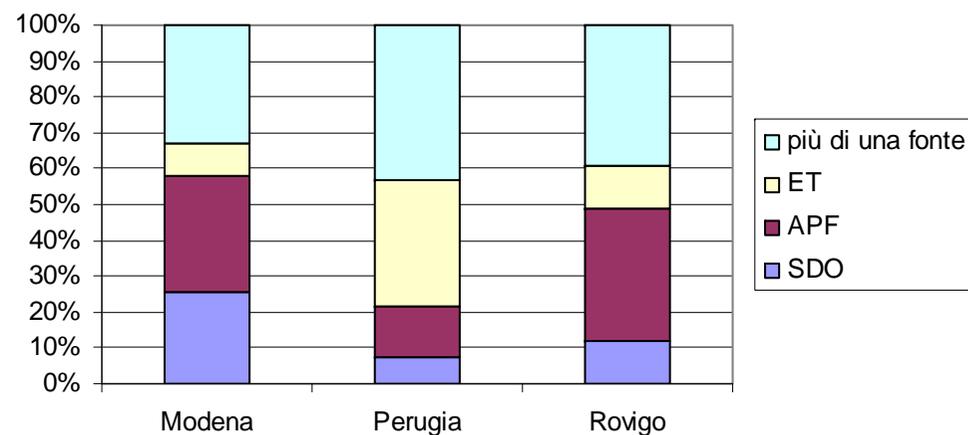
3 produzione di stime di frequenza - CCM

Tassi standardizzati diabete



Gruppo AIE-SISMEC 2002-2004
Range:3.69 (Torino)- 4.18 (Venezia)

Contributi esclusivi fonti, diabete





UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Esempio di utilizzo- validazione stima BPCO attraverso farmaci 2009-2011

COPD JOURNAL OF CHRONIC OBSTRUCTIVE
PULMONARY DISEASE

COPD, 9:1–13, 2012
ISSN: 1541-2555 print / 1541-2563 online
Copyright © Informa Healthcare USA, Inc.
DOI: 10.3109/15412555.2011.654014

informa
healthcare

ORIGINAL RESEARCH

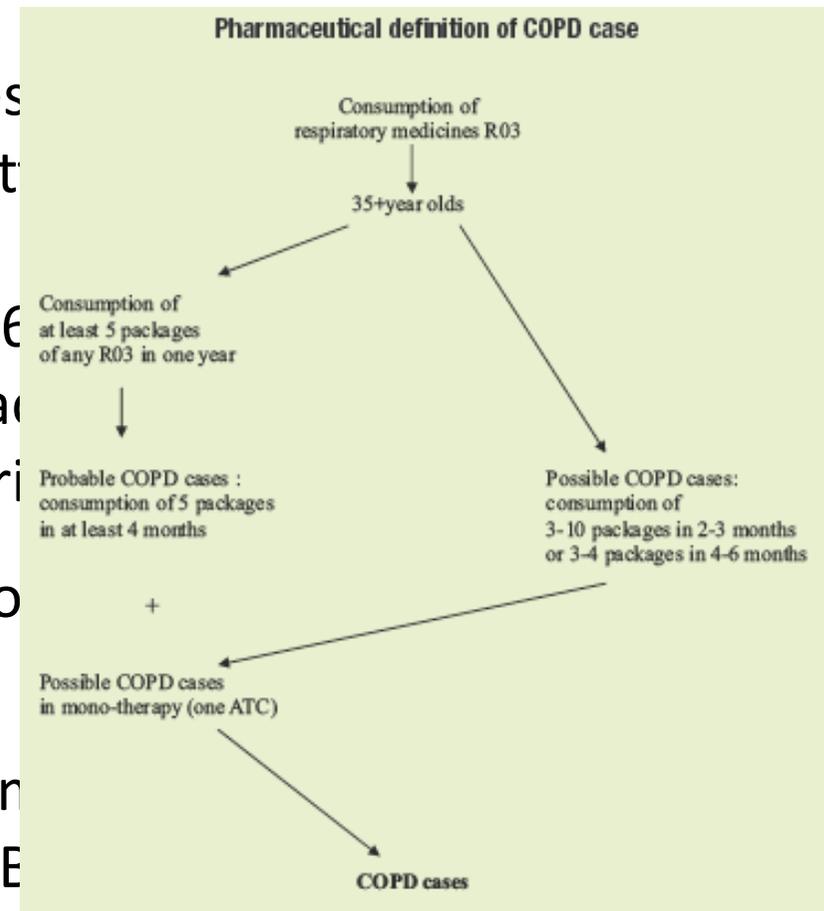
The Reliability of Hospital and Pharmaceutical Data to Assess Prevalent Cases of Chronic Obstructive Pulmonary Disease

Annunziata Faustini,¹ Cristina Canova,² Silvia Cascini,¹ Vincenzo Baldo,³ Karin Bonora,⁴
Gianfranco De Girolamo,⁴ Pierantonio Romor,⁵ Loris Zanier,⁶ and Lorenzo Simonato³

OBIETTIVO: definire il contributo dei dati farmaceutici alla stima di prevalenza della BPCO e per valutare la affidabilità dei registri SDO e CM nell'identificare i casi di BPCO.



- I casi prevalenti di BPCO tra i residenti stimati per il 2006, in quattro amministrazioni, utilizzando i dati di ospedalizzazione dal 2002 al 2006 e successivamente i dati dei farmaci (CINECA) dal 2005 al 2007 (Il algoritmo)
- Per ciascuna area sono stati calcolati i casi di BPCO secondo i due algoritmi.
- La validità interna degli algoritmi è stata valutata confrontando le stime di prevalenza della EPOC nelle aree. I dati farmaceutici sono stati utilizzati per confermare i casi BPCO individuati dai registri SDO e CM e per valutare la possibile selezione e misclassificazione dei casi ospedalieri.





Esempio di utilizzo-metodi 2

Box 1		COPD pharmaceutical cases		
		YES	NO	
COPD hospital or dead cases	YES	a	c	a+c
	NO*	b	??	
		a+b		
<p>a = subject who were discharged or dead with COPD and took medicines following discharge or preceding death, according to pharmaceutical COPD case definition</p> <p>c = COPD hospital or dead cases who did not take medicines</p> <p>b = COPD pharmaceutical cases who had no COPD hospital diagnosis nor died from COPD</p>				
<p>* included those not hospitalised in 2002-2006 or discharged with diagnosis other than COPD, respiratory failure or congestive heart failure.</p>				

Stimata proporzione di casi BPCO con diagnosi ospedaliera (5 anni) o morti nel 2006 tra coloro che erano stati identificati dai farmaci ($a/a+b$) (**sensibilità**) e la proporzione di chi prendeva farmaci tra coloro che erano stati ricoverati o deceduti ($a/a+c$) (VPP) (box 1)



Esempio di utilizzo- risultati 1

Tassi standardizzati per età di BPCO in soggetti residenti, di 35+ anni, per area, algoritmo e genere – 4 aree, 2006

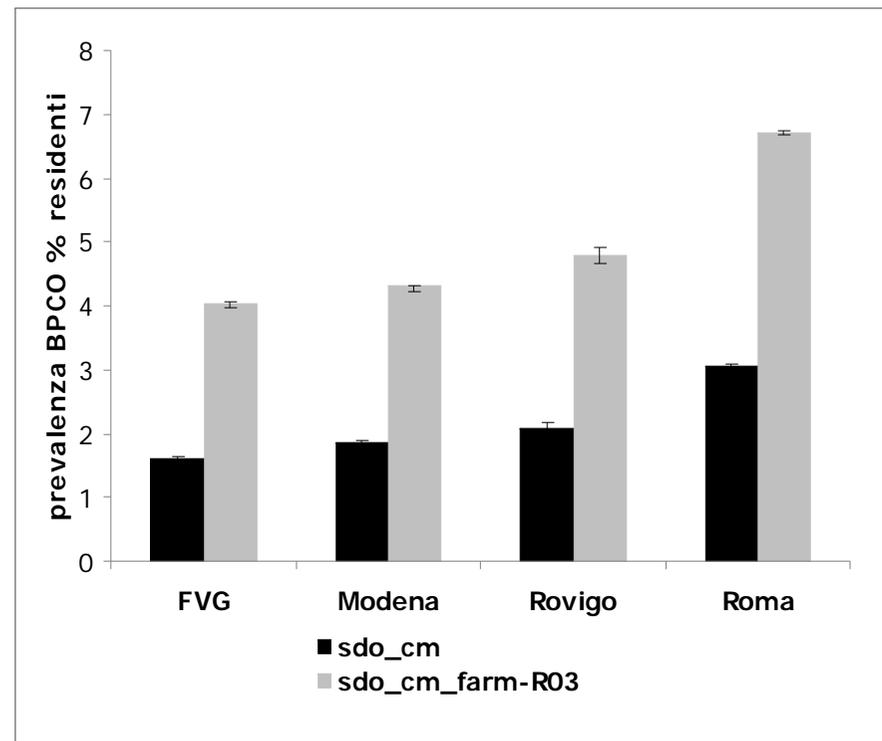




Table 3. Pooled results: % of pharmaceutical COPD cases who were discharged with a COPD or died from COPD, by hospital case definition and age, 2006.

COPD hospital case definition	COPD cases		sensitivity			p-value heterogen ^b	
	all pharm	hosp/dead using drugs	% ^a	95% CI			
	N	N					
Any COPD diagnosis							
age (yrs)	35+	120492	25233	22.7	19.9	25.9	
	35-64	47663	3358	7.2	6.4	8.0	0.832
	65-84	63386	18315	31.0	28.0	34.3	0.425
	85+	9443	3560	40.0	34.5	46.4	0.001
Principal diagnosis of COPD							
age (yrs)	35+	120492	8691	7.3	6.9	7.3	
	35-64	47663	1365	2.6	2.2	3.0	0.220
	65-84	63386	6145	9.7	8.8	10.8	0.545
	85+	9443	1181	13.0	11.0	15.2	0.001

^athe values of proportions are from meta-analysis, they are not immediately derivable from the absolute numbers reported in the table

^b p-value of heterogeneity test across the cities (null hypothesis was perfect homogeneity of city-specific results).



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 ^b	
	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	
	35-64	13104	3358	29.3	23.7	36.3	0.016
	65-84	50414	18315	41.7	34.8	49.9	0.023
	85+	13580	3560	28.5	22.0	36.9	0.001
Principal diagnosis of COPD							
age (yrs)	35+	20465	8691	44.1	38.2	51.0	
	35-64	4371	1365	32.9	25.6	42.2	0.021
	65-84	12708	6145	50.9	45.7	56.7	0.426
	85+	3386	1181	35.1	30.9	39.9	0.001

^athe values of proportions are from meta-analysis, they are not immediately derivable from the absolute numbers reported in the table

^b p-value of heterogeneity test across the cities (null hypothesis was perfect homogeneity of city-specific results).



- L'uso dei dati di ospedalizzazione e mortalità sottostima la prevalenza della BPCO, secondo l'atteso. La probabilità che i casi di BPCO che consumano farmaci respiratori siano stati ospedalizzati nei 5 anni in studio era bassa (23%).
- 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 era bassa (37%).
- Studi differenti richiedono differenti caratteristiche di validità dei dati: stimare la prevalenza richiede maggiore sensibilità, arruolare coorti per studi di follow-up o stimare gli effetti degli inquinanti ambientali richiede maggiore affidabilità.

Grazie per l'attenzione