

AIE, Roma 18 maggio 2007

L'integrazione di fonti informative sanitarie e non sanitarie
per la valutazione delle politiche e per la ricerca

L'uso integrato a livello individuale delle indagini ISTAT con le fonti sanitarie correnti

Nicola Caranci (S. Epidemiologia, ASL5 Torino)

Gabriella Sebastiani, Marilena Pappagallo (ISTAT)

Arduino Verdecchia (ISS)

Lucia Lispi (Ministero Salute)

Patrizia Vittori (Regione Val d'Aosta)

Esperienze italiane di integrazione tra fonti informative ISTAT e Sanitarie

1.

Censimento 1981 → Mortalità 1981- 1982
(6 mesi f.u.)

Censimento 1991 → Mortalità 1991-1992
(12 mesi f.u.)

Basso livello di istruzione VS
laurea: $RR > 2$

% di associazione:
75 circa

2.

Censimento 1991 ←

RT; 1990-1994,
da 6 registri
tumori (Genova,
Parma, Firenze, Macerata,
Ragusa, Sassari)

***" DIFFERENZE DI MORTALITÀ E OSPEDALIZZAZIONE
SECONDO STATO DI SALUTE, STILI DI VITA E
CONSUMO DI SERVIZI SANITARI***

ISTAT SALUTE 2000 " (prog. Ex art. 12)

- ISTAT
- Ministero della Salute
- Val d'Aosta
- Servizio di Epidemiologia ASL 5 Torino

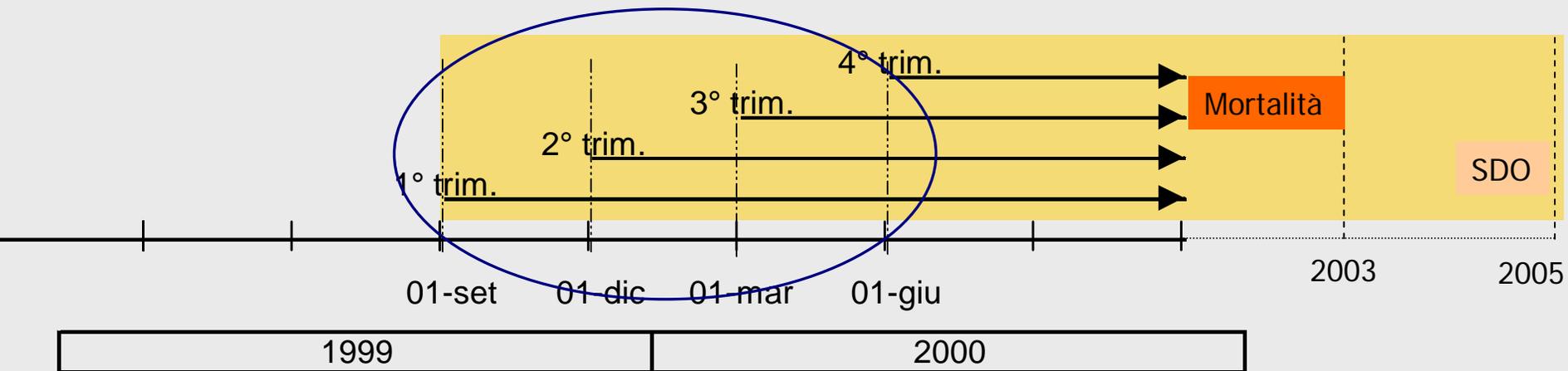
Il disegno dello studio

Follow-up degli intervistati nell'indagine campionaria ISTAT sulle condizioni di salute (edizione 2000):

↓
Record linkage deterministico

con dati correnti di mortalità e ricoveri

Studio di coorte (chiusa)



Metodo ^{1/5}

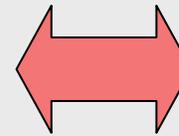
- *Record linkage* deterministico:

ISTAT Salute 2000



Dati nominativi

Ricostruzione del Codice Fiscale **(CF)**



Ricoveri
Piemonte



Mortalità Italia

Metodo ^{2/5}

- *Record linkage* deterministico; preparazione archivio indagine Salute 2000:
 - costruzione chiave di *linkage* per attribuire il CF ai 140.011 intervistati:
trimestre di rilevazione, comune di residenza, codice famiglia, data di nascita, sesso
 - costruzione del CF tramite:
cognome, nome, comune di nascita, data di nascita, sesso

Metodo ^{3/5}

Dati sanitari disponibili da archivi correnti:

- **Mortalità:**

- archivio ISTAT nazionale con cause di decesso (1999-2002)
- archivio ISTAT nazionale deceduti (2003)

CF presente e valido rispettivamente nel 78% e 82% dei decessi

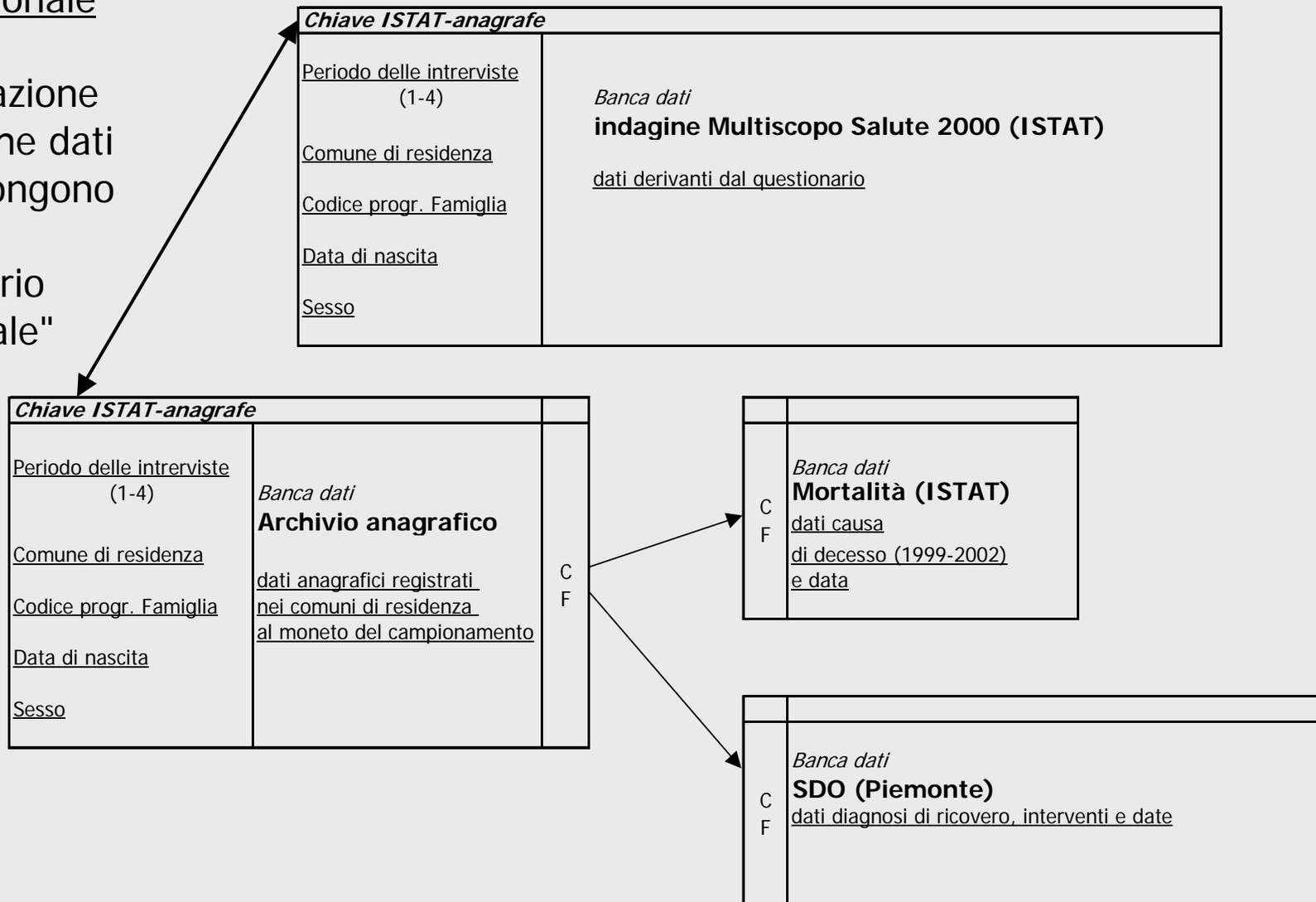
- **Ricoveri:** archivio regionale Piemonte (1999-2005)

CF presente e valido nel 95% delle Schede

NB: studio pilota dell'integrazione a livello nazionale dei ricoveri

Metodo 4/5

Schema del *data base* relazionale generato dall'integrazione delle banche dati che compongono il sistema "campionario longitudinale"



Metodo 5/5

“Sotto-campione” dei soggetti per cui è possibile un’analisi longitudinale degli esiti di salute:

128.967 record con informazioni necessarie per il *linkage* con gli archivi amministrativi di mortalità e ricoveri



Correzione automatica di parti di chiave errate (sesso, nome, cognome, data di nascita, sesso)

NB: il livello di eleggibilità della “coorte teorica” è del 92%

Risultati 1/6

Record linkage dei decessi tramite **CF** (ricostruito nel 20% dei record):

procedura di 29 passi in successione con chiavi con potere discriminante decrescente:

n. chiave	chiave	Pattern	1999-2002		2003	
			Linked	% sul tot. Dei candidati	Linked	% sul tot. Dei candidati
0	pseudo CF completo	ABC XYZ 999 H 01 L219 M	2629	66.36	907	80.12
1	senza sesso	ABC XYZ 999 H 01 L219 -	12	0.30	7	0.62
2	senza comune	ABC XYZ 999 H 01 --- M	372	9.39	45	3.98
3	senza giorno	ABC XYZ 999 H - L219 M	33	0.83	6	0.53
4	senza mese	ABC XYZ 999 - 01 L219 M	25	0.63	10	0.88
5	senza anno	ABC XYZ --- H 01 L219 M	32	0.81	5	0.44
6	senza nome	ABC --- 999 H 01 L219 M	79	1.99	31	2.74
7	senza cognome	--- XYZ 999 H 01 L219 M	46	1.16	14	1.24
8	senza sesso comune	ABC XYZ 999 H 01 ---- -	1	0.03	0	0.00
9	senza sesso giorno	ABC XYZ 999 H - L219 -	3	0.08	1	0.09
10	senza sesso mese	ABC XYZ 999 - 01 L219 -	0	0.00	0	0.00
11	senza sesso anno	ABC XYZ --- H 01 L219 -	10	0.25	1	0.09
12	senza sesso nome	ABC --- 999 H 01 L219 -	32	0.81	7	0.62
13	senza sesso cognome	--- XYZ 999 H 01 L219 -	37	0.93	5	0.44
14	senza comune giorno	ABC XYZ 999 H - ---- M	215	5.43	15	1.33
15	senza comune mese	ABC XYZ 999 - 01 ---- M	2	0.05	1	0.09
16	senza comune anno	ABC XYZ --- H 01 ---- M	239	6.03	11	0.97
17	senza comune nome	ABC --- 999 H 01 ---- M	13	0.33	1	0.09
18	senza comune cognome	--- XYZ 999 H 01 ---- M	0	0.00	0	0.00
19	senza mese giorno	ABC XYZ 999 - - L219 M	9	0.23	1	0.09
20	senza anno giorno	ABC XYZ --- H - L219 M	17	0.43	0	0.00
21	senza nome giorno	ABC --- 999 H - L219 M	1	0.03	0	0.00
22	senza cognome giorno	--- XYZ 999 H - L219 M	4	0.10	0	0.00
23	senza anno mese	ABC XYZ --- - 01 L219 M	94	2.37	41	3.62
24	senza nome mese	ABC --- 999 - 01 L219 M	36	0.91	22	1.94
25	senza cognome mese	--- XYZ 999 - 01 L219 M	1	0.03	0	0.00
26	senza nome anno	ABC --- --- H 01 L219 M	16	0.40	1	0.09
27	senza cognome anno	--- XYZ --- H 01 L219 M	1	0.03	0	0.00
28	senza cognome nome	--- --- 999 H 01 L219 M	3	0.08	0	0.00
Totale			3962	100.0	1132	100.0

Risultati 2/6

Record linkage dei decessi tramite **CF**:

Stima dei decessi attesi per anno di osservazione

Anno	Decessi osservati Italia	Tassi Italia (per 10.000)	Calcolo dei decessi attesi COORTE (A)	Osservati (O)	O/A
1999	556943	97,7			
2000	549721	96,4			
2001	546447	95,9	4089	3962	97%
2002	550185	96,5			
2003	586776	102,9	1378	1132	82%

Risultati 3/6

Le evoluzioni della regolamentazione per la protezione di dati hanno introdotto ostacoli nell'acquisizione dei dati nazionali delle dimissioni; si è usato, in via sperimentale, il flusso informativo delle SDO della Regione Piemonte

Record linkage dei ricoveri tramite **CF** (ricostruito nel 5% dei record):
procedura di 29 passi in successione con chiavi aventi potere discriminante decrescente:
risultato che non supera di molto l'80%; buona capacità della chiave senza comune di nascita (20% di collegamenti)

Risultati 4/6

Stima dei ricoveri attesi per anno di osservazione

Anno	Ricoveri osservati Regione Piemonte	Calcolo dei ricoveri attesi (A)	Attesi cumulati	Osservati "riproporzionati" (O)	O/A
1999	934331	416	416	307	74%
2000	897246	1297	1713	1007	78%
2001	883896	1573	3286	1251	80%
2002	859409	1529	4815	1250	82%
2003	852400	1517	6331	1197	79%
2004	870430	1549	7880	1279	83%
2005	865625	1540	9421	1252	81%

Risultati 5/6

- La banca dati degli esiti sanitari individuati all'interno della coorte è stata collegata dall'ISTAT ai dati dell'indagine Salute 2000
- Gli esiti sanitari si riferiscono alla "sotto-coorte" di 128.967 intervistati con le informazioni anagrafiche necessarie per il CF
- L'ISTAT ha reso disponibile ai soggetti partecipanti alla convenzione * i dati dell'indagine e i relativi eventuali esiti sanitari, rendendo anonimi i *record*

* CONVENZIONE RELATIVA ALLA REALIZZAZIONE DI UNA BASE DATI LONGITUDINALE NELL'AMBITO DEL PROGETTO "DIFFERENZE DI MORTALITA' E DI OSPEDALIZZAZIONE (2000-2004) SECONDO LO STATO DI SALUTE, GLI STILI DI VITA E IL CONSUMO DI SERVIZI SANITARI NEL CAMPIONE DELL'INDAGINE ISTAT SULLA SALUTE 1999-2000" tra ISTAT, Ministero Salute e Regione Valle d'Aosta,

Risultati 6/6

- Consistenza finale complessiva dei dati sanitari nella coorte osservata:
 - 5.094 deceduti tra tutti i soggetti intervistati; 4.666 tra i soggetti appartenenti alla "sotto-coorte"
 - 7.543 ricoveri; 6.936 ricoveri nella "sotto-coorte" seguita per il Piemonte dopo l'intervista
- I differenziali di mortalità e morbosità mostrano risultati coerenti con quanto atteso sulla base della letteratura e delle ipotesi;
cfr.: *follow-up* censimento 1991; Cornella (Spagna, 2005)

Validità ^{1/2}

- Incompleta attribuzione CF al campione degli intervistati in Salute 2000 (cfr. seguito)
- Bassa percentuale di decessi osservati VS attesi nell'anno 2003
- Bassa percentuale di associazione dei dati di ricoveri:
mancano istituzionalizzati

Validità 2/2

- 128.967 soggetti della "sotto-coorte" VS rimanenti 11.044:
 - non differiscono per distribuzione di sesso
 - età lievemente superiore (40,8 VS 39,4 anni)
 - frequenza di disoccupati inferiore di 1 pto. %
 - minore frequenza di alto Indice di Deprivazione
- L'eventuale *bias* introdotto dovrebbe andare nella direzione "conservativa" rispetto ai differenziali di stato socio-economico

Altri esempi in letteratura internazionale

- *The Cornella* (Barcellona, Spagna) *Health Interview Survey follow-up* (AEP 2005):
survey di 2.500 individui (campionati tra 85.000 abitanti)
intervista eseguita nell'anno 1994, 8 anni di f.u.
 - i soggetti con una peggiore percezione del loro stato di salute hanno una probabilità di morire **2,6** volte rispetto a coloro che si sentivano in buona salute
- *National Health and Nutrition Examination Survey* (NHNES follow up, US 1971-1987): malattie coronariche in un campione di 5.800 individui
- Studio MONICA, EPIC

Valore d'uso

Esempi di quesiti a cui è possibile dare risposta (e relative politiche da valutare):

- intensità e direzione delle diseguaglianze sociali nella mortalità a livello nazionale (es.: revisione di età e coefficienti previdenziali)
- predittività dei rischi di morte degli stili di vita (es.: carte del rischio cardiovascolare italiane)
- differenze di mortalità e morbosità per indicatori di qualità ambientale aggiustati per posizione sociale e stili di vita (es.: qualità dell'aria)
- impatto sulla mortalità di interventi di prevenzione (es.: screening spontaneo VS screening programmato)

Grazie per l'attenzione

nicola.caranci@epi.piemonte.it