Le tecniche di record-linkage

Nicola Caranci, Valeria Fano

ISS 3-5 Aprile 2013

Gli anni 2000 e l'informazione statistica

Nell'ultimo quarto di secolo, l'evoluzione **dell'epidemiologia** ha avuto a che fare con:

- grandi moli di **dati crescenti**, registrati per fini amministrativi o statistici, utili anche per **fini epidemiologici**

- con l'evoluzione delle **esigenze di studio** e **osservazione** empirica, evolvono i **quesiti di ricerca**

- questo è il **"secolo dell'integrazione"**; dalla fine degli anni '80 sempre **+ dati** dalla Statistica ufficiale e dalla Sanità e **> potenza di calcolo**

- se le storie di vita sono **bibliografie**, il ***record linkage*** equivale al lavoro certosino di rilegatura delle pagine

Laboratorio dei Sistemi di Babele

Basi dati, record, campi

Esempio fittizio di archivio (**base dati**) anagrafico :

Cognome	Nome	Data_nascita	Com__nas	Com_res	Indir_residenza	Sez_cr	Codice Fiscale
Verde	Aldo	27/01/1967	037006	037006	Via Leopardi, 75	023	VRD LDA 67A27 A944Y
Bianco	Franco	12/04/1973	027019	037012	P.ZZA AZZARITA, 3	001	BNC FNC 73D12 C388 K
Rosso	Maria	02/08/1953	001179	037006	Via Libia, 34	034	RSS MRA 53M42 E379 R
Grigio	Olindo	22/09/1945	037006	037006	Via della Salita, 777	102	GRG LND 45P22 A944 J
Giallo	Tina	28/02/1957	036008	037006	Via XXIV Maggio	075	GLL TNI 45B68 B249 R

Records

campi: caratteristiche rilevate sulle **unità statistiche**, registrate nei loro ***record***, seguendo un determinato **tracciato**

Basi dati, record, campi

… es. di *linkage* con archivio redditi dei residenti**:**

Cognome	Nome	Data_nascita	Codice Fiscale
Verde	Aldo	27/01/1967	VRD LDA 67A27 A944Y
Bianco	Franco	12/04/1973	BNC FNC 73D12 C388 K
Rosso	Maria	02/08/1953	RSS MRA 53M42 E379 R
Grigio	Olindo	22/09/1945	GRG LND 45P22 A944 J
Giallo	Tina	28/02/1957	GLL TNI 45B68 B249 R

omocodia

Grigio Olindo
Gregorio Aleandro
01/07/1953

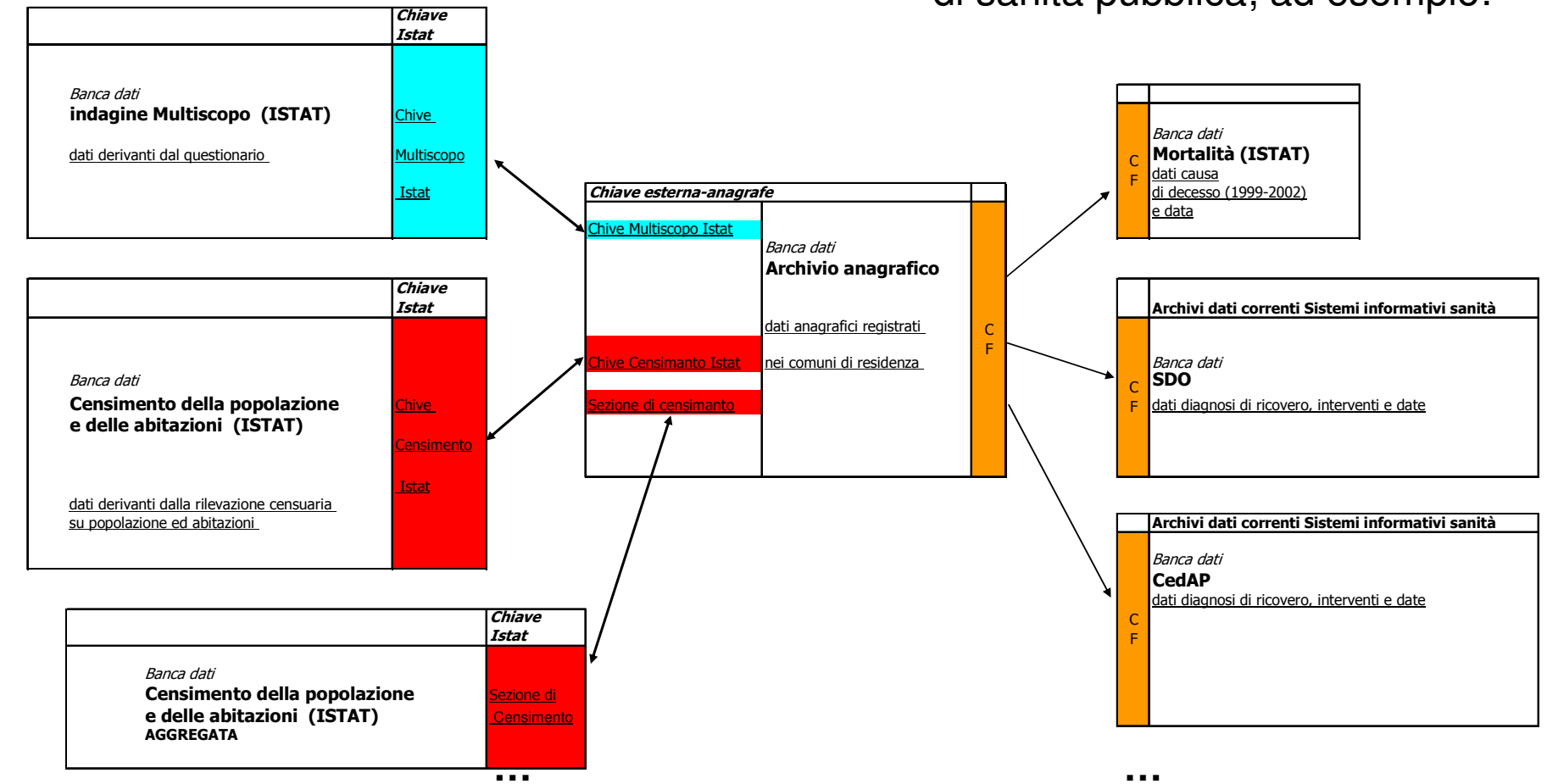
29/02/1957

CF_c	Reddito_2011	N_famil
CTC RFV 58L22 E224 Y	28.760	2
GRG LND 45P22 A944 J	18.700	3
GRG LND 45P22 A944 J	49.200	2
RSS MRA 53L41 E379 R	33.080	5
VRD LDA 67A27 A944Y	37.500	4
GLL TNI 45B-- B249 R	29.450	4

codice incompleto per errore

Concetto “chiavi”

Ricercare tratti di biografie per fini di studio epidemiologico e sorveglianza clinica / di sanità pubblica; ad esempio:



Definizione "storica" di *RL*

"Each person in the word creates a Book of Life. Its pages are made up of the records of principal events. **Record linkage** is the name given to the process of assembling the pages of this book…

It is necessary at times to link the various important records of a person's life."

H.L. Dunn, AJPH 1946

Contenuti

***Record Linkage* (*RL*) (semi-)deterministico**; aspetti teorici e applicazioni in Sanità

- cosa si intende per *RL* deterministico

Qualità dei dati e affidabilità del processo di ***RL***:

- **quali** e quanti **archivi** da collegare:
  - **a.** dai flussi di dati correnti in sanità a quelli della statistica ufficiale (con record di individui)
  - **b.** fino a quelli aggregati (dati di contesto e georeferenziazione)

- **quali livelli di affidabilità** si possono raggiungere in base alla bontà delle chiavi:
  1. chiavi già definite per tutti gli archivi da collegare (es: prog_paz)
  2. chiavi complete da anagrafe con verifica (es.: CF da SOGEI)
  3. chiavi incomplete da anagrafe (es.: nome, cognome e altri dati anagrafici)
  4. chiavi molto incomplete e già parzialmente "non identificative" (sesso, data di nascita e comune)

- l'importanza della qualità delle chiavi; **aspetti pratici**

Tipi di *Record Linkage*

Le **tecniche** principali **di RL** sono:

- **deterministica**: utilizzano l'**accordo esatto** dell'insieme **delle caratteristiche** (**campi**) che costituiscono la **chiave identificativa** di un individuo
- **procedure semi-deterministiche** (o *stepwise*); **sequenza di passi** in cui la **concordanza** è valutata su **sottoinsiemi di campi identificativi**

- **probabilistiche**: nessun accordo o disaccordo singolo tra i campi identificativi è sufficiente per stabilire l'appaiamento, o il non appaiamento, di due campi [ci si basa sulla capacità discriminante e sull'attendibilità dei singoli campi identificativi]

Laboratorio dei Sistemi di Babele

Fornari, 2008

E&P, 2011

Pro e contro delle tecniche di ***RL***

Tecniche non probabilistiche:

- **deterministica**: basandosi sull'accordo esatto della chiave (solitamente dati anagrafici o CF), generalmente anonimizzata, ha una **limitata capacità di riconoscere un appaiamento** in condizioni di incertezza

- le **procedure semi-deterministiche** superano in parte tale limite, usando per l'appaiamento chiavi ridotte, sottraendo campi o parti di essi

  - qualunque sia la tecnica usata, non si possono escludere **errori di identificazione** e **appaiamento** che implichino una distorsione dei risultati dello studio

Fornari, 2008

E&P, 2011

Quali archivi da collegare

- Le esperienze di realizzazione di **sistemi informativi** molto articolati e complessi in alcune realtà **regionali**, dove è possibile correlare archivi di:

- **dimissioni ospedaliere**

- **prescrizione di farmaci**

- **mortalità**

hanno offerto e **possono continuare** ad offrire interessanti opportunità nella realizzazione di grandi studi osservazionali.

Raschetti, 2003

… aggiornamento non esaustivo, con esperienze di *RL* (semi-)deterministico

Quali archivi di dati correnti sanitari...
L'esempio del Sistema infomativo (SEI) **veneto**

- Certificati di Morte (1987-)
- Referti AnatomoPatologici (1981-)
- Schede di Dimissione Ospedaliera (1982-)
- Archivio mobilitàpassiva
- Ricoveri in regime di Day-Hospital (1998-)
- Archivio di consumo Farmaci (1998-)
- Esenzioni Ticket (1998-)

AIE, 2007. http://www.epidemiologia.it/sites/www.epidemiologia.it/files/R.Tessari.pdf

Laboratorio dei Sistemi di Babele

Tessari, 2007

Quali archivi di dati correnti sanitari...
L'esempio del Sistema infomativo (SISEPS) **Emilia-Romagna**

Sistema di accoglienza regionale
portale web

Saluter
il portale del Servizio sanitario regionale dell'Emilia-Romagna

SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA

Il Sistema Informativo Politiche per la Salute e Politiche Sociali

Giovedì 14 giugno 2012

Un sistema informativo che sappia garantire flussi di informazioni validate ed aggiornate rappresenta una risorsa indispensabile per la programmazione e la verifica del Servizio Sanitario Regionale nel suo complesso. È gestito e coordinato dal Servizio Sistema Informativo Sanità e Politiche Sociali.

Area Sanità
- Assistenza Farmaceutica - AFT-AFO-FED
- Assistenza Specialistica Ambulatoriale - ASA
- Certificato di Assistenza al Parto - CedAP
- Cure Primarie - PRIM
- Cure Termali - CT
- Dispositivi Medici - DiMe
- Emergenza Urgenza - PS-118
- Hospice - SDHS
- Laboratori - LAB
- Rilevazione Mortalità - ReM
- Salute Mentale e Dipendenze Patologiche - SISM-SINPIAER-SIDER
- Schede di Dimissione Ospedaliera - SDO
- Screening Colon-Retto - SCR
- Sistema Informativo Consultori - SICO
- Strutture dell'offerta ospedaliera - Posti Letto

Area Politiche Sociali e Integrazione Socio-Sanitaria
- Assegno di Cura Anziani e Disabili - SMAC
- Assistenza Domiciliare Integrata - ADI
- Assistenza Residenziale e Semiresidenziale Anziani - FAR
- Gravissime Disabilità Acquisite - GRAD
- Integrazione Applicativi Sportello Sociale - IASS

Area Economico Finanziaria
- Gestione Costi - COA01

Mobilità
- Mobilità Interregionale
- Mobilità Infraregionale
- *Mobilità Internazionale*

Le Anagrafi
- Anagrafe Strutture Sanitarie, Socio-Sanitarie, Sociali autorizzate e accreditate
- Anagrafe Medici Prescrittori A.R.M.P.
- Anagrafe Assistiti N.A.A.R.

Applicazioni
- AIDA Web
- Metadati
- Medicina di Base
- Reportistica URP
- Rilevazione emergenza calore e influenza A H1N1
- Rilevazione Interruzione di Gravidanza
- Sorveglianza Malattie Infettive

Link
- Ministero della Salute
- Agenzia Nazionale per i Servizi Sanitari Regionali
- Istituto Superiore di Sanità
- Agenzia Sanitaria e Sociale Regionale
- Emilia-Romagna Sociale
- La regione in cifre. Statistica self-service
- Altri siti di interesse

https://siseps.regione.emilia-romagna.it/flussi

Laboratorio dei Sistemi di Babele

a.1. Chiavi *passe-partout* degli archivi da collegare

IDENTIFICATIVO PERSONALE ANONIMO, **EMILIA-ROMAGNA - SISEPS**

- Seguendo la L. 196/2003*, si è introdotto negli archivi contenenti dati sensibili un identificativo personale numerico anonimo (**PROG_PAZ**), in sostituzione dei dati anagrafici. E' un **identificativo personale anonimo, comune a tutte le banche dati**

(NB: nei flussi SDO e Hospice il nuovo identificativo sostituisce quello precedente, introducendo un aumento dei ricoveri ripetuti valutato mediamente inferiore allo 0,5%)

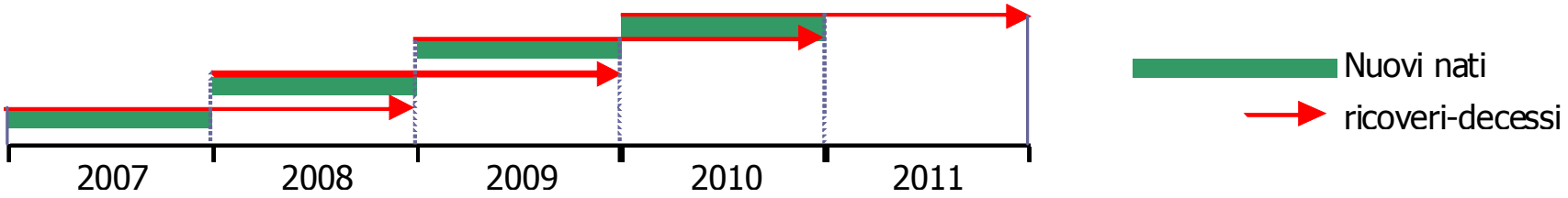
Per coloro che possono accedere ai dati di dettaglio, è possibile ricostruire ed analizzare i percorsi assistenziali nel tempo, in tutto rispetto delle normative vigenti

* Tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali

Laboratorio dei Sistemi di Babele

Esempio di studio possibile con SISEPS 1/2

- Banca dati dei **CedAP**, anno **2007-2010** ($N_{semplici}$=161.571) contiene per ogni nascita:
  informazioni sanitarie e delle
  condizioni socio-demografiche (CSD) della madre
- **ricoveri** nel primo anno di vita **2007-2011** (**SDO**)
- **decessi** nel primo anno di vita **2007-2011** (**REM**)



- Disegno: **coorte di nati vivi**, chiusa e **"seguita"** per un anno **tramite l'archivio SDO** e **REM**

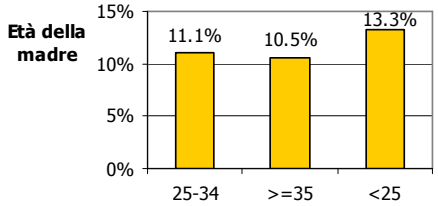
→ $N_{semplici}$ collegati all'archivio SDO=**158.458** (**98,1%***);
- ricoveri di nascita (neonatologia o altro r.): 18.113 (11,4%)
- ricoveri successivi (dopo il 2° giorno di vita): **26.026**

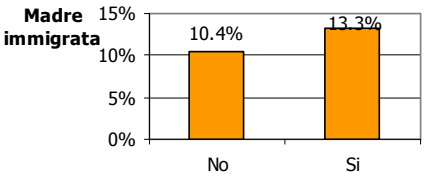
* *linkage* tramite 'numero di riferimento SDO neonato' e altre chiavi applicate *ad hoc*

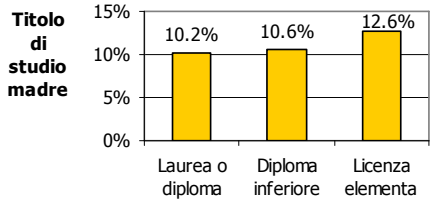
http://www.regione.emilia-romagna.it/sas/cedap/iniziative/2012_11/Caranci.pdf

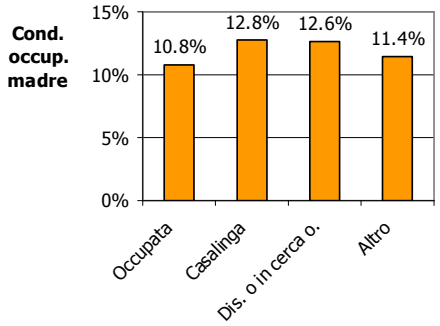
Esempio di studio possibile con SISEPS 2/2

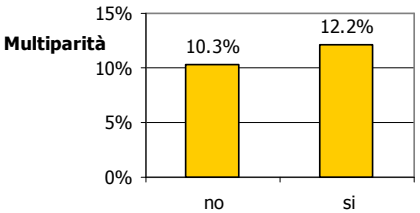
Rischio di **ricovero nel 1° anno dei sani**



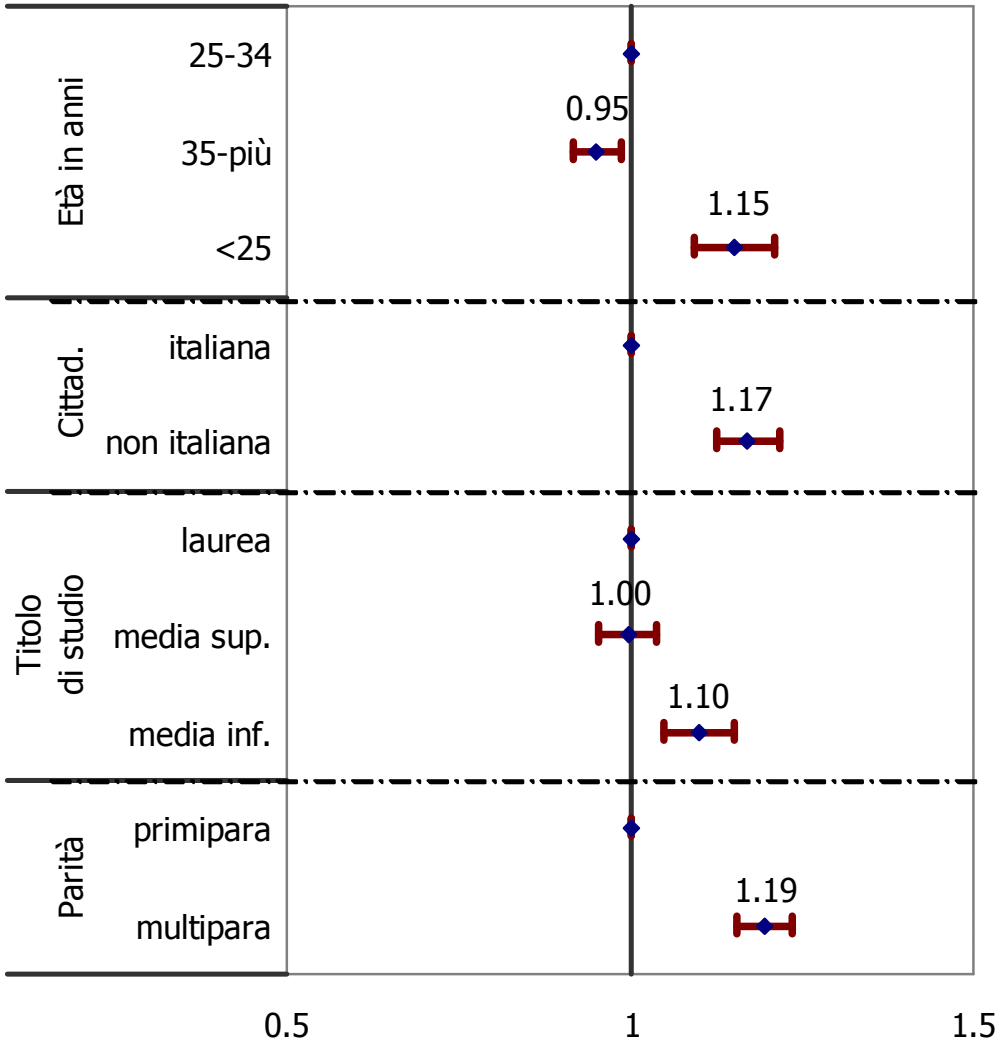








IRR e intervalli di confidenza al 95% (**IC95%**)



IRR N= 136.377

Possibilità di errore nella stima delle prevalenze anche nell'uso di archivi con chiave pre-costituita **1/2**

I **problemi** che si possono incontrare nell'uso per fini epidemiologici degli archivi amministrativi sono:

- **copertura** della popolazione *target* e possibili differenze nei non inclusi (es.: il CedAP è incentrato sull'offerta ed esclude dalla rilevazione le nascite da donne residenti avvenute fuori regione)
- **qualità** dei dati: se i dati non hanno impatto sul processo di gestionale, la qualità può non essere garantita

NB: *spesso è utile collegare diversi archivi informatizzati attraverso la definizioni di chiavi identificative ricavate da sottoinsiemi, più o meno ampi, dei dati anagrafici disponibili*.

Raschetti, 2003

***Sacerdote*, 2003**

Laboratorio dei Sistemi di Babele

Possibilità di errore nella stima delle prevalenze anche nell'uso di archivi con chiave pre-costituita 2/2

- *Se genero una chiave di link con le informazioni anagrafiche che compongono il Codice Fiscale (CF) posso provocare un errore dell'ordine di pochi punti percentuali, ma…*

- … se intendo usare i dati identificativi per una **stima di prevalenza**, **l'errore** nella stima **non è dello stesso ordine** di grandezza **dell'errore presente nell'identificativo** → dipende dalla frequenza di errore nell'identificativo e dal numero di record da registrare per il soggetto;
errori del 5% possono produrre una sovrastima superiore al 100% (per numero prestazioni medio>20).

- **..** il diffondersi della **registrazione elettronica riduce** considerevolmente la **probabilità di errore**;
se invece non si evitano errori di registrazione → scelte di elaborazione (esclusioni)

Laboratorio dei Sistemi di Babele

***Sacerdote*, 2003**

Cislaghi, 2012

aspetti pratici

ci sono alcuni aspetti pratici che di per sé costituiscono dei vincoli nella scelta degli algoritmi di record linkage:

disponibilità di dati completi

qualità dei dati

Laboratorio dei Sistemi di Babele

disponibilità di dati completi

esempio 1

nel Lazio il file delle prescrizioni farmaceutiche contiene il CF ma non i dati anagrafici dei soggetti

nell'integrare più fonti è stato necessario utilizzare le stesse chiavi (in questo caso tutte funzioni del codice fiscale) per garantire a tutte le fonti la stessa probabilità di riuscita del linkage

Laboratorio dei Sistemi di Babele

disponibilità di dati completi

esempio 2

il Registro di mortalità del Lazio non ha il codice fiscale → è necessario utilizzare quello ricalcolato, cosa non possibile per tutto il data set per mancanza (es. luogo di nascita) e/o per inaccuratezza di informazioni (nomi/cognomi errati)

per aumentare la probabilità di trovare i deceduti, oltre ad utilizzare il CF quando completo, si usano anche chiavi basate sulla normalizzazione del nome+cognome (eliminando spazi e caratteri speciali) e unendo la data di nascita

Laboratorio dei Sistemi di Babele

qualità dei dati

struttura del codice fiscale:

cognome	nome	data nasc	luogo nascita	ultima cifra
3	3	5	4	1

FANO	VALERIA	07/02/1967	ROMA	ultima cifra
FNA	VLR	67B47	H501	C

cosa succede se c'è un errore di trascrizione?

FANO	VALERIA	07/02/1961	ROMA	ultima cifra
FNA	VLR	61B47	H501	?

FANO	VALERIA	07/02/1961	ROMA	ultima cifra
FNA	VLR	61B47	H501	W

CORRETTO	FNAVLR67B47H501C
ERRATO	FNAVLR61B47H501W

Laboratorio dei Sistemi di Babele

cosa succede se c'è un errore di trascrizione?

basta una cifra errata per rendere il record non individuabile come riferito alla persona

CORRETTO	FNAVLR67B47H501C
ERRATO	FNAVLR61B47H501W

se cambia una cifra tra le prime 15 → cambia anche l'ultima

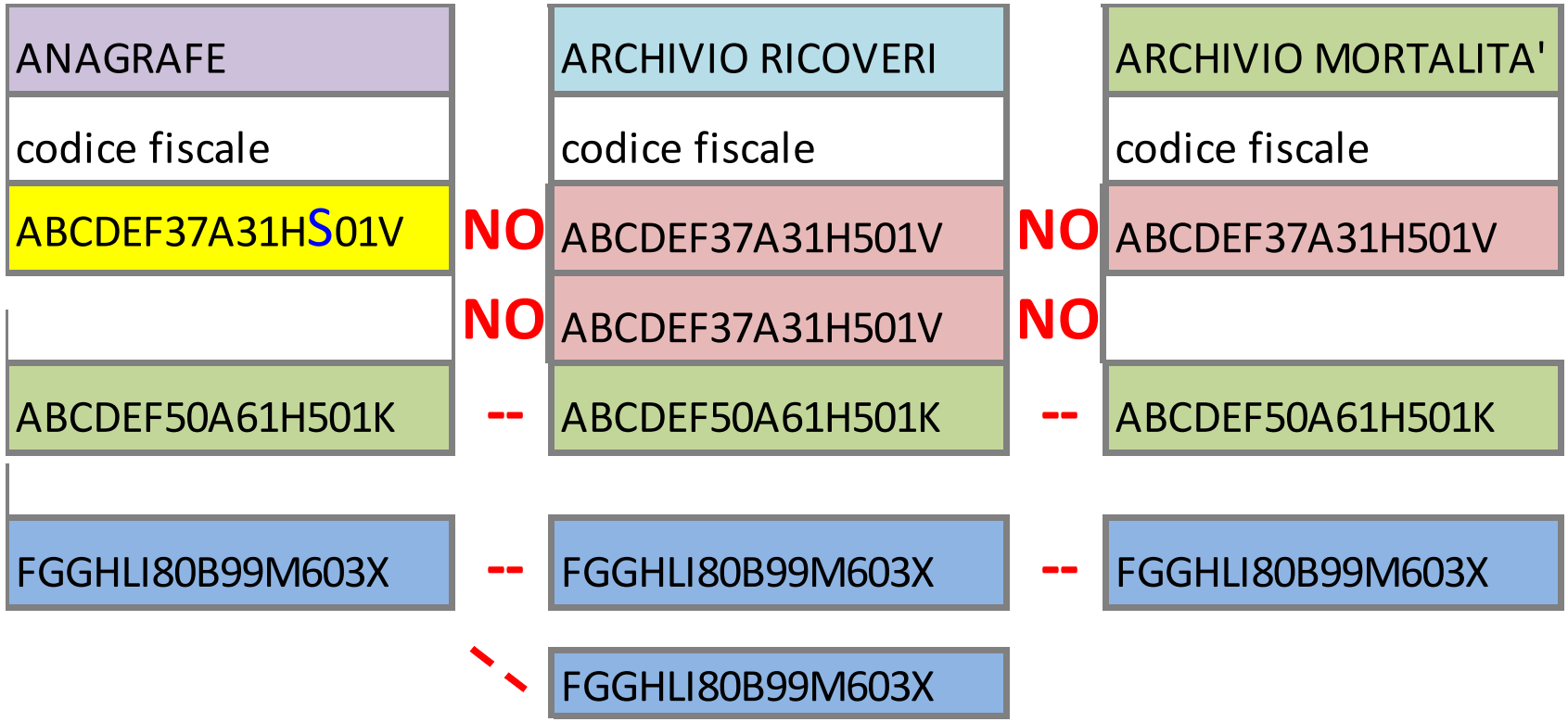
Laboratorio dei Sistemi di Babele

esempio di linkage con l'archivio di mortalità:

ANAGRAFE		ARCHIVIO MORTALITA'
codice fiscale		codice fiscale
ABCDEF37A31H501V	--	ABCDEF37A31H501V
ABCDEF50A61H501K	--	ABCDEF50A61H501K
FGGHLI80B99M603X	--	FGGHLI80B99M603X

ANAGRAFE		ARCHIVIO MORTALITA'
codice fiscale		codice fiscale
ABCDEF37A31HS01V	NO	ABCDEF37A31H501V
ABCDEF50A61H501K	--	ABCDEF50A61H501K
FGGHLI80B99M603X	--	FGGHLI80B99M603X

esempio di linkage con più archivi:



codici fiscali non corretti in tutti gli archivi

CONSEGUENZE

- epidemiologiche
  - mancate chiamate allo screening
  - mancato raggiungimento obiettivi per le campagne di vaccinazione regionali
  - difficoltà nei follow-up
  - datawarehouse

Laboratorio dei Sistemi di Babele

CONSEGUENZE

- **economiche**

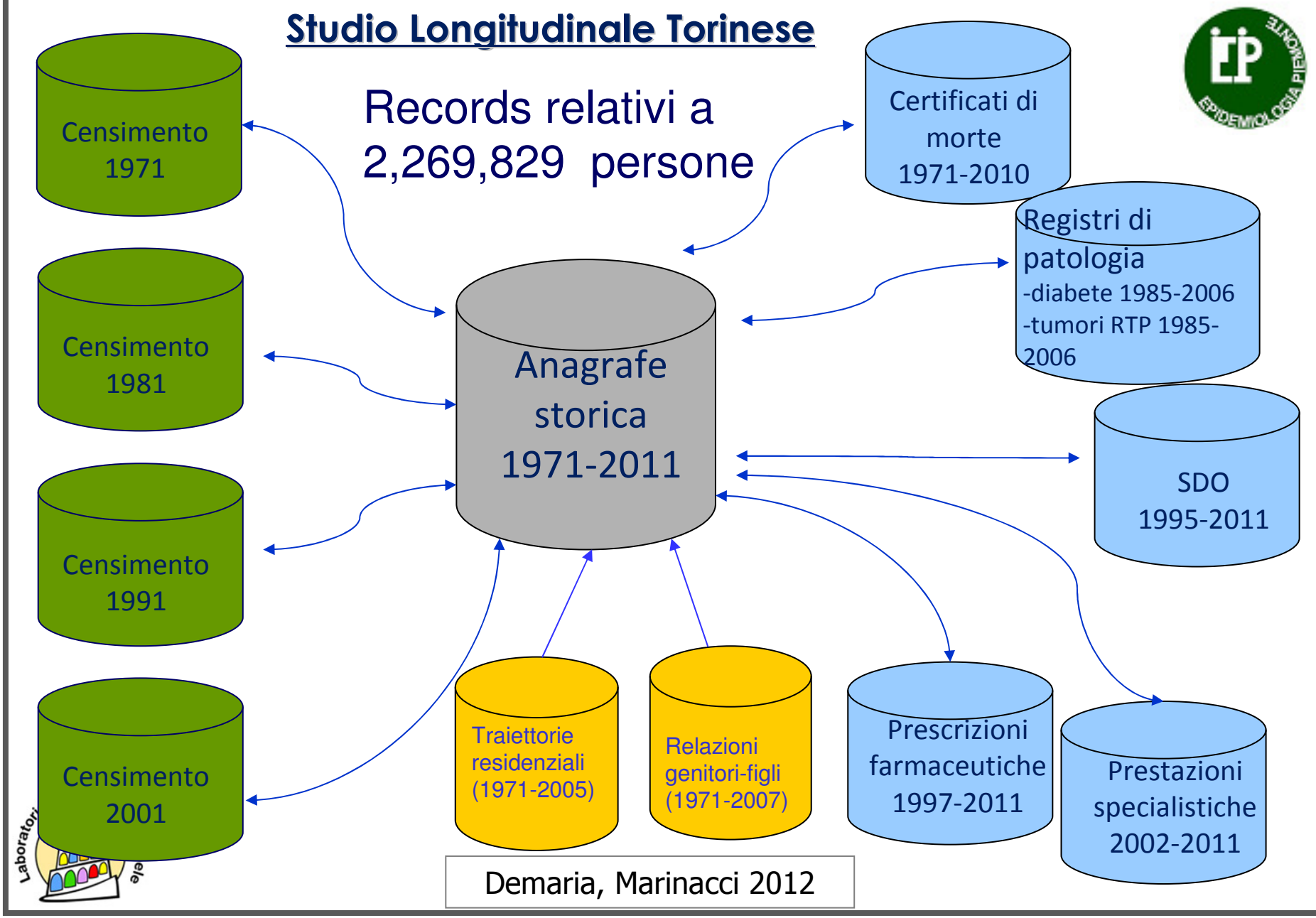
es. assistito deceduto nel 1994 e cancellato solo nel 2010 → costa in media 75€ l'anno

1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
75 €	75 €	75 €	75 €	75 €	75 €	75 €	75 €	75 €	75 €	75 €	75 €	75 €	75 €	75 €	75 €

totale = 1200€

nella Roma D sommando tutte le cancellazioni segnalate negli ultimi 4 anni→ spesa cumulata (evitabile) di oltre 600.000€

a.2-3. Chiavi di diversa natura per gli archivi da collegare



a.2-3. un esempio da fonte campionaria

"***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
- Sevizio di Epidemiologia ASL 5 Torino

Il disegno dello studio

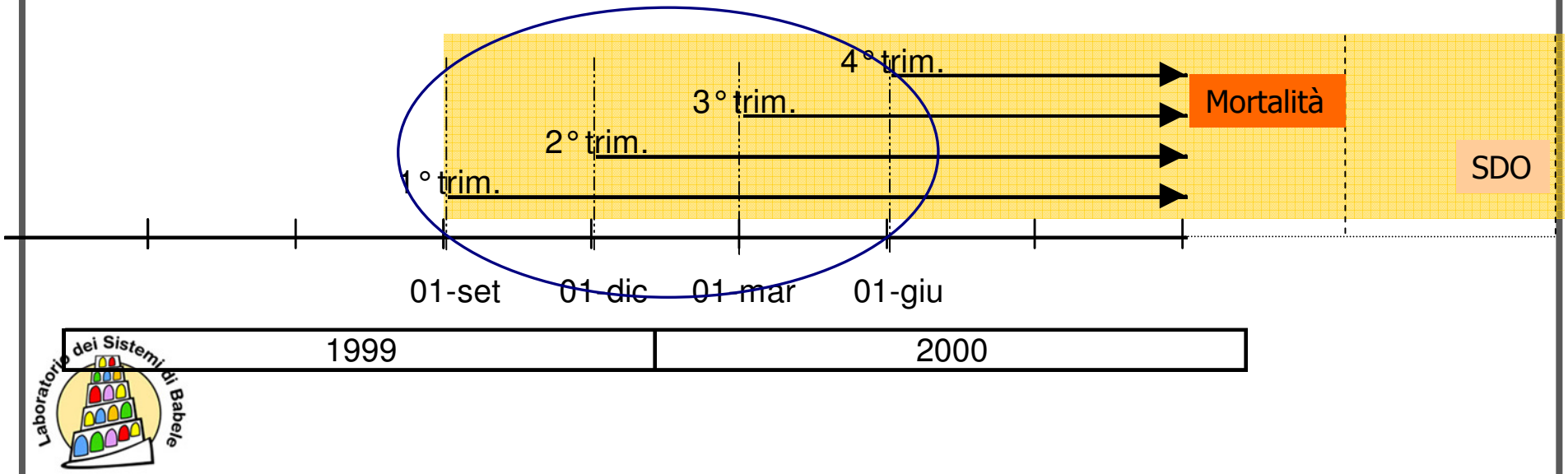
Follow-up degli intervistati nell'indagine campionaria ISTAT sulle condizioni di salute (edizione 2000):

↓

Record linkage semi-deterministico

con dati correnti di mortalità e ricoveri

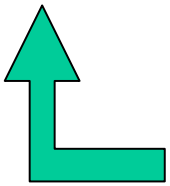
Studio di coorte (chiusa)



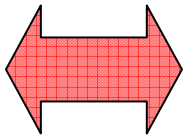
Metodo 1/2

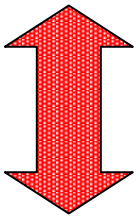
- *Record linkage* deterministico:

ISTAT Salute 2000



Dati nominativi

Ricostruzione del Codice Fiscale **(CF)**  Ricoveri

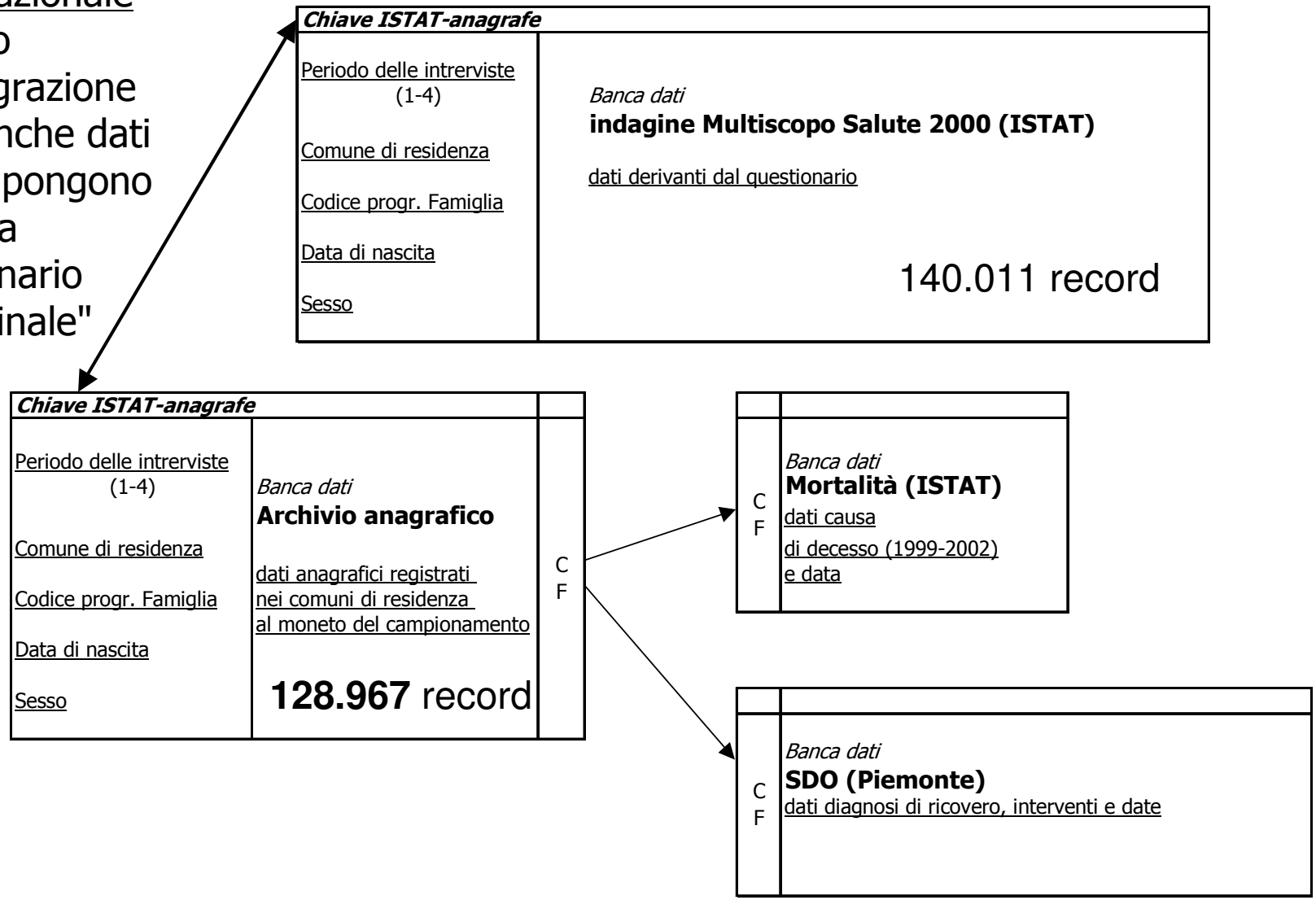


Mortalità

Laboratorio dei Sistemi di Babele

Metodo 2/2

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



Risultati 1/2

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

procedura di 29 passi in successione con chiavi di potere discriminante decrescente (procedura Demaria M.):

			1999-2002		2003	
n. key	chiave	Pattern	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

Laboratorio dei Sistemi di Babele

Risultati 2/2

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%

AIE, 2007:
http://www.epidemiologia.it/sites/www.epidemiologia.it/files/N.Caranci.pdf, si veda anche:
http://www.epidemiologia.it/sites/www.epidemiologia.it/files/A.Bena_P.Crosignani_M.Giraudo_R.Leombruni.pdf

a.3. chiavi da anagrafe (e verifica SOGEI)
Es1: analisi redditi dichiarati e ospedalizzazione

- I dati di reddito sono gestiti dalla SOGEI, che li archivia per il MEF
- All'interno di un progetto min. ex art. 12* è stato possibile ricavare misure aggregate del reddito per quattro città italiane, come dichiarato nell'anno 1998
- Le informazione del reddito sono state studiate in relazione all'ospedalizzazione generale e per particolari trattamenti

* **Diseguaglianze socio economiche di accesso e di trattamento**

Laboratorio dei Sistemi di Babele

Indicatore di reddito

Anagrafi ⟷ Record Linkage ⟷ **Registro dichiarazione redditi**

Popolazione residente all'1/1/'98 ⇨ redditi dichiarati nel 1998

Reddito familiare disponibile

⇩

Reddito pro capite disponibile equivalente (scala Carbonaro)

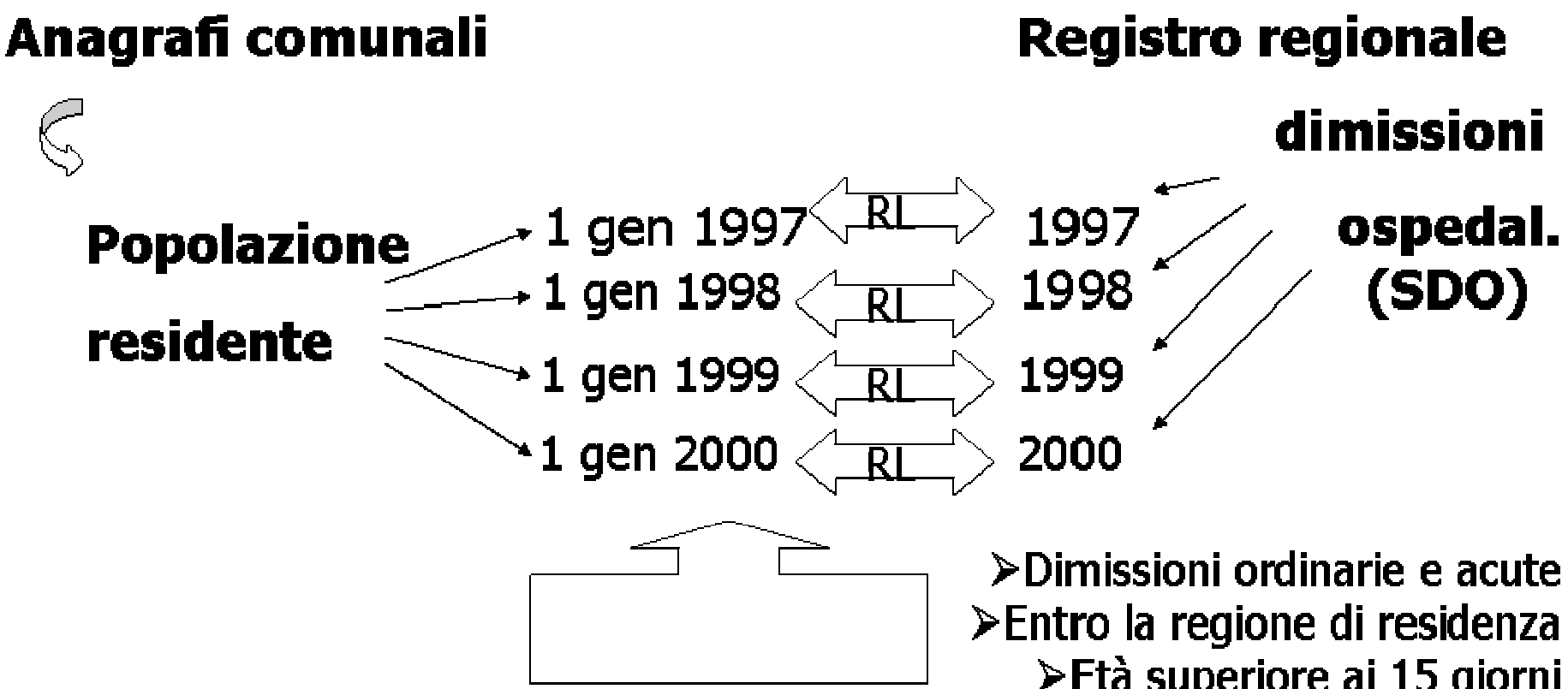
⇩

Reddito mediano per sezione di censimento delle famiglie

Attività svolta da SOGEI

Laboratorio dei Sistemi di Babele

Estrazioni ricoveri



- Dimissioni ordinarie e acute
- Entro la regione di residenza
- Età superiore ai 15 giorni

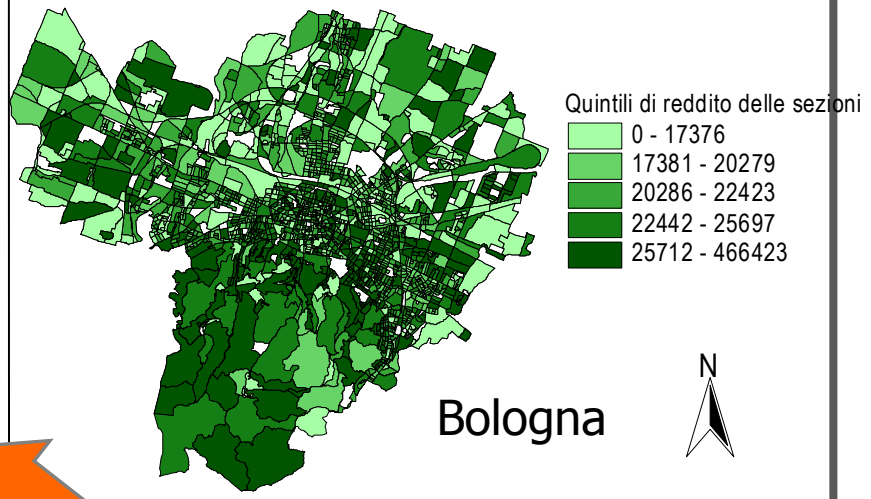
Reddito della sezione di censimento di ogni paziente (classificato in quintili)

b. Attribuzione di dati ecologici
Es1: fine

SDO

Anagrafe Comune

SOGEI
(Ministero Economia e Finanza)



Dati anagrafici individuali nominativi

95,1%

Dati nominativi '97-2000 (279.330)

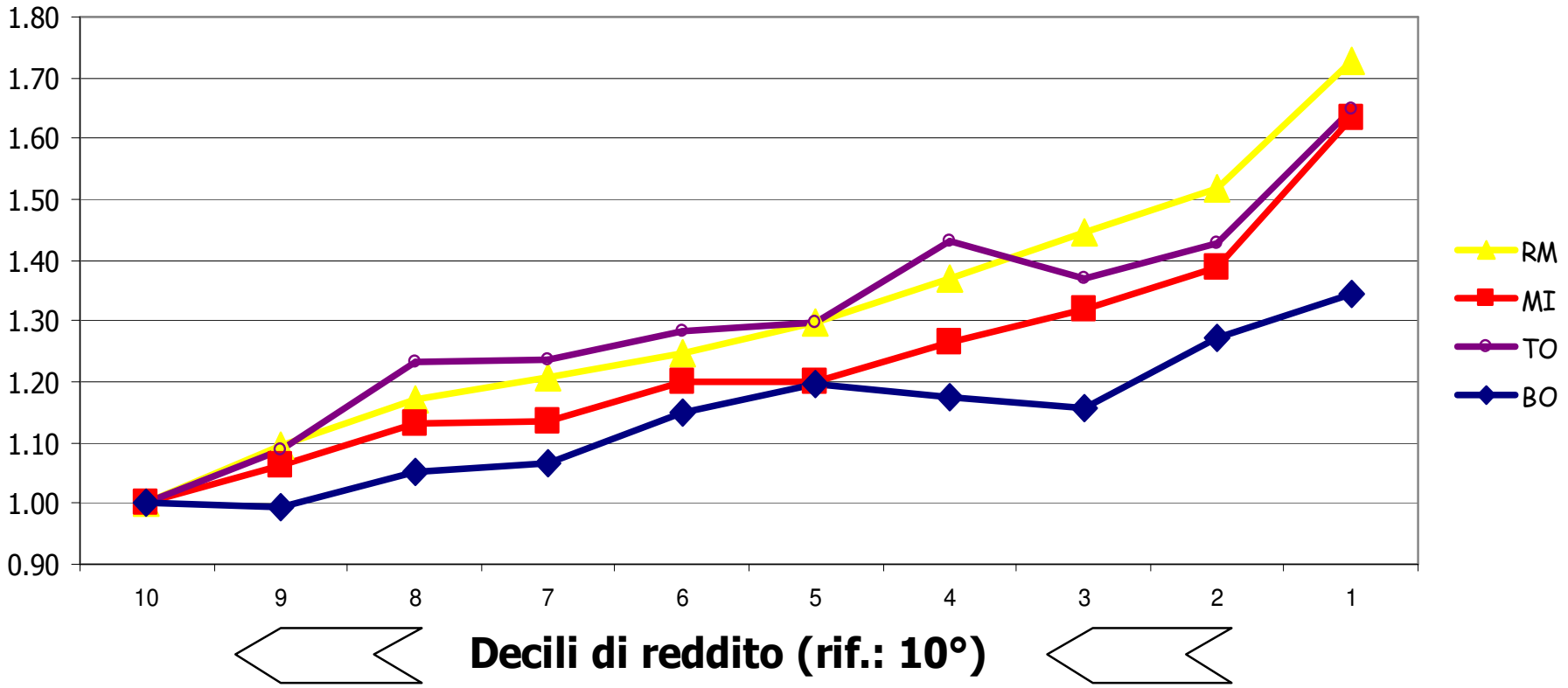
90,8%

Reddito mediano sezioni di censimento (1998)

archivio informatico storico: Bologna

Rapporto tra tassi di ospedalizzazione nei decili di reddito; Roma, Milano, Torino, Bologna

Maschi 1998



b. Attribuzione di dati ecologici
Es2: acquisizione dell'indice di deprivazione

Dati sanitari

(SDO, Mortalità…)

Anagrafe Comune

Dati anagrafici individuali nominativi

X%

Dati nominativi

Sezione di cens.

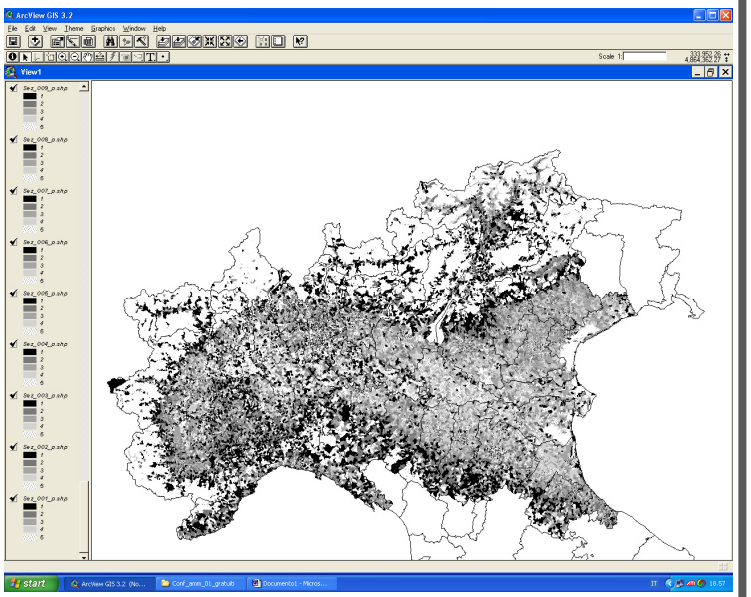
dei residenti,

o preferibilmente:

Georeferenziazione

ISTAT

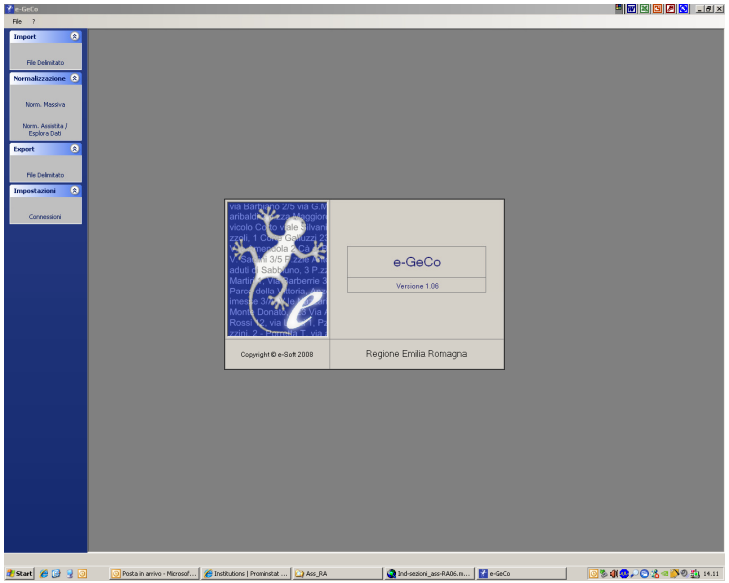
(Censimento 2001)



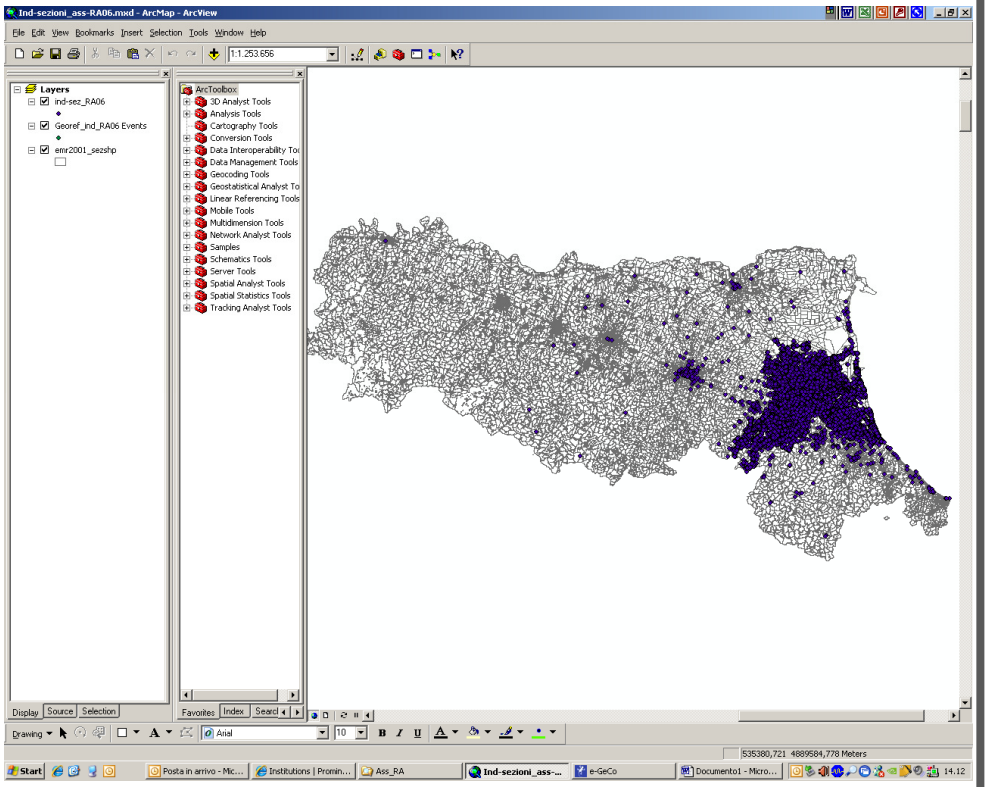
Frequenze per sezione

Indicatori sullo stato socio-demografico (es.: indice di deprivazione)

Georeferenziazione; *es.: Anagrafe degli Assisti* nell'ASL di Ravenna



I passo: normalizzazione e attribuzione delle coordinate spaziali degli indirizzi (comune, toponimo e n° civico). L'uso del programma **eGeCo** (stradario del 2007-2009) consente di georeferenziare il 90% di 310.302 assistiti



II passo: ***Join* spaziale** delle coordinate assegnate agli indirizzi con la cartografia (poligoni delle sezioni di censimento 2001).

L'attribuzione della zona geografica avviene, in questo caso, con qualche approssimazione. Es.: disallineamento dell'informazione del comune (116) nell'1 per mille (301 indirizzi), corrispondente ad un errore di circa 3 metri

Alternativa alla georeferenziazione: *Occorre che* ***anche i dati sanitari siano disaggregati a livello almeno di sezione di censimento*** (Comba, 2007)

a.4. Chiavi molto incomplete e parzialmente identificative

“Un metodo per presidiare l’equità nell’appropriatezza e nella continuità dei percorsi assistenziali” ***(prog. Min. Sal.)***

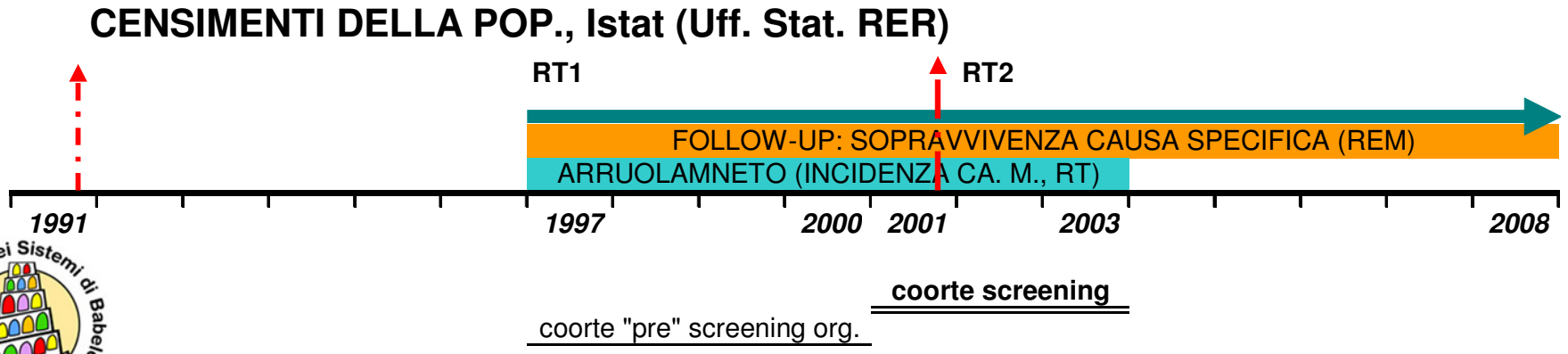
"Studio sulla sopravvivenza per tumore alla mammella in Emilia-Romagna in relazione alle condizioni socio-economiche e allo screening".

Definizione delle variabili di Stato Socio Economico

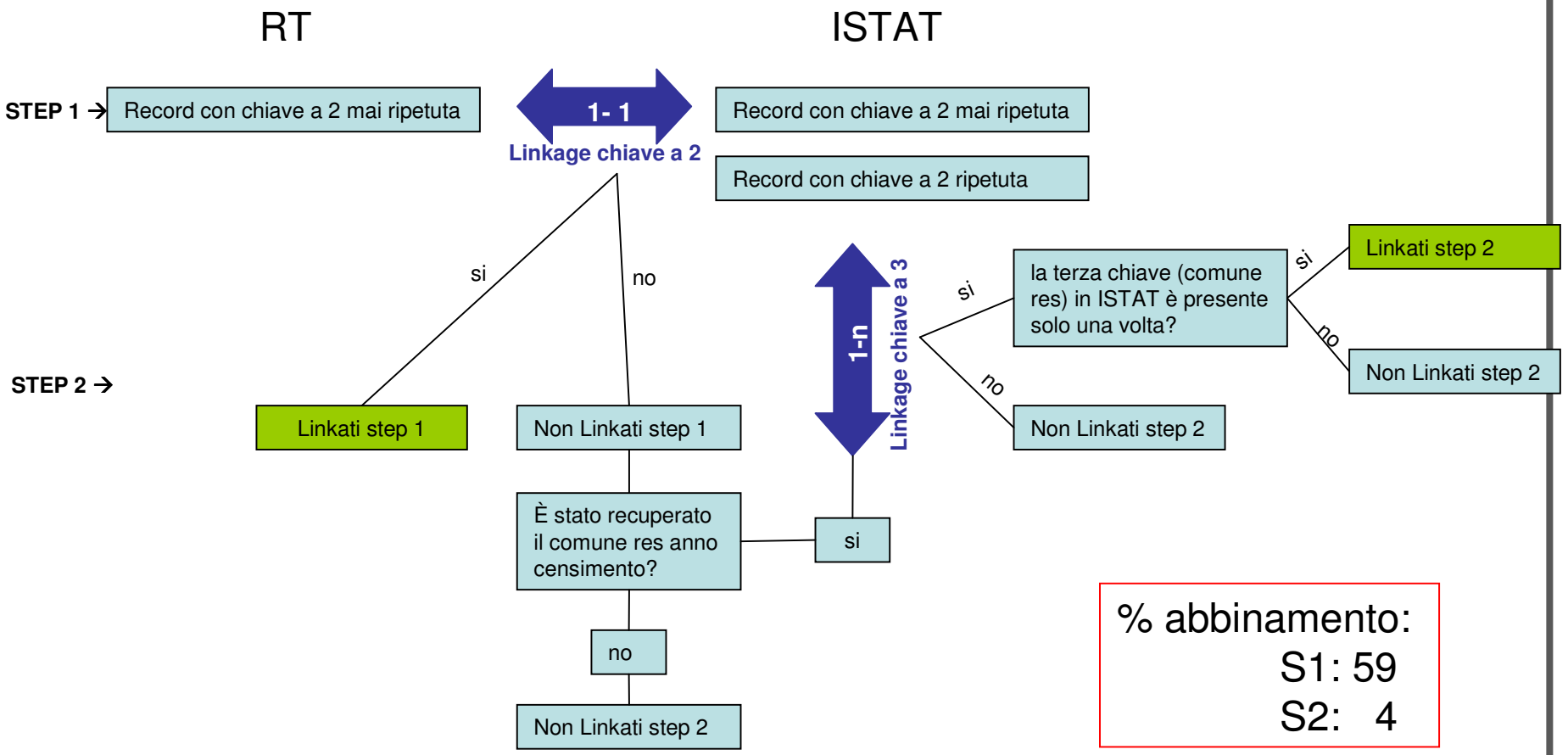
Il disegno dello studio

Coorte:

- arruolamento: casi incidenti di tumore alla mammella 1997-2003
- *follow-up* fino al 2008 (sopravvivenza a 5 anni causa spec.)
- attribuzione Stato Socio Economico (SES):



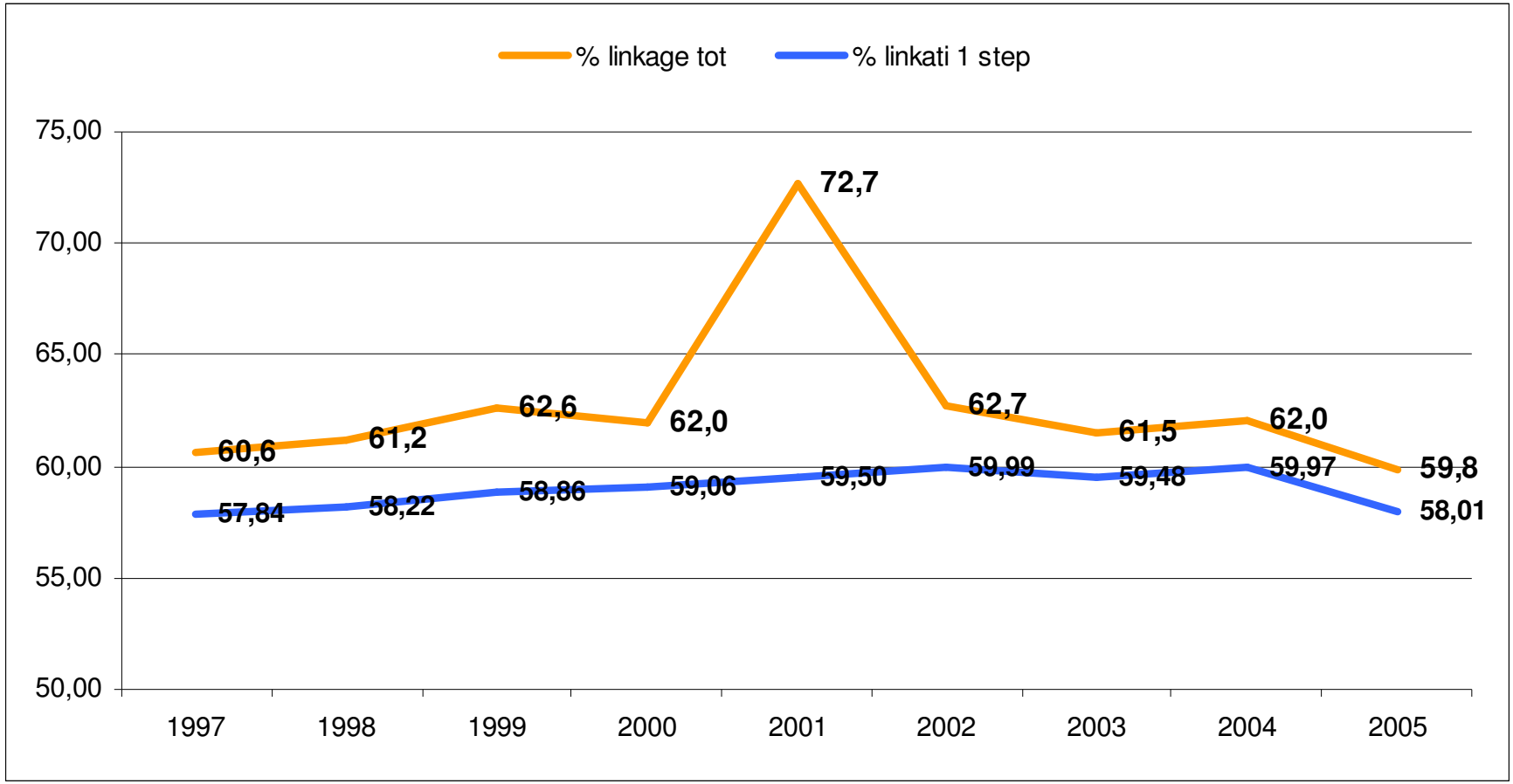
LINKAGE A 2 STEP tra registro tumore mammella - Censimento



% abbinamento:
S1: 59
S2: 4

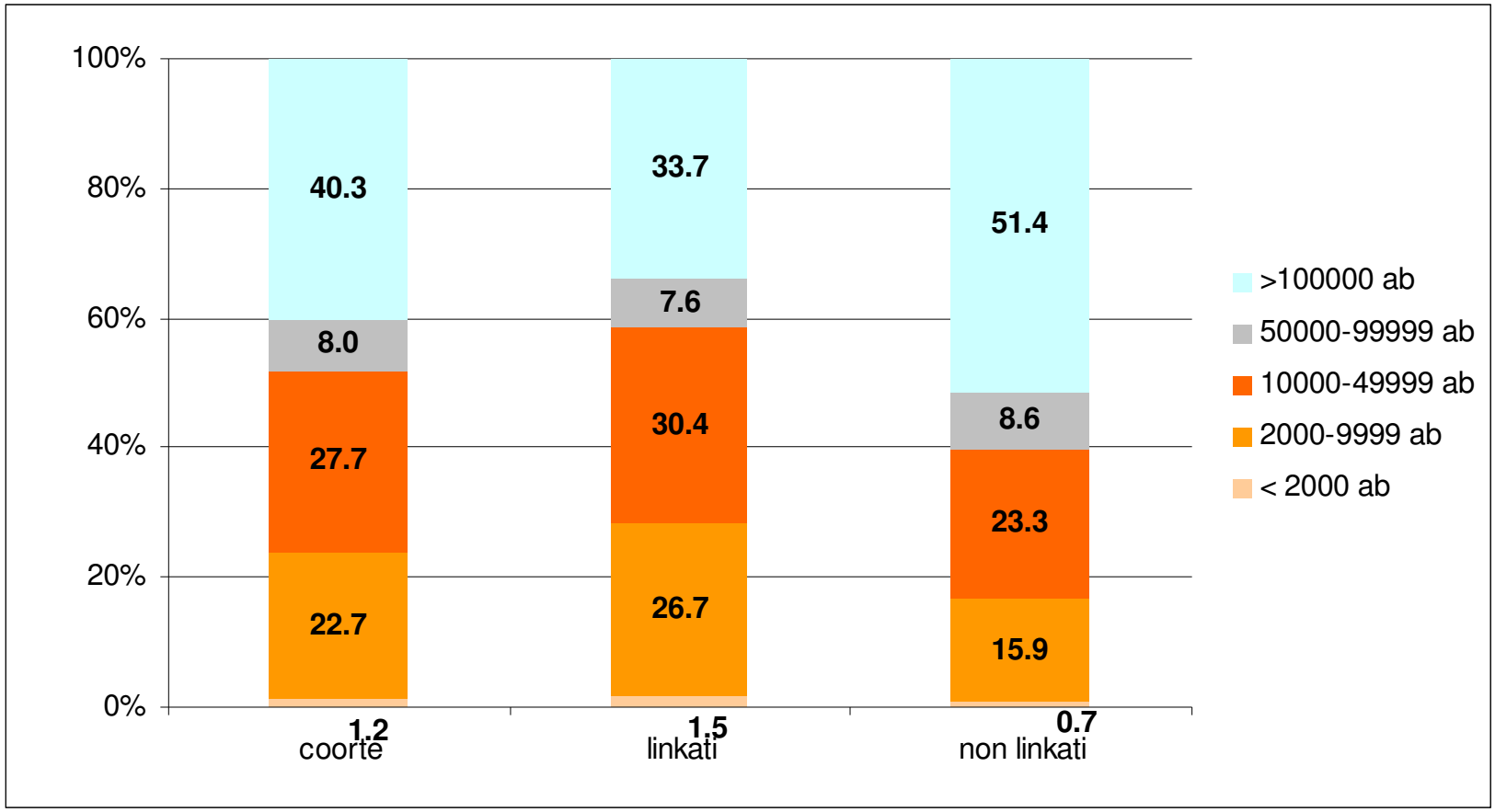
CHIAVE A 2: data nascita + comune nascita
CHIAVE A 3: data nascita + comune nascita + comune residenza anno censimento

LINKAGE A 2 STEP tra registro tumore mammella - Censimento



CHIAVE A 2: data nascita + comune nascita
CHIAVE A 3: data nascita + comune nascita + comune residenza anno censimento

LINKAGE A 2 STEP tra registro tumore mammella - Censimento



confronto record abbinati, record non abbinati e totale coorte per classi di ampiezza dei comuni di residenza all'incidenza

Bibliografica

- Dunn HL. Record linkage. Am J Public Health. 1946, 36: 1312-16.

- Fornari C, Madotto F, Demaria M, Romanelli A, Pepe P, Raciti M, Tancioni V, Chini F, Trerotoli P, Bartolomeo N, Serio G, Cesana G, Corrao G. Record-linkage procedures in epidemiology: an Italian multicentre study. Epidemiol Prev. 2008; 32(3 Suppl): 79-88.

- E&P, 2011: http://www.epiprev.it/materiali/2011/Supplemento_ESITI_full.pdf

- Raschetti R. Editoriale. Inserto BEN – Not Ist Super Sanità 2003; 16 (1) i.

- AIE, 2007. Convegno di primavera: L'integrazione di archivi elettronici per l'epidemiologia e la sanità pubblica, ISS 17-18 maggio:

  http://www.epidemiologia.it/?q=node/230

  - http://www.epidemiologia.it/sites/www.epidemiologia.it/files/R.Tessari.pdf
  - http://www.epidemiologia.it/sites/www.epidemiologia.it/files/N.Caranci.pdf
  - http://www.epidemiologia.it/sites/www.epidemiologia.it/files/A.Bena_P.Crosignani_M.Giraudo_R.Leombruni.pdf
  - http://www.epidemiologia.it/sites/www.epidemiologia.it/files/P.Comba_2.pdf

- Sacerdote C, Dalmasso M, Ciccone G, Demaria M, Gnavi R. Utilizzo di diverse chiavi identificative di soggetti presenti in diversi archivi. Inserto BEN – Not Ist Super Sanità 2003; 16 (1) i-iii.

- Cislaghi C, Zocchetti C, Russo A. Errori nell'identificazione personale e conseguenza sulla stima di prevalenza. Epidemiol Prev. 2012; 36(2): 126-8.

Laboratorio dei Sistemi di Babele

Grazie per l'attenzione

ncaranci@regione.emilia-romagna.it

valeria.fano@aslromad.it