



**Identificazione, sperimentazione e validazione di alcuni indicatori di qualità delle attività sanitarie e socio-sanitarie del territorio**

**Roma 14, 15 settembre 2006**



**Casi di studio**

Applicazione su gold standard  
Applicazione a livello regionale

## “Indicatori di salute per il territorio e per l’attività ospedaliera”

### sottoprogetto: “Modelli statistici di linkage SDO ReNCaM”

Collegamento tra schede di dimissione ospedaliera (SDO) e registri dei certificati delle cause di morte (ReNCaM) con tecniche di record linkage probabilistico

- operazione preliminare di validazione su gold standard
- applicazione delle tecniche ai dati di ricovero e alle schede di morte a livello regionale

Casi di studio – *Applicazione su gold standard*

## Variabili identificative dei soggetti

### ReNCaM

- cognome
- nome
- sesso
- data nascita (per dati regionali solo mese e anno)
- luogo nascita
- luogo residenza
- cittadinanza

### SDO

- codice fiscale
- sesso
- data di nascita
- luogo di nascita
- luogo di residenza
- cittadinanza

Casi di studio – *Applicazione su gold standard*

## Variabili comuni scelte per il linkage

cognome	3 lettere del codice fiscale
nome	3 lettere del codice fiscale
data nascita	mese e anno
luogo nascita	codice Istat

In questa fase non sono state considerate il luogo di residenza e la cittadinanza poiché soggette a cambiamenti nel tempo.

Casi di studio – *Applicazione su gold standard*

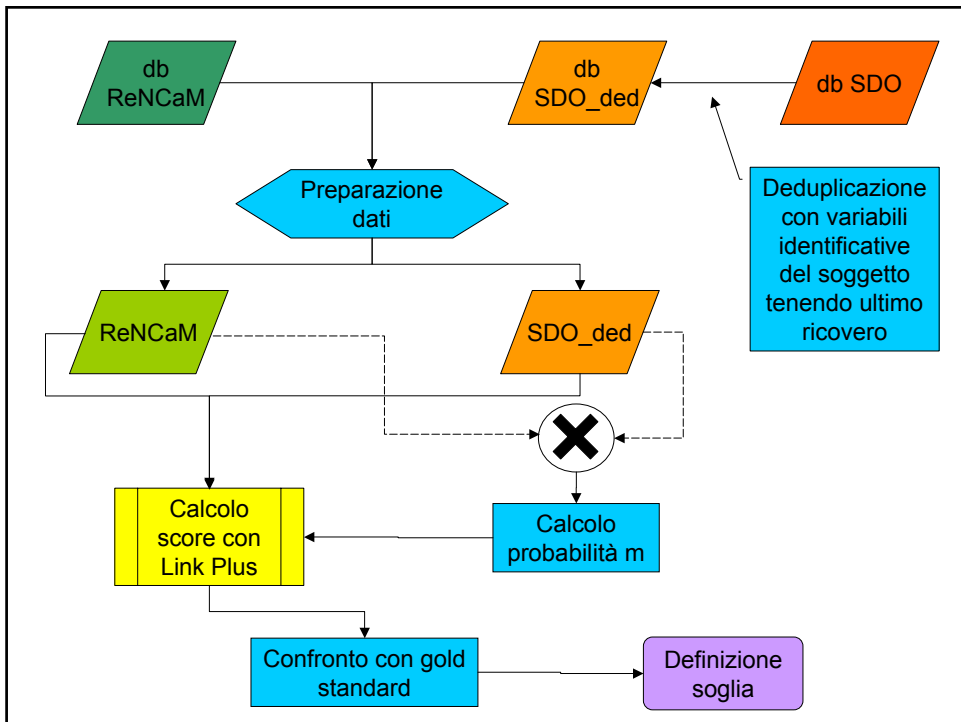
## Strumenti per il record linkage

Si è scelto di utilizzare il pacchetto freeware Link Plus integrandolo con alcune routine sas

Alcune caratteristiche del linkage probabilistico realizzato dal pacchetto :

- OR blocking
- 1-many linkage
- possibilità di usare stime di m fornite dall'utente

Casi di studio – *Applicazione su gold standard*



Variabili di blocking	AND blocking su: sesso, soundex del nome (3 lettere), soundex del cognome (3 lettere)
Variabili di linkage	cognome (3 lettere), nome (3 lettere), mese + anno nascita, luogo nascita
Stima delle probabilità m	Effettuata utilizzando l'algoritmo EM con programma appositamente sviluppato. Partendo da un valore iniziale di 0.9 per tutte le variabili sono state stimate queste probabilità:  $m_{\text{cog}}=0.9998$ $m_{\text{nom}}=0.9998$ $m_{\text{m+a}}=0.7732$ $m_{\text{luogo}}=0.955$  Valutazione della stabilità delle stime

## Studio su gold standard

Numero di SDO (1 anno): 213'902

Dopo deduplicazione deterministica: 152'320

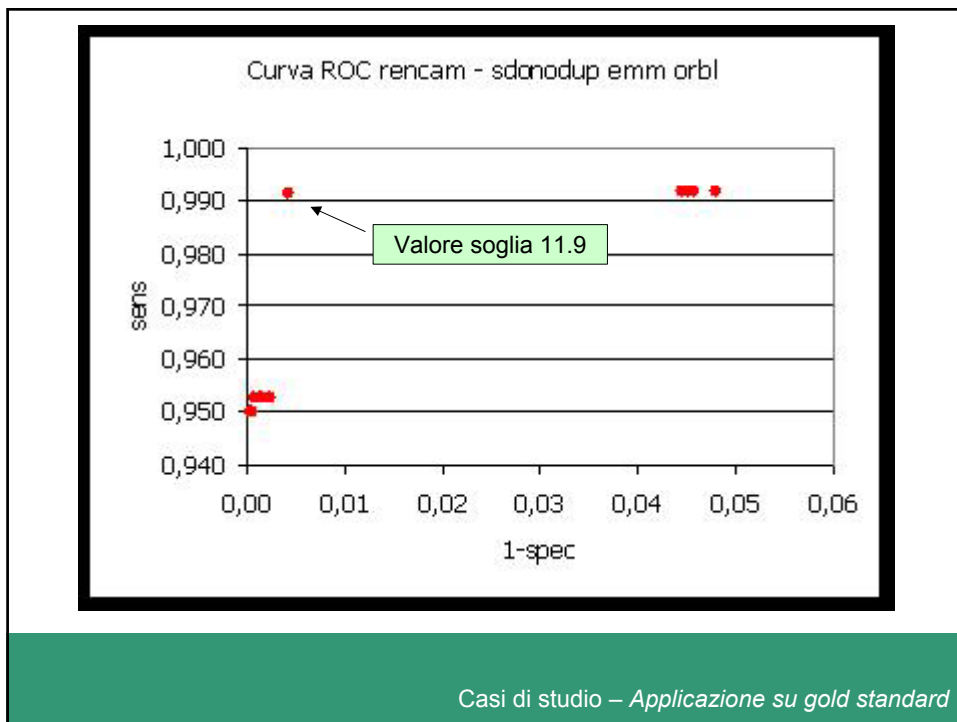
Numero di schede di morte (5 anni); 76'438

Casi di studio – Applicazione su gold standard

score	VP	FN_1	FN_2	FP	VN	SENS	SPEC
18,3	28527	224	1268	49	122252	0,950	1,000
17,2	28527	224	1268	59	122242	0,950	1,000
17	28527	224	1268	67	122234	0,950	0,999
16,6	28527	224	1268	68	122233	0,950	0,999
16,3	28527	224	1268	69	122232	0,950	0,999
16,2	28602	226	1191	90	122211	0,953	0,999
16,1	28602	226	1191	179	122122	0,953	0,999
15,4	28602	226	1191	180	122121	0,953	0,999
15,1	28602	226	1191	189	122112	0,953	0,998
15	28602	226	1191	193	122108	0,953	0,998
14,9	28602	226	1191	206	122095	0,953	0,998
14,5	28602	226	1191	208	122093	0,953	0,998
14,2	28602	226	1191	209	122092	0,953	0,998
14	28602	226	1191	306	121995	0,953	0,997
11,9	29762	235	22	513	121788	0,991	0,996
11,6	29765	235	19	5442	116859	0,992	0,956
10,8	29765	235	19	5527	116774	0,992	0,955

FN\_1 :il linkage probabilistico ha assegnato un match diverso da quello assegnato nel gold standard

FN\_2 :il linkage probabilistico non ha assegnato un match ma nel gold standard si



### Confronto con metodo deterministico

	appaia corrett.	appaia ≠ da gs	appaia ma gs no	Non appaia ma gs si	Non appaia e nemmeno gs
<b>SDO non duplicate</b>					
<b>Determ.</b>	28527	224	49	1268	122252
<b>Prob</b>	29762	235	513	22	121788
<b>Pr vs det</b>	+1235	+11	+464	-1246	-464

Casi di studio – Applicazione su gold standard

### Analisi esito ultimo ricovero

		Prob		Det		Gs	
		No scheda morte	Si scheda morte	No scheda morte	Si scheda morte	No scheda morte	Si scheda morte
Esito SDO	Non dec	121792	22173	123181	20784	122266	21699
	Dec	18	8337	339	8013	35	8320

		prob vs gs		Det vs gs	
		No scheda morte	Si scheda morte	No scheda morte	Si scheda morte
Esito SDO	Non dec	-474	+474	+915	-915
	Dec	-17	+17	+304	-304

*Casi di studio – Applicazione su gold standard*

### Distribuzione dei tempi al decesso per i dimessi vivi deceduti dopo la dimissione

	N	Min	P25	P50	P75	Max	Media	Dev std
Prob	22104	0	279	685	1155	1968	735.7	514.4
Det	20769	0	274	675	1143	1968	727.3	512
gs	21673	0	277	682	1152	1968	733.3	514.1

### Distribuzione per classi del tempo al decesso

	0 g	1-3 g	4-7 g	8-14 g	15-21 g	22-30 g	>30 g
Prob	46	217	246	344	231	326	20694
Det	44	203	233	333	216	313	19427
gs	46	216	244	343	228	325	20271

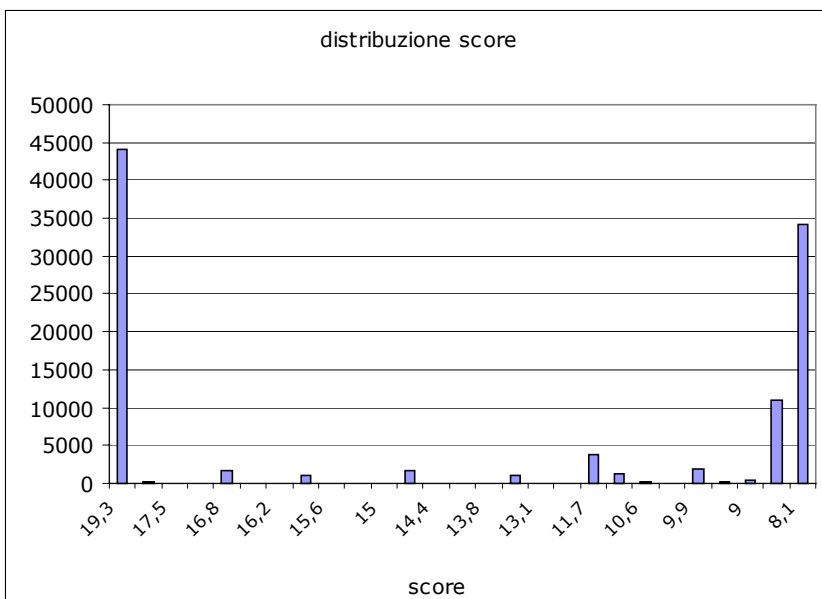
*Casi di studio – Applicazione su gold standard*

## Studio su dati regionali

		Deceduti
Numero di SDO	1'552'323	32'472
Dopo deduplicazione deterministica	1'137'823	32'364
Numero di schede di morte	86'471	

Casi di studio – Applicazione su dati regionali

distribuzione score

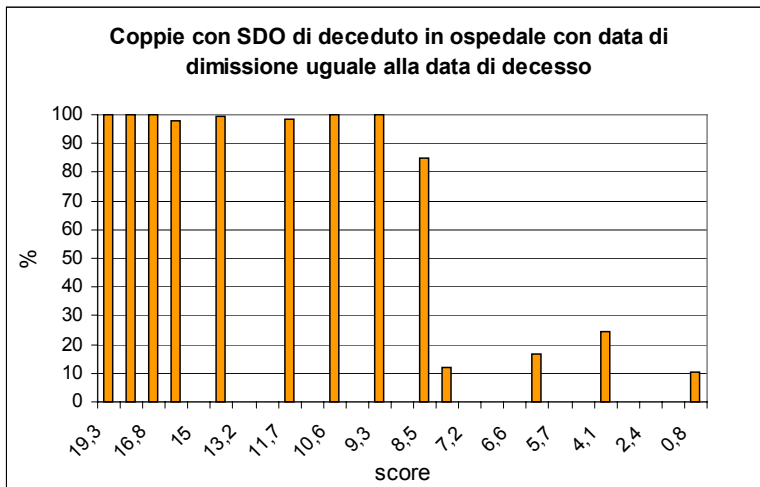


Casi di studio – Applicazione su dati regionali

SDO appaiate			
Totale SDO	Determ.	Prob. con soglia 11,9	Prob. con soglia 13,2
1'137'823	44'020	50'021	48'868

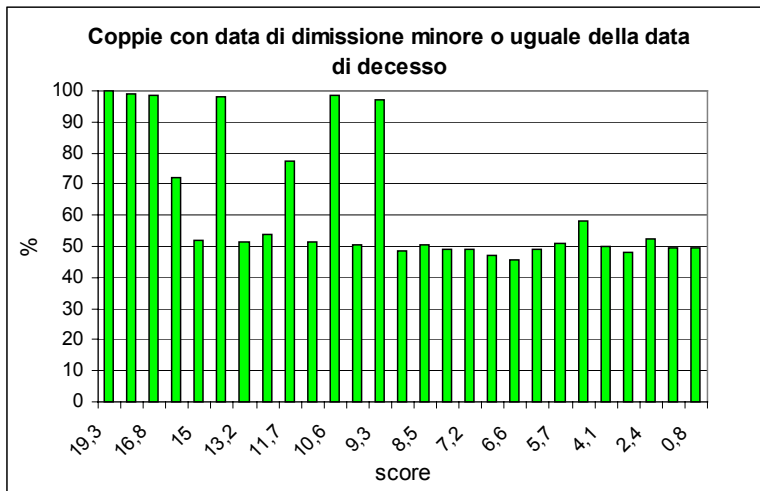
SDO appaiate			
Deceduti SDO	Determ.	Prob. con soglia 11,9	Prob. con soglia 13,2
32'364	24'194	26'289	26'270

Casi di studio – Applicazione su dati regionali



Score con almeno 20 coppie con SDO esitata in decesso

Casi di studio – Applicazione su dati regionali



Score con almeno 50 coppie