

Integrazione delle evidenze generate da basi di dati di diverse regioni o nazioni.

*Dipartimento di Statistica e Metodi Quantitativi
Laboratorio di Biostatistica, Epidemiologia e Salute Pubblica
Università degli Studi di Milano-Bicocca*

14 aprile 2016

Convegno ARCHES 2016



Integrare dati sanitari: perchè?

- Incrementare la potenza dello studio
- Miglioramento della precisione delle stime
- Identificazione di segnali non comuni
- Analisi stratificata

Di speciale interesse negli studi sulla sicurezza dei farmaci



Integrare dati sanitari: un esempio



- Obiettivo dello studio: studiare la associazione tra uso di NSAIDs e rischio di HF
- Lo studio viene condotto da cinque centri (in 4 paesi EU)
- Si desidera avere un unico risultato complessivo



Convegno ARCHES 2016

Integrare dati sanitari: Problemi e soluzioni



Problemi	Soluzioni
Diverse modalità nella raccolta dati	Protocollo comune per la definizione delle variabili d'interesse (esposizione, outcome, covariate)
Diversa qualità dei dati	Protocollo comune per il controllo di qualità
Diversa natura dei dati (e.g., GP-DB, DB amministrativi)	Common Data Model (CDM)

Anche con protocolli comuni ed un CDM:
Diversa quantità di dati disponibili da ogni DB

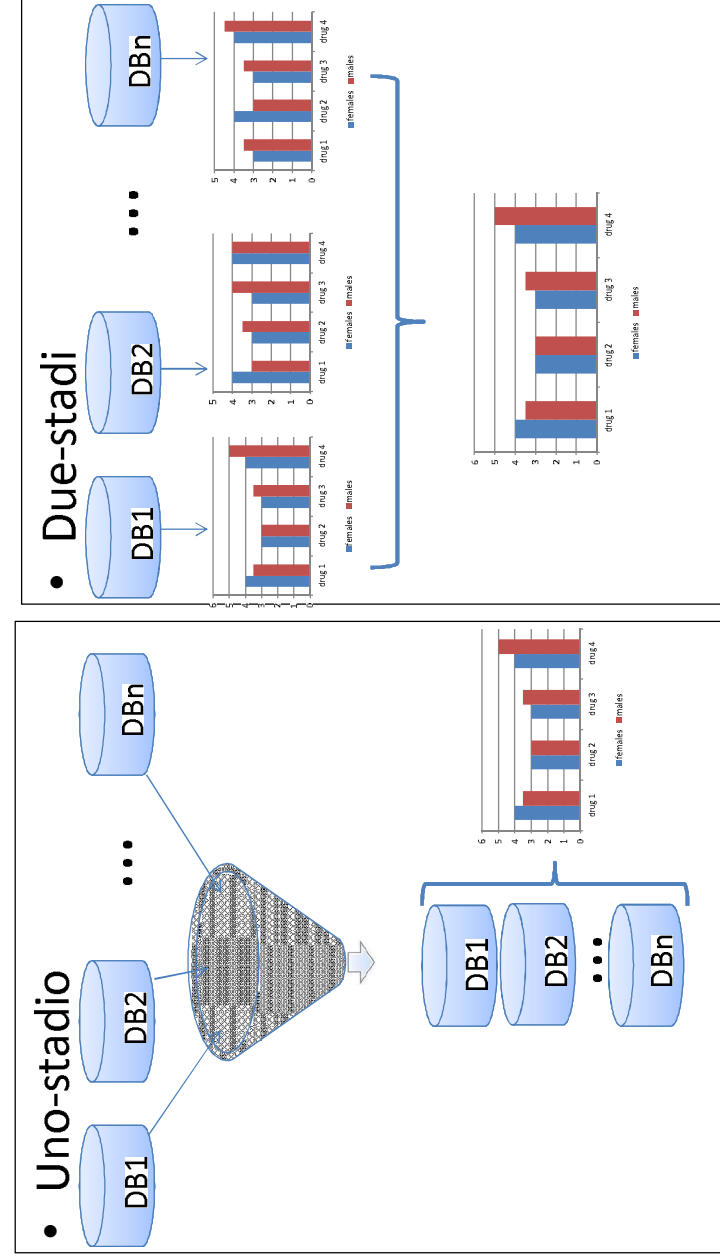


Convegno ARCHES 2016

Integrare dati sanitari: come? Metodi per l'integrazione delle evidenze

- Pooled analysis: integrazione di dati individuali da diversi database {uno-stadio}
- Meta-analisi: integrazione di risultati ottenuti da database diversi {due-stadi}

Metodi per l'integrazione di evidenze: schema illustrativo



Metodi per l'integrazione delle evidenze: dalla letteratura

Anche con protocolli comuni ed un CDM

- due-stadi è asintoticamente equivalente a uno-stadio in assenza di eterogeneità tra database (Lin & Zeng, 2010)
- due-stadi è almeno tanto efficiente quanto uno-stadio in presenza di eterogeneità tra database (Lin & Zeng, 2015)



Convegno ARCHES 2016

Metodi per l'integrazione delle evidenze: una simulazione

simulazione	DB1	DB2	DB3	DB4	DB5	uno-stadio	due-stadi
1	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}	b_{1-1st}	b_{1-2st}
2	b_{21}	b_{22}	b_{23}	b_{24}	b_{25}	b_{2-1st}	b_{2-2st}
...	...	...	...	...	...	...	...
1000	b_{10001}	b_{10002}	b_{10003}	b_{10004}	b_{10005}	$b_{1000-1st}$	$b_{1000-2st}$
media (sd)						$b_{1st} (s_{1st})$	$b_{2st} (s_{2st})$



Convegno ARCHES 2016

Una simulazione: due approcci a confronto

Scenario / Sub-scenario (parameter values)	True $\log(OR_{\infty})$	Bias		Estimator variance	
		One-stage	Two-stage	One-stage	Two-stage
→ 1 $OR_{\infty} = 0.7$	-0.36	-0.01	0.00	0.0127	0.0108
$OR_{\infty} = 1.5$	0.41	-0.02	-0.01	0.0127	0.0115
$OR_{\infty} = 5.0$	1.61	-0.01	0.00	0.0127	0.0112
2 $Var(\log OR_{\infty}) = 0$	0.41	-0.01	0.00	0.0024	0.0024
$Var(\log OR_{\infty}) = 0.01$	0.41	-0.01	0.00	0.0049	0.0047
$Var(\log OR_{\infty}) = 0.04$	0.41	-0.02	-0.01	0.0127	0.0115
$Var(\log OR_{\infty}) = 0.09$	0.41	-0.03	-0.01	0.0247	0.0210
$Var(\log OR_{\infty}) = 0.16$	0.41	-0.05	-0.01	0.0441	0.0373
$Var(\log OR_{\infty}) = 0.25$	0.41	-0.07	-0.01	0.0625	0.0511

Una simulazione: due approcci a confronto

Scenario / Sub-scenario (parameter values)	True $\log(OR_{\infty})$	Bias		Estimator variance	
		One-stage	Two-stage	One-stage	Two-stage
→ 1 $OR_{\infty} = 0.7$	-0.36	-0.01	0.00	0.0127	0.0108
$OR_{\infty} = 1.5$	0.41	-0.02	-0.01	0.0127	0.0115
$OR_{\infty} = 5.0$	1.61	-0.01	0.00	0.0127	0.0112
2 $Var(\log OR_{\infty}) = 0$	0.41	-0.01	0.00	0.0024	0.0024
$Var(\log OR_{\infty}) = 0.01$	0.41	-0.01	0.00	0.0049	0.0047
$Var(\log OR_{\infty}) = 0.04$	0.41	-0.02	-0.01	0.0127	0.0115
$Var(\log OR_{\infty}) = 0.09$	0.41	-0.03	-0.01	0.0247	0.0210
$Var(\log OR_{\infty}) = 0.16$	0.41	-0.05	-0.01	0.0441	0.0373
$Var(\log OR_{\infty}) = 0.25$	0.41	-0.07	-0.01	0.0625	0.0511

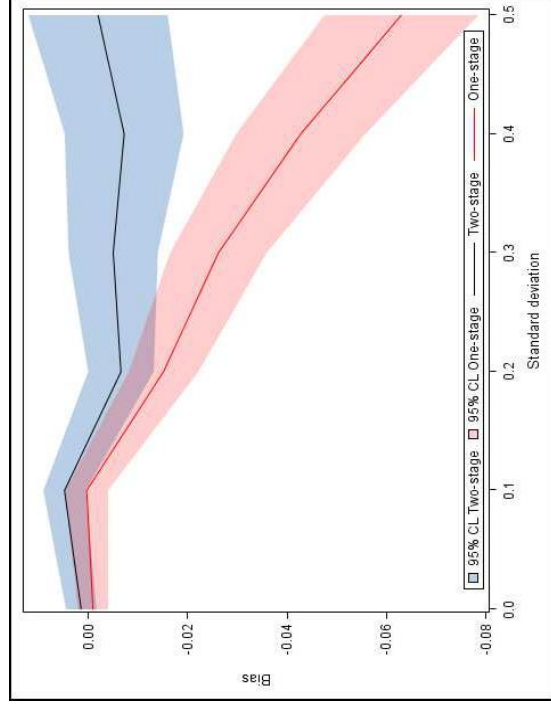
Una simulazione: due approcci a confronto

	Scenario / Sub-scenario (parameter values)	True $\log(\text{OR}_{\infty})$	Bias		Estimator variance	
			One-stage	Two-stage	One-stage	Two-stage
1	$\text{OR}_{\infty} = 0.7$	-0.36	-0.01	0.00	0.0127	0.0108
	$\text{OR}_{\infty} = 1.5$	0.41	-0.02	-0.01	0.0127	0.0115
	$\text{OR}_{\infty} = 5.0$	1.61	-0.01	0.00	0.0127	0.0112
2	$\text{Var}(\log \text{OR}_{\infty}) = 0$	0.41	-0.01	0.00	0.0024	0.0024
	$\text{Var}(\log \text{OR}_{\infty}) = 0.01$	0.41	-0.01	0.00	0.0049	0.0047
	$\text{Var}(\log \text{OR}_{\infty}) = 0.04$	0.41	-0.02	-0.01	0.0127	0.0115
	$\text{Var}(\log \text{OR}_{\infty}) = 0.09$	0.41	-0.03	-0.01	0.0247	0.0210
	$\text{Var}(\log \text{OR}_{\infty}) = 0.16$	0.41	-0.05	-0.01	0.0441	0.0373
	$\text{Var}(\log \text{OR}_{\infty}) = 0.25$	0.41	-0.07	-0.01	0.0625	0.0511

Una simulazione: due approcci a confronto

	Scenario / Sub-scenario (parameter values)	True $\log(\text{OR}_{\infty})$	Bias		Estimator variance	
			One-stage	Two-stage	One-stage	Two-stage
1	$\text{OR}_{\infty} = 0.7$	-0.36	-0.01	0.00	0.0127	0.0108
	$\text{OR}_{\infty} = 1.5$	0.41	-0.02	-0.01	0.0127	0.0115
	$\text{OR}_{\infty} = 5.0$	1.61	-0.01	0.00	0.0127	0.0112
2	$\text{Var}(\log \text{OR}_{\infty}) = 0$	0.41	-0.01	0.00	0.0024	0.0024
	$\text{Var}(\log \text{OR}_{\infty}) = 0.01$	0.41	-0.01	0.00	0.0049	0.0047
	$\text{Var}(\log \text{OR}_{\infty}) = 0.04$	0.41	-0.02	-0.01	0.0127	0.0115
	$\text{Var}(\log \text{OR}_{\infty}) = 0.09$	0.41	-0.03	-0.01	0.0247	0.0210
	$\text{Var}(\log \text{OR}_{\infty}) = 0.16$	0.41	-0.05	-0.01	0.0441	0.0373
	$\text{Var}(\log \text{OR}_{\infty}) = 0.25$	0.41	-0.07	-0.01	0.0625	0.0511

Una simulazione: due approcci a confronto



Scenario 2: confronto dell'andamento del bias tra uno- e due-stadi quando aumenta la eterogeneità dell'effetto dell'esposizione.

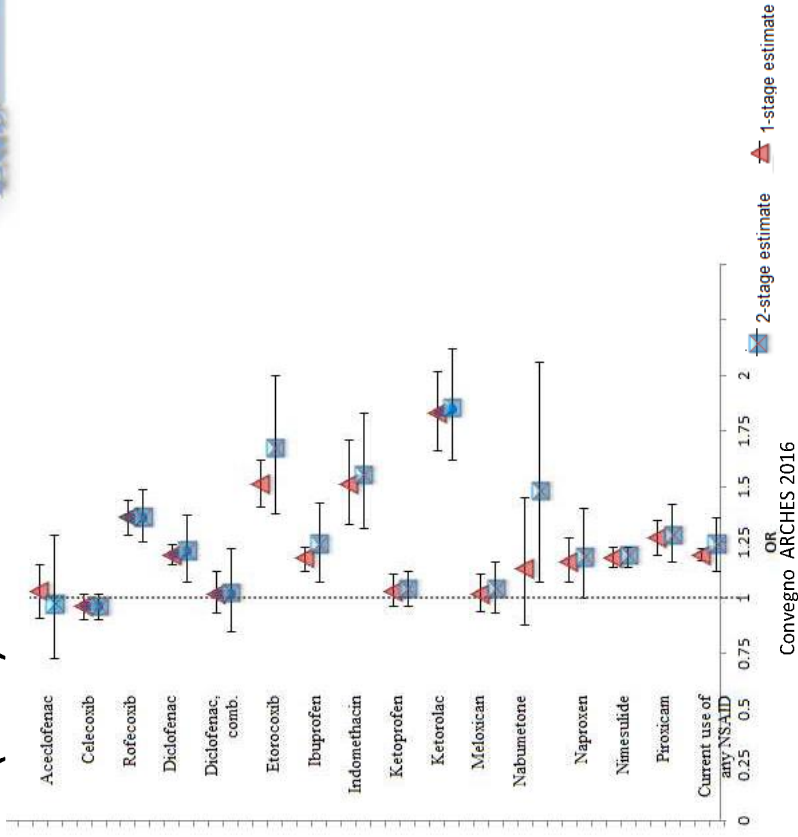
Una simulazione: conclusioni

Anche con protocolli comuni ed un CDM

Studi osservazionali caso-controllo con matching:

- La stima a uno-stadio dell'effetto dell'esposizione è tanto più distorta quanto maggiore l'eterogeneità tra database
- La stima a due- stadi è non distorta in presenza di eterogeneità
- Ulteriore vantaggio dell'approccio a due-stadi: aggiustamento per variabili confondenti

Metodi per l'integrazione delle evidenze: un caso empirico (SOS)



Conclusioni

- Il metodo a due-stadi non richiede la condivisione dei dati individuali ma solo dei dati aggregati -> **PRIVACY**
- Anche considerando i vantaggi del metodo a uno-stadio, l'approccio meta-analitico (due-stadi) dovrebbe essere considerato come analisi principale negli studi osservazionali
- Dal punto di vista statistico, il metodo a uno-stadio è preferibile quando i dati sono scarsi

References

- Lièvre M, Cucherat M and Leizorovicz A. Pooling, meta-analysis, and the evaluation of drug safety. *Current Controlled Trials in Cardiovascular Medicine* 2002, **3**:6
- Lin DY, Zeng D. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika* 2010;97:321-32
- Zeng D, Lin DY. On random-effects meta-analysis. *Biometrika* 2015;102:281-94