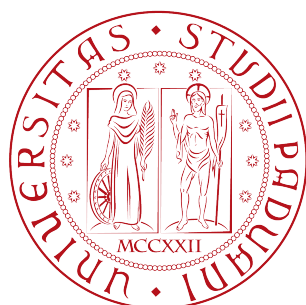


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze



RELAZIONE FINALE
**META-ANALISI IN PRESENZA DI RISCHI
COMPETITIVI**

Relatore Prof.ssa Giuliana Cortese
Dipartimento di Scienze Statistiche

Laureando: Daniela Corbetta
Matricola n° 1149610

Anno Accademico 2018/2019

Indice

1	Introduzione	3
2	Meta-analisi	5
2.1	Modelli per la meta-analisi	5
2.1.1	Modello a effetto fisso	6
2.1.2	Modello a effetti casuali	6
2.1.3	Statistiche per valutare l'eterogeneità	7
2.1.4	Stimatori per Δ^2	11
3	Analisi di sopravvivenza e rischi competitivi	19
3.1	Analisi di sopravvivenza	19
3.1.1	Stimatore di Kaplan-Meier	20
3.2	Rischi competitivi	21
3.2.1	Stima della funzione di incidenza cumulata	22
4	I dati: presentazione degli studi e prime analisi esplorative	25
4.1	I dati	25
4.2	Analisi esplorative	29
4.2.1	Analisi univariate	29
4.2.2	Analisi bivariate	35
5	Applicazione: meta-analisi per quantificare la sovrastima dello stimatore di Kaplan-Meier	37
5.1	Modello a effetto fisso	38
5.2	Modello a effetti casuali	38
6	Conclusioni	45

A	Codice R	49
A.1	Codice R per le analisi esplorative	49
A.1.1	Analisi univariate	49
A.1.2	Analisi bivariate	50
A.2	Codice R per la meta-analisi	51

Capitolo 1

Introduzione

La meta-analisi è una tecnica statistica che permette di integrare in maniera formale i risultati provenienti da diversi studi condotti su un medesimo argomento. Il primo a parlare di meta-analisi fu Glass (1976) in *Primary, Secondary and Meta-Analysis of Research*, un articolo in cui presentava tale tecnica come un'alternativa rigorosa alla discussione narrativa e casuale delle ricerche che caratterizzava i tentativi del periodo di far fronte alla quantità crescente di materiale disponibile in letteratura.

Il metodo proposto da Glass riceve da subito grande attenzione, e il numero di articoli pubblicati che ne fanno uso dal momento della sua introduzione ad oggi cresce esponenzialmente. Il motivo per cui la meta-analisi conosce una tale popolarità è il fatto che la produzione di lavori su uno stesso argomento è sempre più vasta, e si rende quindi indispensabile trovare una misura che li sintetizzi. Inoltre in campo medico capita spesso di lavorare con campioni di bassa numerosità, il che porta a studi con una potenza bassa e che possono essere discordanti. È quindi necessario raggruppare i risultati e fornire un effetto comune. Tra i vantaggi della meta-analisi c'è quindi, oltre alla possibilità di fornire il quadro riassuntivo di un argomento, il fatto che porta ad ottenere un campione più cospicuo con una conseguente maggiore potenza. Il primo passo di una meta-analisi è la stesura di un protocollo di ricerca, che deve essere il più possibile simile ai protocolli dei singoli studi. Sulla base del protocollo stilato si definiscono poi criteri di inclusione ed esclusione degli studi. Si procede a una ricerca bibliografica esaustiva di tutti i lavori sull'argomento presenti in letteratura e si includono nella meta-analisi gli studi che rispondono a tutti i requisiti specificati.

Bisogna poi valutare se i risultati degli studi scelti possono essere combinati in maniera ragionevole e se eventuali risultati discordanti possono essere imputati alle differenti caratteristiche degli studi. Infine si produce una misura comune di risultato a partire dai risultati dei singoli studi.

Per quanto la meta-analisi riduca l'errore casuale c'è la possibilità di incorrere in un errore sistematico. La prima causa di errore sistematico è il *selection bias*, strettamente collegato ai criteri di inclusione scelti e di conseguenza alla qualità degli studi coinvolti che possono portare a risultati falsamente negativi a causa della ridotta potenza. Un'altra causa di errore sistematico è l'eterogeneità tra gli studi, che può derivare da differenze cliniche note o da differenze metodologiche non note. Si rende necessaria un'analisi statistica che ha l'obiettivo di valutare se le variazioni tra i risultati delle diverse pubblicazioni possono essere o meno imputabili al caso. In presenza di eterogeneità bisogna considerare un modello a effetti casuali, mentre se l'eterogeneità non risulta confermata si può procedere con un modello a effetto fisso. Vi è poi il *publication bias*: studi con risultati negativi hanno minor probabilità di essere pubblicati rispetto a studi con risultati positivi, il che porta a una sovrastima dell'effetto terapeutico.

A fronte di queste cause d'errore possiamo individuare come limiti delle meta-analisi la loro dipendenza dalla esaustività della ricerca degli studi rilevanti, la loro dipendenza dalla qualità degli studi inclusi e il fatto che sovrastimino l'effetto positivo.

In questa tesi si utilizzerà una meta-analisi allo scopo di identificare i limiti dell'analisi di sopravvivenza condotta tramite lo stimatore di Kaplan-Meier in presenza di rischi competitivi (Lacny *et al.*, 2018), con attenzione ai modelli statistici da utilizzare.

Capitolo 2

Meta-analisi

2.1 Modelli per la meta-analisi

Un'importante decisione da prendere quando si procede a una meta-analisi è il tipo di modello con cui combinare i dati, cioè se adottare un modello a effetto fisso o un modello a effetti casuali. Un modello con effetto fisso assume l'omogeneità degli studi e di conseguenza considera come unica fonte di variabilità la varianza interna agli studi. Significa che le eventuali discordanze che emergono dai risultati sono dovute al caso, non si considerano eventuali differenze nel disegno dello studio, nella popolazione d'interesse e nel modo in cui il trattamento viene somministrato.

Un modello a effetti casuali tiene invece conto dell'eterogeneità e ingloba le caratteristiche di ogni studio in una variabile casuale detta effetto casuale. Effetti casuali relativi ad unità diverse sono assunti indipendenti. Le fonti di variabilità in questo tipo di modello sono quindi sia la varianza interna ai *trials* (varianza *within*) sia la varianza tra i *trials* (varianza *between*).

Condizioniamoci a una meta-analisi in ambito medico in cui interessa studiare l'effetto di un trattamento. Scopo di entrambi gli approcci è fornire una misura comune dell'effetto che interessa studiare, attraverso una media ponderata degli effetti osservati nei singoli studi. Si pone quindi il problema di come combinare gli effetti: la soluzione adottata nel contesto meta-analitico prevede di calcolare tale effetto assegnando ad ogni studio un peso proporzionale alla sua precisione, in maniera tale da dare più peso agli studi che portano maggior informazione. I pesi assegnati variano quindi in base all'approccio scelto perché dipendono dalle fonti di variabilità che entrano in gioco.

2.1.1 Modello a effetto fisso

Sia k il numero di diversi *trials* che entrano nella meta-analisi, e sia Y_i la variabile di interesse che rappresenta l'effetto del trattamento, $i = 1, \dots, k$. In un modello a effetto fisso si suppone ci sia un unico effetto comune a tutti gli studi che interessa stimare, cioè tutti gli studi vengono campionati da una popolazione con la stessa media μ . L'unica fonte di variabilità è data dalla varianza interna agli studi, che dipende dalla loro numerosità campionaria; studi che considerano popolazioni più ampie avranno quindi maggior precisione e di conseguenza un ruolo più importante nella stima dell'effetto combinato.

Il modello è quindi, per $i = 1, \dots, k$

$$Y_i = \mu + e_i$$

dove e_i rappresenta l'errore campionario, che si suppone avere media nulla e varianza v_i da stimare a partire dai dati e la cui espressione dipende dalla natura della misura di effetto Y_i (*risk ratio*, *odds ratio*,...). Si ha quindi

$$E[Y_i] = \mu$$

e

$$\text{var}(Y_i) = v_i$$

La misura comune dell'effetto μ viene stimata come

$$\hat{\mu} = \frac{\sum_{i=1}^k w_{i,fe} y_i}{\sum_{i=1}^k w_{i,fe}}$$

dove $w_{i,fe}$ rappresenta l'inverso della varianza di Y_i in un modello a effetto fisso, ossia $w_{i,fe} = (v_i)^{-1}$, e $\text{var}(\hat{\mu}) = (\sum_{i=1}^k w_{i,fe})^{-1}$.

2.1.2 Modello a effetti casuali

In un modello a effetti casuali si assume invece che ci sia una distribuzione per l'effetto di interesse e si utilizzano gli effetti osservati nei diversi studi per stimare tale distribuzione.

Anche in questo caso, sia k il numero di diversi *trials* che entrano nella meta-analisi e sia Y_i la variabile di interesse che rappresenta l'effetto del trattamento, $i = 1, \dots, k$. Il modello a effetti casuali divide la misura

dell'effetto del trattamento Y_i nella somma di due componenti: il vero effetto del trattamento θ_i e l'errore casuale e_i a media nulla. La varianza di e_i , v_i , viene calcolata a partire dagli studi. Si assume che θ_i possa essere scritto come $\theta_i = \mu + \delta_i$, dove μ rappresenta l'effetto medio del trattamento e δ_i è l'effetto casuale a media nulla che descrive lo scostamento dalla media per l' i -esimo studio. La popolazione di interesse è data dall'insieme di studi condotti sull'argomento analizzato e i *trials* che entrano nella meta-analisi vengono considerati come un campione casuale estratto da tale popolazione. Si utilizzano gli effetti osservati per stimare l'effetto medio μ e la varianza della popolazione ($\text{var}(\delta_i) = \Delta^2$, $i = 1, \dots, k$).

Il modello è quindi, per $i = 1, \dots, k$

$$Y_i = \mu + \delta_i + e_i$$

con

$$E[Y_i] = \mu$$

e

$$\text{var}(Y_i) = \Delta^2 + v_i$$

La varianza dell'effetto osservato del trattamento può cioè essere scissa nella somma tra la varianza rispetto all'effetto medio dovuta alle diversità degli studi, Δ^2 , e una varianza interna agli studi, v_i .

La prima fase dell'analisi consiste nella stima della variabilità tra studi, Δ^2 , che può essere ottenuta con il metodo dei momenti o con metodi legati alla teoria della verosimiglianza.

Si passa poi alla stima di μ :

$$\hat{\mu} = \frac{\sum_{i=1}^k w_{i,re} y_i}{\sum_{i=1}^k w_{i,re}}$$

con $w_{i,re} = (v_i + \hat{\Delta}^2)^{-1}$ e $\text{var}(\hat{\mu}) = (\sum_{i=1}^k w_{i,re})^{-1}$.

2.1.3 Statistiche per valutare l'eterogeneità

Per capire quale approccio adottare, se cioè un modello a effetto fisso o un modello a effetti casuali, è necessario valutare se sia presente o meno eterogeneità tra gli studi. Se infatti gli studi presentano caratteristiche simili,

ossia non emergono differenze per quanto riguarda la numerosità campionaria, la popolazione considerata, il modo in cui l'esperimento viene condotto, appare sensato avvalersi di un modello a effetto fisso e considerare che essi facciano parte di una popolazione accumulata dalla stessa media. Se al contrario gli studi appaiono diversi si può ipotizzare che l'effetto medio cambi da studio a studio e quindi considerare un modello ad effetti casuali. In altre parole, valutare la presenza di eterogeneità degli studi significa valutare se la varianza tra studi Δ^2 sia significativamente diversa da zero, ossia effettuare il test d'ipotesi

$$\begin{cases} H_0 : \Delta^2 = 0 \\ H_1 : \Delta^2 > 0 \end{cases} \quad (2.1)$$

A questo scopo la statistica più utilizzata è la statistica Q di Cochran, introdotta nel 1954, e che viene calcolata tramite la somma degli scarti quadratici dell'effetto osservato per ogni studio dall'effetto comune stimato, pesando ogni studio per l'inverso della sua varianza:

$$Q = \sum_{i=1}^k w_{i,fe} (y_i - \hat{\mu})^2 \quad (2.2)$$

Sotto H_0 , Q si distribuisce come un χ_{k-1}^2 . Si accetta quindi l'ipotesi di omogeneità per valori bassi di Q , mentre si rifiuta a favore dell'eterogeneità quando Q assume valori alti.

È stato evidenziato però (Hardy *et al.*, 1998) come tale statistica sia poco potente quando il numero di studi inclusi nella meta-analisi è ridotto, mentre tenda a portare al rifiuto dell'ipotesi nulla se il numero di studi inclusi nella meta-analisi è alto, anche quando in realtà l'eterogeneità è trascurabile. Il rischio è quindi di adottare erroneamente un modello a effetto fisso in una meta-analisi con pochi studi o al contrario di adottare, sempre erroneamente, un modello a effetti casuali in una meta-analisi in cui k è consistente. Inoltre Q fornisce solo un'indicazione dicotomica sulla presenza o assenza di eterogeneità, e non dà un'idea del grado con cui tale eterogeneità, se presente, influisce sui risultati del processo di meta-analisi.

Allo scopo di superare i limiti di questa statistica, Higgins e Thompson (2002) hanno proposto tre indici, il cui scopo è quantificare l'eterogeneità

e di conseguenza fornire anche un'indicazione di quanto i risultati ottenuti possano essere generalizzati ad altri contesti. Inoltre, nel formulare questi tre indici, gli autori si sono preoccupati di far sì che la loro bontà non dipendesse strettamente dal numero di studi inclusi.

Higgins e Thompson hanno considerato il caso particolare in cui $v_i = v$ per ogni studio, e si sono prefissati l'obiettivo di trovare una misura di eterogeneità $f(\mu, \Delta^2, v, k)$ che fosse funzione del valor medio dell'effetto del trattamento μ , della varianza *between* Δ^2 , della varianza *within* v e del numero di studi k . Nella ricerca di tale misura si sono preposti tre criteri:

1. Dipendenza dal grado di eterogeneità, ossia

$$f(\mu, \Delta_1^2, v, k) > f(\mu, \Delta^2, v, k) \quad \text{se} \quad \Delta_1^2 > \Delta^2$$

Questo criterio implica che la misura debba dipendere da Δ^2 , e che f debba crescere all'aumentare di Δ^2 .

2. Invarianza di scala, ossia

$$f(a + b\mu, b^2\Delta^2, b^2v, k) = f(\mu, \Delta^2, v, k) \quad \forall a, b$$

In questo modo è possibile confrontare meta-analisi con diverse scale di misura. Questo criterio implica che la misura non debba dipendere da μ .

3. Invarianza di grandezza, ossia

$$f(\mu, \Delta^2, v, k_1) = f(\mu, \Delta^2, v, k) \quad \forall k_1, k$$

Grazie a questa proprietà, le misure ricercate dai due autori superano il problema della dipendenza dal numero di studi che presentava la statistica Q . Il numero di studi infatti, anche se è legato all'eterogeneità, non deve condizionarne la portata.

Sono state proposte tre funzioni che rispettano questi criteri:

- $H^2 = \frac{Q}{k-1}$;
- $R^2 = \frac{v_{re}}{v_{fe}}$, dove $v_{re} = v + \Delta^2$ e $v_{fe} = v$
- $I^2 = \frac{\hat{\Delta}^2}{\hat{\Delta}^2 + \hat{\sigma}}$

Generalizzando ora al caso di eteroschedasticità, considerando quindi la varianza *within* v_i diversa da studio a studio, si rende necessario per il calcolo di I^2 scegliere un valore riassuntivo per \hat{v} che inglobi le diverse varianze. La soluzione proposta dai due autori è di considerare

$$s^2 = \frac{\sum_{i=1}^k w_{i,fe}(k-1)}{(\sum_{i=1}^k w_{i,fe})^2 - \sum_{i=1}^k w_{i,fe}^2}$$

che permette l'instaurarsi di una relazione tra H^2 e I^2 : $I^2 = \frac{H^2-1}{H^2}$.

Per quanto riguarda l'interpretazione, la statistica H^2 è il rapporto tra la statistica Q e i suoi gradi di libertà. Considerando che in assenza di eterogeneità $E[Q] = k - 1$, $H^2 = 1$ è indice di omogeneità.

Vale un'interpretazione analoga per quanto riguarda il valore unitario anche per la statistica R^2 , che viene calcolata nel modo seguente:

$$R^2 = \frac{\sum_{i=1}^k w_{i,fe}}{\sum_{i=1}^k w_{i,re}} = \frac{\sum_{i=1}^k (v_i)^{-1}}{\sum_{i=1}^k (v_i + \Delta^2)^{-1}}$$

Se $R^2 = 1$, allora si è in presenza di omogeneità ed è sufficiente adottare un modello ad effetti fissi.

Per quanto riguarda I^2 , essa può essere vista come il rapporto tra la varianza tra studi e la varianza totale delle Y_i , $i = 1, \dots, k$, il che suggerisce un'interpretazione di tale quantità come percentuale di variabilità dovuta all'eterogeneità. Maggiore è I^2 dunque, maggiore è l'impatto che l'eterogeneità ha nel determinare la varianza totale degli effetti. La valutazione di quanto la varianza tra studi sia importante è soggettiva, ma Higgins propone le seguenti linee guida per valutarne la portata:

- Per $I^2 \in (0, 0.3)$ l'eterogeneità osservata è bassa;
- Per $I^2 \in (0.3, 0.6)$ l'eterogeneità osservata è moderata;
- Per $I^2 \in (0.5, 0.9)$ l'eterogeneità osservata è sostanziale;
- Per $I^2 \in (0.75, 1)$ l'eterogeneità osservata è considerevole.

2.1.4 Stimatori per Δ^2

Nel caso in cui l'eterogeneità risulti non trascurabile e si decida di adottare un modello a effetti casuali, si rende necessario decidere come stimare la varianza tra studi Δ^2 . Sono stati proposti diversi stimatori che soddisfano diverse caratteristiche. Per quanto riguarda la soluzione, possono fornire soluzioni esplicite o rendere necessario ricorrere a procedure iterative per arrivare ad un risultato. Si dividono inoltre in stimatori che portano a una stima della varianza sempre positiva e in stimatori che invece forniscono una stima non negativa. Nel seguito si analizzeranno sette diversi stimatori e per ognuno di essi si valuterà la bontà in termini di distorsione (ossia si valuterà se $E[\hat{\Delta}^2] = \Delta^2$) ed errore quadratico medio ($MSE(\hat{\Delta}^2) = E[(\hat{\Delta}^2 - \Delta^2)^2]$) (Veroniki *et al.*, 2016). Gli stimatori considerati e le loro caratteristiche sono presentati in Tabella 2.1.

Tabella 2.1: Stimatori e loro caratteristiche

Stimatore	Soluzione	Positivo/Non-negativo
DerSimonian-Laird	Forma chiusa	Non negativo
Hedges-Olkin	Forma chiusa	Non negativo
Paule-Mandel	Iterativo	Non negativo
Hunter-Schmidt	Forma chiusa	Non negativo
Sidik-Jonkman	Forma chiusa	Positivo
Massima verosimiglianza	Iterativo	Non negativo
Massima verosimiglianza ristretta	Iterativo	Non negativo

Stimatore di DerSimonian e Laird

Il modello DerSimonian-Laird è il modello a effetti casuali più comunemente adottato nell'ambito delle meta-analisi, proposto da DerSimonian e Laird (1986) nella rivista *Contemporary Clinical Trials*. La popolarità di questo modello va ricercata nella sua semplicità, nel fatto che non richieda l'utilizzo di algoritmi iterativi per pervenire alle stime di media e varianza e nel fatto che pesi gli studi con quantità che riflettono il valore informativo e non solo per l'inverso della varianza. Fu inoltre il primo modello proposto che introduceva una metodologia statistica formale per incorporare l'eterogeneità. Fornisce uno stimatore per la varianza Δ^2 basato sul metodo dei momenti, al quale si perviene uguagliando il valore atteso della statistica Q di Cochran al

suo valore osservato:

$$E(Q) = \Delta^2 \left(\sum_{i=1}^k w_{i,fe} - \frac{\sum_{i=1}^k w_{i,fe}^2}{\sum_{i=1}^k w_{i,fe}} \right) + (k-1)$$

Si ottiene quindi

$$\hat{\Delta}_{DL}^2 = \max \left\{ 0, \frac{Q - (k-1)}{\sum_{i=1}^k w_{i,fe} - \frac{\sum_{i=1}^k w_{i,fe}^2}{\sum_{i=1}^k w_{i,fe}}} \right\} \quad (2.3)$$

Studi di simulazione suggeriscono che lo stimatore DerSimonian-Laird per la varianza tra studi porti a risultati distorti negativamente quando la vera varianza tra studi Δ^2 è alta. Jackson *et al.* (2010) hanno valutato l'efficienza asintotica di $\hat{\Delta}_{DL}^2$, ossia hanno verificato se per $k \rightarrow \infty$ esso raggiunge il limite inferiore di Rao-Cramer. Hanno concluso che lo stimatore DerSimonian-Laird è inefficiente quando gli studi inclusi nella meta-analisi hanno numerosità campionarie diverse e quando Δ^2 non è bassa; suggeriscono tuttavia che si possa usare quando la meta-analisi include molti studi.

Stimatore di Hedges e Olkin

Anche conosciuto come stimatore di Cochran, lo stimatore di Hedges e Olkin (Hedges e Olkin, 1985) è stato introdotto per la prima volta da Cochran nel 1954 e discusso da Hedges nel 1983. È anch'esso uno stimatore al quale si perviene tramite il metodo dei momenti e porta a una stima in forma chiusa della varianza tra studi che si ottiene ponendo la varianza campionaria $S^2 = \frac{1}{k-1} \sum_{i=1}^k (y_i - \bar{y})^2$ uguale al suo valore atteso.

La formula per lo stimatore è:

$$\hat{\Delta}_{HO}^2 = \max \left\{ 0, \frac{1}{k-1} \sum_{i=1}^k (y_i - \bar{y})^2 - \frac{1}{k} \sum_{i=1}^k v_i \right\} \quad (2.4)$$

con \bar{y} media non pesata degli effetti y_i , $i = 1, \dots, k$.

$\hat{\Delta}_{HO}^2$ è non distorto quando le varianze campionarie possono essere stimate in maniera non distorta. È stato inoltre messo a confronto con lo stimatore DerSimonian-Laird, con lo stimatore di massima verosimiglianza e con lo stimatore di massima verosimiglianza ristretta da DerSimonian e Laird (1986)

e da Thorlund *et al.* (2011), ed è stato dimostrato come in media produca stime della varianza tra studi più alta degli altri stimatori considerati. Dallo studio di Sidik e Jonkman (2007) è emerso come per Δ^2 alta lo stimatore $\hat{\Delta}_{HO}^2$ sia il meno efficiente in termini di errore quadratico medio, tuttavia se il numero di studi che entrano a far parte della meta-analisi è alto, è quello che presenta una minore distorsione. Nonostante questa proprietà e la facilità di calcolo, non viene comunemente utilizzato.

Differisce da $\hat{\Delta}_{DL}^2$ perchè si basa sulla varianza non pesata delle stime dell'effetto del trattamento, mentre lo stimatore DerSimonian-Laird si basa sulla stima pesata (DerSimonian, Laird, 1986).

Entrambi gli stimatori sono infatti casi particolari dello stimatore generalizzato del metodo dei momenti

$$\hat{\Delta}_{GMM}^2 = \max \left\{ 0, \frac{Q_a - \left(\sum_{i=1}^k a_i v_i - \frac{\sum_{i=1}^k a_i^2 v_i}{\sum_{i=1}^k a_i} \right)}{\sum_{i=1}^k a_i - \frac{\sum_{i=1}^k a_i^2}{\sum_{i=1}^k a_i}} \right\} \quad (2.5)$$

al quale si perviene uguagliando la statistica di Cochran generalizzata per la varianza *between* al suo valore atteso:

$$E[Q_a] = \Delta^2 \left(\sum_{i=1}^k a_i - \frac{\sum_{i=1}^k a_i^2}{\sum_{i=1}^k a_i} \right) + \left(\sum_{i=1}^k a_i v_i - \frac{\sum_{i=1}^k a_i^2 v_i}{\sum_{i=1}^k a_i} \right) \quad (2.6)$$

Le quantità a_i sono i pesi che vengono assegnati ad ogni studio. Lo stimatore DerSimonian-Laird considera $a_i = w_{i,fe}$, mentre lo stimatore Hedges-Olkin considera $a_i = 1/k$

Stimatore di Paule e Mandel

Il metodo di Paule-Mandel è un metodo iterativo introdotto da Paule e Mandel (1982) che si basa sulla statistica Q_a generalizzata, con $a_i = w_{i,re} = (v_i + \Delta^2)^{-1}$:

$$Q_{gen} = \sum_{i=1}^k w_{i,re} (y_i - \hat{\mu}_{re}(\Delta^2))^2 \sim \chi_{k-1}^2 \quad (2.7)$$

Come per i due casi precedenti, per ottenere lo stimatore si eguaglia la statistica Q_{gen} al suo valore atteso. In questo caso non è però possibile ottenere una stima esplicita per $\hat{\Delta}_{PM}$ e si procede con una procedura iterativa.

Questo metodo è più robusto del metodo DerSimonian-Laird anche quando le assunzioni sottostanti non sono soddisfatte. È asintoticamente non distorto, mentre nel caso finito produce stime distorte positivamente per Δ^2 bassa e stime distorte in senso negativo per Δ^2 alta. Studi di simulazione hanno dimostrato che offre prestazioni migliori rispetto a $\hat{\Delta}_{DL}^2$ e $\hat{\Delta}_{REML}^2$ in termini di distorsione, mentre per quanto riguarda l'efficienza gli altri due stimatori forniscono risultati migliori.

Stimatore di Hunter e Schmidt

Lo stimatore Hunter-Schmidt (Hunter e Schmidt, 2004) è stato introdotto nel 2004 ed è così definito:

$$\hat{\Delta}_{HS}^2 = \max\left\{0, \frac{Q - k}{\sum_{i=1}^k w_{i,fe}}\right\} \quad (2.8)$$

È uno stimatore distorto in senso negativo (Viechtbauer, 2005). Ha invece buone proprietà per quanto riguarda l'errore quadratico medio: è stato dimostrato infatti che è simile a quello dello stimatore di massima verosimiglianza e che entrambi sono inferiori a quelli degli stimatori DerSimonian-Laird, massima verosimiglianza ristretta e Hedges-Olkin (Viechtbauer, 2005).

Stimatore di Sidik e Jonkman

Introdotta da Sidik e Jonkman (2005), è uno stimatore che rende disponibile una forma esplicita per il calcolo della varianza tra studi e si basa sui minimi quadrati pesati. Sia $\Delta_0^2 = \frac{1}{k} \sum_{i=1}^k (y_i - \bar{y})^2$ una stima iniziale per la varianza tra studi e sia $\hat{q}_i = v_i / \Delta_0^2 + 1$. L'espressione dello stimatore è:

$$\hat{\Delta}_{SJ}^2 = \frac{1}{k-1} \sum_{i=1}^k \hat{q}_i^{-1} (y_i - \hat{\mu}_{\hat{q},re})^2 \quad (2.9)$$

dove $\hat{\mu}_{\hat{q},re}$ è la stima pesata dell'effetto considerando gli effetti casuali:

$$\hat{\mu}_{\hat{q},re} = \frac{\sum_{i=1}^k \hat{q}_i^{-1} y_i}{\sum_{i=1}^k \hat{q}_i^{-1}} = \frac{\sum_{i=1}^k w_{i,re} y_i}{\sum_{i=1}^k w_{i,re}}, \quad (2.10)$$

con $w_{i,re} = (\Delta^2 + v_i)^{-1}$. Si noti come tale stimatore a differenza dei precedenti sia sempre positivo e non assuma mai il valore zero.

Per quanto riguarda la distorsione, è stato dimostrato che tra gli stimatori

considerati è quello al quale, per valori bassi di Δ^2 , è associato un *bias* maggiore, che diminuisce gradualmente al crescere della varianza tra studi, mentre l'errore quadratico medio è inferiore rispetto a quello dello stimatore DerSimonian-Laird per valori alti di k e Δ^2 e inferiore rispetto a quello dello stimatore Hedges-Olkin indipendentemente dal numero di studi inclusi e dal valore della varianza (Sidik e Jonkman, 2007).

I metodi appena presentati forniscono stimatori che vengono calcolati tramite il metodo dei momenti, e non richiedono quindi che vengano fatte ipotesi sulla distribuzione dell'effetto di interesse Y_i , $i = 1, \dots, k$. Si presentano ora due stimatori che si basano sulla teoria della verosimiglianza, lo stimatore di massima verosimiglianza e lo stimatore di massima verosimiglianza ristretta, che poggiano sulle assunzioni $Y_i \sim N(\theta_i, v_i)$ e $\theta_i \sim N(\mu, \Delta^2)$.

Stimatore di massima verosimiglianza

Per calcolare lo stimatore di massima verosimiglianza (Hardy e Thompson, 1996; Thompson e Sharp, 1999) si parte dalla funzione di verosimiglianza, se ne calcola il logaritmo e si massimizza la funzione di log-verosimiglianza. Si tratta quindi di massimizzare la funzione di log-verosimiglianza di una variabile casuale Normale di media μ e varianza totale data dalla somma della varianza campionaria e della varianza tra studi, $\text{var}(Y_i) = v_i + \Delta^2$. La funzione di log-verosimiglianza è:

$$\ell(\mu, \Delta^2) = -\frac{k}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k \log(v_i + \Delta^2) - \frac{1}{2} \sum_{i=1}^k \frac{(y_i - \mu)^2}{(v_i + \Delta^2)} \quad (2.11)$$

Le stime per i parametri di interesse sono:

$$\hat{\mu}_{re}(\hat{\Delta}_{ML}^2) = \frac{\sum_{i=1}^k w_{i,re} y_i}{\sum_{i=1}^k w_{i,re}} \quad (2.12)$$

e

$$\hat{\Delta}_{ML}^2 = \max \left\{ 0, \frac{\sum_{i=1}^k w_{i,re}^2 ((y_i - \hat{\mu}_{re}(\hat{\Delta}_{ML}^2))^2 - v_i)}{\sum_{i=1}^k w_{i,re}^2} \right\}, \quad (2.13)$$

con $w_{i,re} = 1/(v_i + \hat{\Delta}_{ML}^2)$.

Per arrivare alle stime finali bisogna avvalersi di una procedura iterativa. Si parte assegnando alla varianza $\hat{\Delta}_{ML}^2$ un valore plausibile sulla base delle

conoscenze che si hanno riguardo la diversità tra gli studi che fanno parte della meta-analisi o, in alternativa, il valore derivante dalla stima ottenuta tramite un metodo non iterativo, e si ottiene così una stima iniziale per $\hat{\mu}_{re}(\hat{\Delta}_{ML}^2)$ con la quale si aggiorna il valore di $\hat{\Delta}_{ML}^2$. Si itera fino a convergenza e si perviene alle stime finali di $\hat{\Delta}_{ML}^2$ e $\hat{\mu}_{re}(\hat{\Delta}_{ML}^2)$.

In alternativa ci si può avvalere di metodi numerici per la massimizzazione della funzione di log-verosimiglianza, per esempio utilizzando l'algoritmo di Newton-Raphson.

Lo stimatore così ottenuto gode delle proprietà degli stimatori di massima verosimiglianza: è quindi asintoticamente non distorto e asintoticamente efficiente (ossia raggiunge il limite inferiore di Rao-Cramer). Tuttavia studi di simulazione (Sidik e Jonkman, 2007; Viechtbauer, 2005) hanno messo in luce come per campioni finiti, quando cioè k è basso, se Δ^2 è alto presenta una forte distorsione negativa. A causa di questa forte distorsione l'uso di $\hat{\Delta}_{ML}^2$ è sconsigliato.

Per quanto riguarda invece l'errore quadratico medio, è stato dimostrato che lo stimatore di massima verosimiglianza gode di un minor errore quadratico medio se confrontato agli stimatori di Hedges-Olkin, Sidik-Jonkman e massima verosimiglianza ristretta (Sidik e Jonkman, 2007; Thompson e Sharp, 1999). Un problema da tenere in considerazione quando si utilizza uno stimatore basato su una procedura iterativa è che la convergenza non è assicurata, e c'è il rischio che non si riesca a pervenire a una stima di $\hat{\Delta}_{ML}^2$ a causa di una funzione di log-verosimiglianza piatta o che presenta massimi locali.

Stimatore di massima verosimiglianza ristretta

Per ovviare alla distorsione negativa che comporta l'utilizzo dello stimatore di massima verosimiglianza si può utilizzare lo stimatore di massima verosimiglianza ristretta, che si basa sull'ottimizzazione della funzione di log-verosimiglianza marginale dei residui $r_i = y_i - \hat{\mu}$ (Raudenbush, 2009):

$$\begin{aligned} \ell(\Delta^2) = & -\frac{k}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k \log(v_i + \Delta^2) \\ & - \frac{1}{2} \sum_{i=1}^k \frac{(y_i - \hat{\mu}_{re}(\hat{\Delta}_{ML}^2))^2}{v_i + \Delta^2} \\ & - \frac{1}{2} \log \left(\sum_{i=1}^k \frac{1}{(v_i + \Delta^2)} \right) \end{aligned} \quad (2.14)$$

Risolvendo, si ottiene la seguente espressione per $\hat{\Delta}_{REML}^2$:

$$\hat{\Delta}_{REML}^2 = \max \left\{ 0, \frac{\sum_{i=1}^k w_{i,re}^2 ((y_i - \hat{\mu}_{re}(\hat{\Delta}_{REML}^2))^2 - v_i)}{\sum_{i=1}^k w_{i,re}^2} + \frac{1}{\sum_{i=1}^k w_{i,re}} \right\}, \quad (2.15)$$

con $w_{i,re} = 1/(v_i + \hat{\Delta}_{REML}^2)$.

Come per lo stimatore precedente si ricorre a una procedura iterativa, partendo dall'assegnare a $\hat{\Delta}_{REML}^2$ un valore non negativo e aggiornando la stima ad ogni passo, fino al momento in cui da un'iterazione all'altra la variazione del valore della stima è al di sotto di una soglia prestabilita.

Anche questo stimatore gode delle proprietà degli stimatori di massima verosimiglianza e quindi è asintoticamente corretto e asintoticamente efficiente. È stato dimostrato che lo stimatore di massima verosimiglianza ristretta ha un errore quadratico medio inferiore a quello dello stimatore di Hedges-Olkin, superiore a quello degli stimatori di massima verosimiglianza e di Hunter-Schmidt e simile a quello dello stimatore di DerSimonian-Laird. Inoltre se gli studi che entrano a far parte della meta-analisi sono studi di grandi dimensioni, è lo stimatore che porta a risultati migliori (Viechtbauer, 2005). È disponibile una versione approssimata di tale stimatore (Sidik e Jonkman,

2007; Thompson e Sharp, 1999):

$$\hat{\Delta}_{AREML}^2 = \max \left\{ 0, \frac{\sum_{i=1}^k w_{i,re}^2 \left(\left(\frac{k}{k-1} \right) (y_i - \hat{\mu}_{re}(\hat{\Delta}_{AREML}^2)) \right)^2 - v_i}{\sum_{i=1}^k w_{i,re}^2} \right\}, \quad (2.16)$$

con $w_{i,re} = 1/(v_i + \hat{\Delta}_{AREML}^2)$.

Le stime ottenute con lo stimatore di massima verosimiglianza ristretta e con quello di massima verosimiglianza ristretta approssimata sono praticamente identiche.

Capitolo 3

Analisi di sopravvivenza e rischi competitivi

3.1 Analisi di sopravvivenza

Nell'ambito dell'analisi di sopravvivenza, è di interesse studiare il tempo di attesa al verificarsi di uno specifico evento. Tuttavia nella realtà può capitare che non si riesca ad osservare l'evento di interesse durante il periodo dello studio a causa del fatto che l'unità statistica non sperimenta l'evento in tale periodo o a causa del fatto che è uscita dallo studio; in questo caso si è in presenza di un dato censurato.

Sia X la variabile aleatoria che descrive il tempo di attesa all'evento e C la variabile aleatoria per la censura. Nella pratica è possibile osservare solo $T = \min(X, C)$ e un indice ϵ che indica se è stato possibile osservare l'evento o se si è in presenza di censura, $\epsilon = I(X \leq C)$ (Klein, Moeschberger, 2003). Le funzioni di interesse sono:

- la funzione di rischio, $h(t)$: è una probabilità istantanea condizionata, che rappresenta la probabilità che l'evento si verifichi nell'istante di tempo t dato che un individuo è sopravvissuto fino a t . L'espressione matematica di tale funzione è:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

- la funzione di sopravvivenza, $S(t)$, che è la funzione fondamentale per l'analisi di sopravvivenza, ed è la probabilità che un individuo non abbia

ancora sperimentato l'evento al tempo t . Si ottiene come complemento a 1 della funzione di ripartizione di X :

$$S(t) = P(X > t) = 1 - F(t)$$

In quanto complemento a 1 di una funzione di ripartizione, la funzione di sopravvivenza è non crescente, tale che $S(0) = 1$ e $\lim_{t \rightarrow \infty} S(t) = 0$

Le due funzioni sono legate dalla relazione

$$h(t) = -\frac{d \log(S(t))}{dt}$$

dalla quale si può ottenere la relazione inversa

$$S(t) = \exp \left[- \int_0^t h(u) du \right]$$

3.1.1 Stimatore di Kaplan-Meier

Lo stimatore più utilizzato per la stima di $S(t)$ è lo stimatore di Kaplan-Meier.

Siano t_1, t_2, \dots, t_D i tempi distinti ordinati nei quali si verifica almeno un evento. Ad ogni tempo t_i siano n_i il numero di soggetti a rischio, dove un soggetto si intende a rischio se non è stato censurato e non ha sperimentato l'evento fino all'istante appena precedente t_i , e d_i il numero di eventi osservati. Lo stimatore di Kaplan-Meier è:

$$\hat{S}(t) = \prod_{i: t_i \leq t} \left(1 - \frac{d_i}{n_i} \right) \quad (3.1)$$

dove $\hat{h}_i = \frac{d_i}{n_i}$ è la stima per la funzione di rischio a t_i .

Lo stimatore di Kaplan-Meier è uno stimatore non crescente, continuo a destra e continuo a tratti con salti pari a $\hat{h}_i = \frac{d_i}{n_i}$ in presenza dei tempi in cui si verifica almeno un evento. È inoltre ottenuto come stimatore di massima verosimiglianza, e in quanto tale gode delle proprietà di consistenza, correttezza asintotica, efficienza asintotica e normalità asintotica. In virtù di quest'ultima proprietà è possibile quantificare l'incertezza che accompagna la stima sfruttando gli intervalli di confidenza alla Wald, $\hat{S}(t) \pm z_{1-\alpha/2} \sqrt{\hat{V}(\hat{S}(t))}$. Per ottenere una stima della varianza si può utilizzare la formula di Greenwood:

$$\hat{V}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}.$$

Si noti che lo stimatore di Kaplan-Meier discrimina solo tra eventi di interesse e censure, senza considerare la differente informazione che si ha quando si verifica un rischio competitivo. Rischi competitivi e censura vengono trattati allo stesso modo. Formalizziamo ora il problema di analisi di sopravvivenza in presenza di rischi competitivi.

3.2 Rischi competitivi

Un rischio competitivo è un evento che compete con l'evento di interesse e la cui comparsa preclude l'osservazione dell'evento di interesse. Sia M il numero di rischi competitivi e sia X_j , $j = 1, \dots, M$, il tempo non osservabile in cui si verifica il j -esimo rischio. Per ogni soggetto è possibile osservare $T = \min(X_1, \dots, X_M)$ e un indice ϵ che indica che evento ha sperimentato il soggetto, $\epsilon = j$ se $T = X_j$.

Le quantità di interesse nelle condizioni di rischi competitivi sono:

- la funzione di rischio causa specifica per il rischio j , $h_j(t)$, che rappresenta la probabilità istantanea che un soggetto sperimenti l'evento j al tempo t :

$$h_j(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t, \epsilon = j | T \geq t)}{\Delta t}$$

$$j = 1, \dots, M.$$

- la funzione di incidenza cumulata per il rischio j , $F_j(t)$, che è la probabilità congiunta che un soggetto sperimenti l'evento j entro il tempo t :

$$F_j(t) = P(T \leq t, \epsilon = j) \quad (3.2)$$

$F_j(t)$ non è una funzione di ripartizione propria in quanto $F_j(\infty) = P(\epsilon = j)$. È non decrescente, $F_j(0) = 0$ e $F_j(\infty) < 1$.

Le due funzioni sono unite dalla relazione

$$F_j(t) = \int_0^t S(u-) h_j(u) du$$

dove $S(u) = \exp(-\int_0^u (\sum_{h=1}^M h_j(s)) ds)$ è la funzione di sopravvivenza marginale, viene scritta come $S(u-)$ per indicare che la si considera fino all'istante appena precedente u e rappresenta la probabilità che al tempo u un individuo non abbia sperimentato alcun evento.

3.2.1 Stima della funzione di incidenza cumulata

Siano t_1, t_2, \dots, t_D i tempi distinti ordinati nei quali si verifica uno qualsiasi degli M eventi possibili, e siano $t_1^*, t_2^*, \dots, t_{D_j}^*$ i tempi distinti ordinati nei quali si verifica l'evento di interesse. Ad ogni tempo t_i siano n_i il numero di soggetti a rischio, r_i il numero di soggetti che sperimentano l'evento di interesse e d_i il numero di soggetti che sperimentano uno tra gli eventi competitivi. La censura casuale non viene considerata come un rischio competitivo ma modifica solo n_i . Lo stimatore non parametrico per la funzione di incidenza cumulata è:

$$\hat{F}_j(t) = \begin{cases} 0 & \text{se } t \leq t_1 \\ \sum_{t_i^* \leq t} \left\{ \prod_{k=1}^{i-1} \left[1 - \frac{d_k + r_k}{n_k} \right] \right\} \frac{r_i}{n_i} & \text{se } t_1 \leq t \end{cases}$$

Si noti che può essere scritto come una funzione dello stimatore di Kaplan-Meier $\hat{S}(t_i)$:

$$\hat{F}_j(t) = \sum_{t_i^* \leq t} \hat{S}(t_i^* -) \frac{r_i}{n_i},$$

dove $\hat{S}(t_i)$ viene calcolato utilizzando gli istanti di tempo fino a quello appena precedente t_i e rappresenta la funzione di sopravvivenza che si ottiene considerando tutti i rischi competitivi come un unico evento. Questo stimatore porta a una funzione di incidenza stimata non decrescente e continua a tratti, con salti in corrispondenza dei tempi in cui si verifica l'evento di interesse. La varianza di $\hat{F}(t_i)$ viene calcolata tramite la formula:

$$\hat{V}(F(t)_j) = \sum_{t_i^* \leq t} \hat{S}(t_i^*)^2 \left\{ [F(t) - \hat{F}(t_i^*)]^2 \frac{r_i + d_i}{n_i^2} + [1 - 2(F(t) - \hat{F}(t_i^*))] \frac{r_i}{n_i} \right\}.$$

In presenza di rischi competitivi non è quindi corretto appoggiarsi all'analisi di sopravvivenza condotta tramite Kaplan-Meier. Tale stimatore

tende infatti a sovrastimare la funzione di incidenza cumulata, intesa come complemento a 1 della funzione di sopravvivenza, in quanto il verificarsi di un rischio competitivo viene considerato come una censura. Nei capitoli successivi verrà condotta una meta-analisi per dimostrare e quantificare la sovrastima.

Capitolo 4

I dati: presentazione degli studi e prime analisi esplorative

4.1 I dati

Per quantificare di quanto lo stimatore di Kaplan-Meier sovrastimi l'incidenza cumulata sono stati selezionati 16 studi che presentavano la stima dell'incidenza cumulata calcolata sia tramite il complemento a uno dello stimatore di Kaplan-Meier sia tramite stimatori per la funzione di incidenza cumulata causa-specifica (equazione 3.2). Di questi 16 studi, quattro hanno svolto l'analisi separatamente su due coorti distinte. Nella meta-analisi sono quindi entrati 20 studi.

Per ogni studio sono state rilevate caratteristiche relative al disegno, ossia il tipo di studio, la popolazione di interesse, la numerosità campionaria e il tempo di follow-up medio. È stato rilevato poi quale fosse l'evento di interesse e quanti e quali fossero gli eventi competitivi. Sono stati rilevati il numero di soggetti a rischio all'inizio dello studio e il numero totale di eventi di interesse che si sono verificati durante il periodo di osservazione, la stima dell'incidenza calcolata tramite Kaplan-Meier e la stima dell'incidenza calcolata tramite metodi *ad hoc* per rischi competitivi. Sono stati considerati i valori delle due funzioni corrispondenti all'ultimo tempo di follow-up disponibile.

Gli studi sono stati divisi per area clinica trattata. Sono state considerate tre aree: ossa e legamenti (5 studi), cardiologia (7 studi) e nefrologia (4 studi). Si riportano le caratteristiche degli studi divisi per area clinica nelle Tabelle 4.1, 4.2 e 4.3.

Tabella 4.1: Caratteristiche degli studi in ossa e legamenti

Studio	Caratteristiche	Evento di interesse	Evento competitivo	Tempo di follow-up	Soggetti a rischio	Eventi osservati	Stima KM(%)	Stima CIF(%)	RR
Berry et al.	Studio di coorte; 481 soggetti dal Framingham Heart Study che hanno avuto una prima frattura tra il 1948 e il 2003	Seconda frattura dell'anca	Morte	Mediano: 4.2 anni	481	72	21	12	1.75
Kremers et al.	Studio di coorte; 22864 pazienti sottoposti ad artroplastica primaria totale del ginocchio	Revisione dell'artroplastica	Morte	Medio: 7.8 anni	22864	1640	20	13	1.54
Schuh et al.	Studio di coorte; 247 pazienti con modular endoprosthesis	Complicazioni dell'endoprotesi	Morte	Medio: 86 mesi	247	122	77	65	1.18
Pritchett	Studio di coorte; 537 pazienti sottoposti ad artroplastica totale del ginocchio tra il 1989 e il 1992 per un totale di 639 artroplastiche	Revisione dell'artroplastica	Morte	Medio: 23 anni	639	21	11	6	1.83
Martin et al. (1)	Studio di coorte; 109 Charnley total hip procedures in pazienti con meno di 50 anni	Revisione per fallimento dell'impianto asettico	Morte; Revisione per infezione o rottura	Massimo: 20 anni	109	21	23	19	1.21
Martin et al. (2)	Studio di coorte; 649 Charnley total hip procedures in pazienti con piùdi 50 anni	Revisione per fallimento dell'impianto asettico	Morte; Revisione per infezione o rottura	Massimo: 20 anni	649	33	8	5	1.60

Tabella 4.2: Caratteristiche degli studi in cardiologia

Studio	Caratteristiche	Evento di interesse	Evento competitivo	Tempo di follow-up	Soggetti a rischio	Eventi osservati	Stima KM(%)	Stima CIF(%)	RR
Austin et al.	Studio di coorte; 16237 pazienti ricoverati per insufficienza cardiaca	Morte per cause cardiovascolari	Morte per altre cause	5 anni	16237	5970	43	36.8	1.17
Grunkemeier et al. (2007)	Studio di coorte; 18596 pazienti sottoposti a CABG dal 1986 al 2003	Nuovo intervento CABG	Morte	30 anni	18596	2975(16%)	61	30	2.03
Grunkemeier et al. (2001) (1)	Studio di coorte; 246 pazienti operati per impianto di valvola aortica	Tromboembolismo	Morte	Massimo: 30 anni; Medio: 12 anni	246	80	63	34	1.85
Grunkemeier et al. (2001) (2)	Studio di coorte; 115 pazienti operati per impianto di valvola mitrale	Tromboembolismo	Morte	Massimo: 30 anni; Medio: 11.4 anni	115	47	58	42	1.38
Kampchen et al.	Studio di coorte; 257 pazienti sottoposti a impianto di valvola mitrale tra il 1974 e il 2000	Nuovo intervento	Morte	Massimo: 24 anni	257	27	45	17	2.65
Sobolev et al. (1)	Studio di coorte; 6405 pazienti gravi in lista d'attesa per CABG	Morte prima dell'intervento	Intervento	Massimo: 52 settimane	6405	54	3.7	0.9	4.11
Sobolev et al. (2)	Studio di coorte; 1920 pazienti meno gravi in lista d'attesa per CABG	Morte prima dell'intervento	Intervento	Massimo: 52 settimane	1920	26	3.4	1.5	2.27
Southern et al.	Studio di coorte; 21624 pazienti che si sottopongono a cateterizzazione cardiaca	CABG	Intervento coronarico percutaneo; Morte	1 anno	21624	4502	28.8	20.8	1.38
Wolbers et al.	Studio di coorte; 442 soggetti con ICD	Terapia per ICD	Morte	Mediano: 3.3 anni; Massimo: 5 anni	442	180	51	49	1.04

Tabella 4.3: Caratteristiche degli studi in nefrologia

Studio	Caratteristiche	Evento di interesse	Evento competitivo	Tempo di follow-up	Soggetti a rischio	Eventi osservati	Stima KM (%)	Stima CIF (%)	RR
Alberti et al.	Studio di coorte; 657 destinatari di trapianto renale	Kidney graft loss	Morte	Mediano: 77 mesi	657	116	27.5	26	1.058
Beuscart et al.	Studio di coorte; 383 pazienti sottoposti a dialisi peritoneale	Emodialisi	Morte durante la dialisi peritoneale; Trapianto renale	Mediano: 1.3 anni	383	109	59	32	1.84
Evans et al.	Studio di coorte; 8559 pazienti dal Registre de Dialyse Peritoneale de Langue Francaise	Primo episodio di peritonite	Morte; Emodialisi; Trapianto renale; Recupero funzioni renali	Massimo: 8 anni; Medio: 1.5 mesi	8559	3137	54	36	1.50
Grams et al. (1)	Studio di coorte; 1498 partecipanti	ESRD	Morte	Massimo: 18.9 anni; Medio: 7.2 anni	1498	974	66	54	1.22
Grams et al. (2)	Studio di coorte; 224 partecipanti	ESRD	Morte	Massimo: 18.9 anni; Medio: 7.2 anni	224	125	66	62	1.02

4.2 Analisi esplorative

4.2.1 Analisi univariate

Scopo di questa meta-analisi è stabilire le differenze nella stima dell'incidenza cumulata calcolata con Kaplan-Meier e tramite la funzione di incidenza cumulata. Per questo motivo è stata presa come quantità di interesse il *risk ratio* dato dal rapporto tra queste due quantità:

$$RR = \frac{1 - \hat{S}(t)}{\hat{F}(t)} \quad (4.1)$$

dove $\hat{S}(t)$ e $\hat{F}(t)$ sono, rispettivamente, la quantità in (3.1) e la stima della funzione in (3.2) presentate nel Capitolo 3, calcolate in corrispondenza del massimo tempo di follow-up osservato.

I valori di RR per ogni studio sono riportati nelle Tabelle 4.1, 4.2 e 4.3.

Le analisi sono state svolte tramite il software statistico R, e il codice è riportato in Appendice A, sezione A.1.

Si riportano di seguito le statistiche di sintesi per i *risk ratio*:

Tabella 4.4: Statistiche di sintesi per i *risk ratio*

Min	1st Qu	Mediana	Media	3rd Qu	Max	sd
1.02	1.20	1.52	1.68	1.84	4.11	0.72

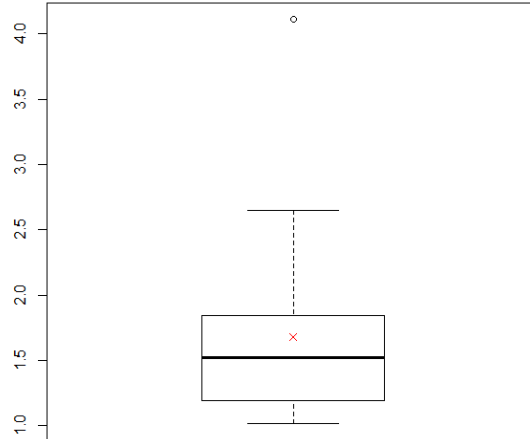


Figura 4.1: Boxplot per i *risk ratio*

La distribuzione dei *risk ratio* appare asimmetrica a destra, anche se lo scarto tra primo quartile e mediana è uguale allo scarto tra mediana e terzo quartile (Figura 4.1). Vi è un unico valore anomalo, al quale corrisponde $RR = 4.11$. Tale valore viene osservato in corrispondenza dello studio di Sobolev *et al.* (2006), uno studio di cardiologia nel quale interessa stimare l'incidenza di morte per pazienti gravi, prima che vengano sottoposti a bypass aorto-coronarico.

Si noti che in nessun caso si osserva un *risk ratio* uguale o inferiore a 1 (il minimo è 1.02 dello studio di Grams *et al.*, 2012), il che indica che in tutti gli studi considerati lo stimatore di Kaplan-Meier sovrastima l'incidenza cumulata.

Si verifica ora se i *risk ratio* ottenuti seguono una distribuzione normale, al fine di valutare se nello svolgimento della meta-analisi ci si può avvalere dell'assunzione di normalità per il calcolo degli stimatori di massima verosimiglianza, $\hat{\Delta}_{ML}^2$, e di massima verosimiglianza ristretta, $\hat{\Delta}_{REML}^2$.

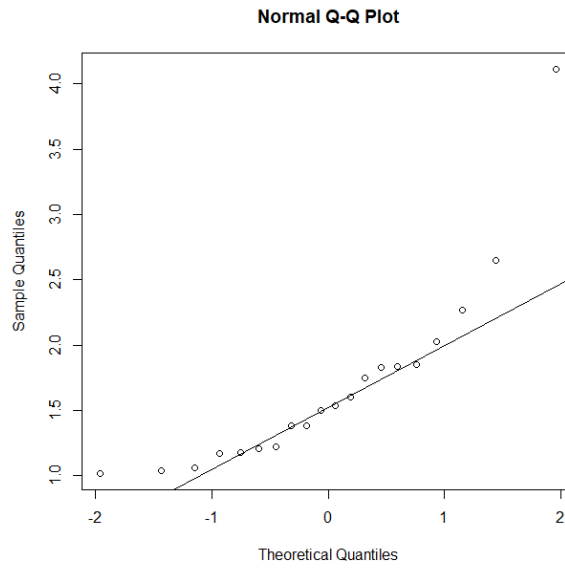


Figura 4.2: Grafico quantile-quantile

Dal grafico quantile-quantile in Figura 4.2 emerge un allontanamento per asimmetria in entrambe le code, che sembra essere dovuto principalmente al valore anomalo osservato. Il p -value associato al test di Shapiro-Wilk per la normalità è $p = 0.000429$. Tuttavia, se si ripete l'analisi escludendo lo studio di Sobolev, si ottiene il grafico in Figura 4.3.

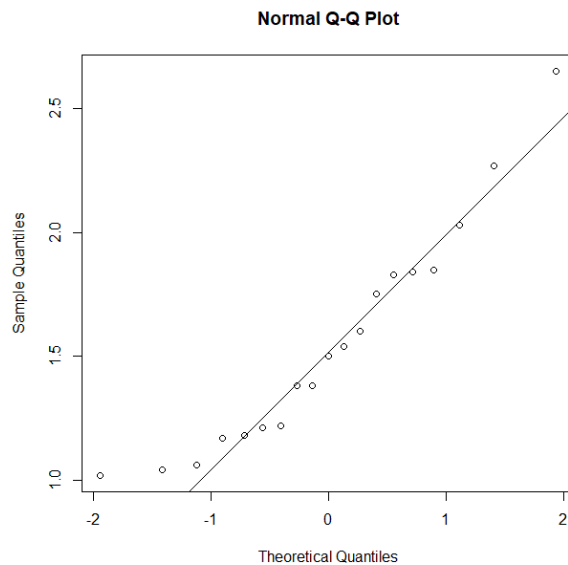


Figura 4.3: Grafico quantile-quantile senza il valore anomalo

In questo caso la retta interquartilica interpola bene i punti del grafico. Il p -value associato al test di Shapiro-Wilk è $p = 0.1367$. La non-normalità è quindi causata dall'unico studio per il quale si è rilevato un valore anomalo. Per questo motivo nel seguito si adotterà la normalità.

Si analizzano ora le caratteristiche degli studi (tipo di studio, numerosità campionaria, tempo di follow-up medio e numero di rischi competitivi) per vedere se da un punto di vista esplorativo sembra esserci omogeneità tra essi. Innanzitutto, dal punto di vista del disegno si può notare che tutti i 20 studi considerati sono studi di coorte.

Si riportano in Tabella 4.5 le statistiche di sintesi per la numerosità campionaria:

Tabella 4.5: Statistiche di sintesi per la numerosità campionaria

Min	1st Qu	Mediana	Media	3rd Qu	Max	sd
109.0	254.5	644.0	5107.6	6943.5	22864.0	7948.4

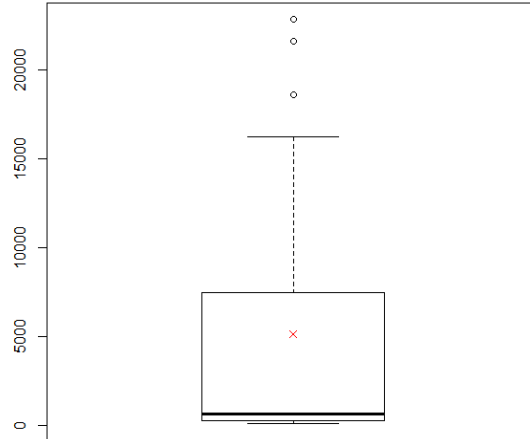


Figura 4.4: Boxplot della numerosità campionaria

Si può notare la presenza di una forte asimmetria a destra: la mediana assume infatti un valore di molto inferiore rispetto a quello assunto dalla media (indicata nel boxplot con una x rossa, Figura 4.4). I tre valori anomali corrispondono agli studi di Grunkemeier *et al.* (2007), di Southern *et al.* (2006) e di Kremers *et al.* (2016), che presentano numerosità campionarie, rispettivamente, di 18596, 21624 e 22864.

Stratificando per area clinica, risulta che gli studi facenti capo all'area clinica nefrologia hanno in media una numerosità campionaria minore (2264.2), mentre quelli svolti in cardiologia hanno la numerosità campionaria media maggiore (7315.8). Gli studi su ossa e legamenti hanno invece una numerosità campionaria media di 4164.8.

Si riportano in Tabella 4.6 le statistiche di sintesi per il tempo di follow-up medio.

Tabella 4.6: Statistiche di sintesi per il tempo di follow-up

Min	1st Qu	Mediana	Media	3rd Qu	Max	sd
1.00	2.8	7.2	9.7	14.0	30.0	8.9

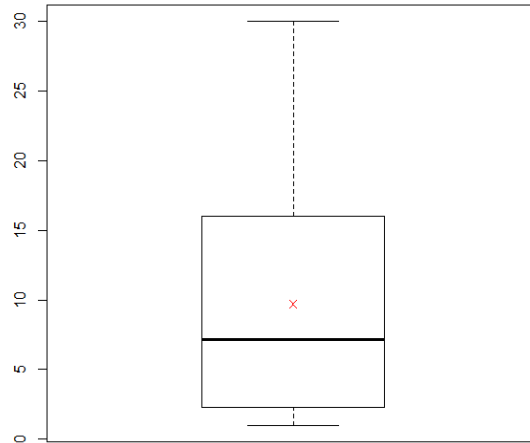


Figura 4.5: Boxplot del tempo di follow-up

Anche in questo caso la distribuzione è asimmetrica a destra, tuttavia l'asimmetria è meno presente rispetto a quanto lo era per la variabile *numerosità campionaria*. Stratificando per area clinica risulta che gli studi relativi alla nefrologia hanno in media un tempo di follow-up medio minore (4.62), mentre quelli relativi a cardiologia e ossa e legamenti hanno tempi di follow-up medi pari rispettivamente a 9.86 e 13.70.

Per quanto riguarda il numero di eventi competitivi, 15 studi ne considerano uno solo, quattro studi ne considerano due e uno (Evans *et al.*, 2010, uno studio in nefrologia in cui si studia l'incidenza del primo episodio di peritonite) ne considera quattro (Figura 4.6).

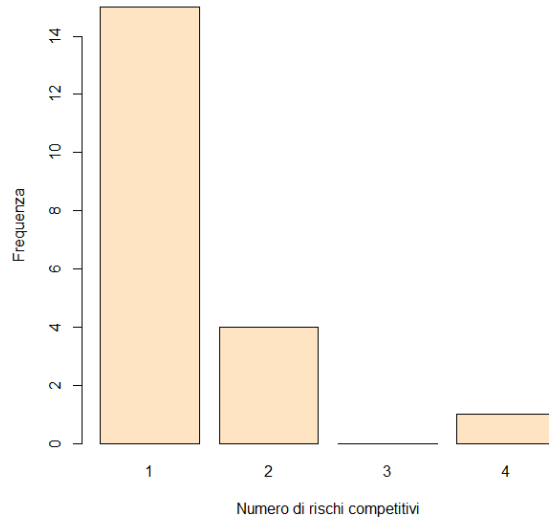


Figura 4.6: Frequenze per il numero di rischi competitivi

La morte è il rischio competitivo per eccellenza, considerato in 19 casi su 20. L'unico studio in cui la morte non rappresenta un rischio competitivo è quello di Sobolev *et al.* (2006), in cui tale evento coincide con l'evento di interesse.

4.2.2 Analisi bivariate

Si analizzano ora i rapporti che sussistono tra *risk ratio* e caratteristiche degli studi, al fine di valutare se e come tali caratteristiche influenzano la portata della sovrastima dello stimatore di Kaplan-Meier.

In Figura 4.7 è riportato il diagramma di dispersione di numerosità campionaria e *risk ratio*. Dal grafico non appaiono andamenti sistematici del valore del *risk ratio* al variare della numerosità campionaria. La correlazione stimata è $\rho = 0.021$ e risulta non significativa (intervallo di confidenza al 5%: $[-0.426, 0.459]$). Il *risk ratio* non risulta quindi influenzato dalla numerosità campionaria.

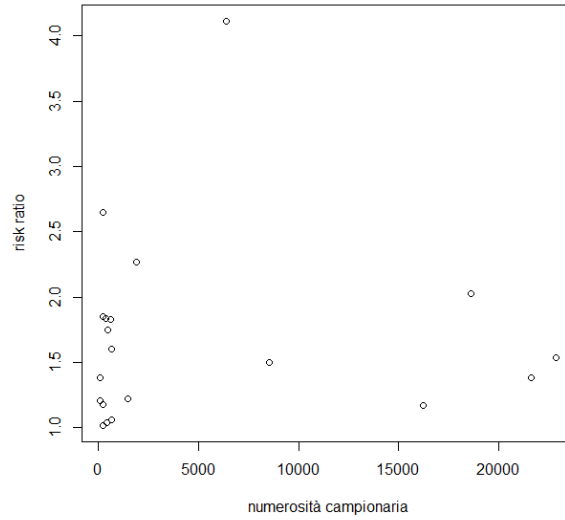


Figura 4.7: Diagramma di dispersione di numerosità campionaria e *risk ratio*

Il diagramma di dispersione tra tempo di follow-up medio e *risk ratio* è riportato in Figura 4.8. Anche in questo caso non si evidenzia alcuna relazione tra le due quantità. La correlazione è pari a $\rho = 0.051$, con intervallo di confidenza $[-0.400, 0.483]$. È quindi non significativamente diversa da zero.

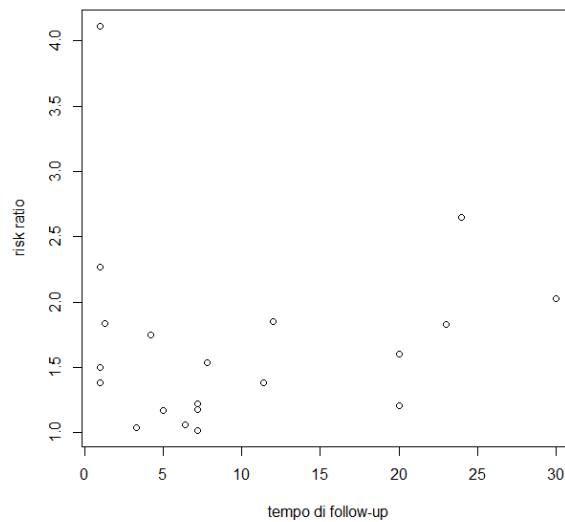


Figura 4.8: Diagramma di dispersione di tempo di follow-up e *risk ratio*

Capitolo 5

Applicazione: meta-analisi per quantificare la sovrastima dello stimatore di Kaplan-Meier

Si procede ora alla meta-analisi al fine di trovare una misura per il *risk ratio* che sintetizzi le 20 misure osservate.

Il primo passo è stabilire come approssimare la varianza dei singoli *risk ratio*, al fine di calcolare le varianze *within* \hat{v}_i , $i = 1, \dots, 20$, necessarie alla stima dell'effetto comune:

$$\hat{\mu} = \frac{\sum_{i=1}^k w_i y_i}{\sum_{i=1}^k w_i},$$

dove w_i può essere $w_{i,fe} = (v_i)^{-1}$ nel caso di un modello a effetto fisso o $w_{i,re} = (v_i + \Delta^2)^{-1}$ nel caso di un modello a effetti casuali.

È stata usata la seguente approssimazione per la varianza del logaritmo del *risk ratio*, proposta da Williamson *et al.* (2002):

$$V(\log(RR)) \simeq \frac{1}{\frac{\text{numero osservato di eventi di interesse}}{1}} - \frac{1}{\text{numero di individui a rischio}} \quad (5.1)$$

La meta-analisi è stata svolta tramite il software statistico **R**, utilizzando la libreria **metafor**. Il codice è riportato in Appendice A, sezione A.2. Per

l'adattamento dei modelli è stata utilizzata la funzione `rma`.

5.1 Modello a effetto fisso

Inizialmente è stato adattato un modello a effetto fisso. Il modello è stato costruito per il logaritmo del *risk ratio*, in modo da poter sfruttare l'approssimazione per le varianze *within*. L'output di R è riportato in Tabella 5.1:

Tabella 5.1: Stima del modello a effetto fisso

	Stima	Std error	z-value	p-value
$\log(\hat{\mu})$	0.3229	0.0058	55.71	<0.001
$\ell(\hat{\mu})$:	-494.6483			
AIC:	991.2965			

L'effetto comune stimato dal modello è $e^{0.3229} = 1.3811$, e un intervallo di confidenza calcolato tramite il metodo Delta partendo da un intervallo di confidenza alla Wald per il logaritmo di tale effetto è $[1.3655, 1.3969]$. Tale intervallo non contiene l'1 che è il valore neutro per un rapporto tra rischi *RR*, indicando che l'incidenza cumulata calcolata tramite Kaplan-Meier è significativamente più alta rispetto a quella calcolata tramite la funzione di incidenza cumulata in (3.2).

Tale modello non è però appropriato: il valore assunto dalla statistica Q di Cochran è $Q = 1063.397$, il che porta ad un *p-value* per il test di omogeneità inferiore allo 0.0001. Il test Q rileva quindi una forte eterogeneità nonostante il numero di studi inclusi nella meta-analisi non sia troppo alto, il che suggerisce che l'eterogeneità sia effettivamente presente. Si adatta quindi un modello a effetti casuali.

5.2 Modello a effetti casuali

Sono stati calcolati sette modelli ad effetti casuali per mettere in luce le differenze nel calcolo della varianza *between* tra gli stimatori presentati nella sezione 1.2.4. Le diverse varianze sono riportate nel grafico in Figura 5.1.

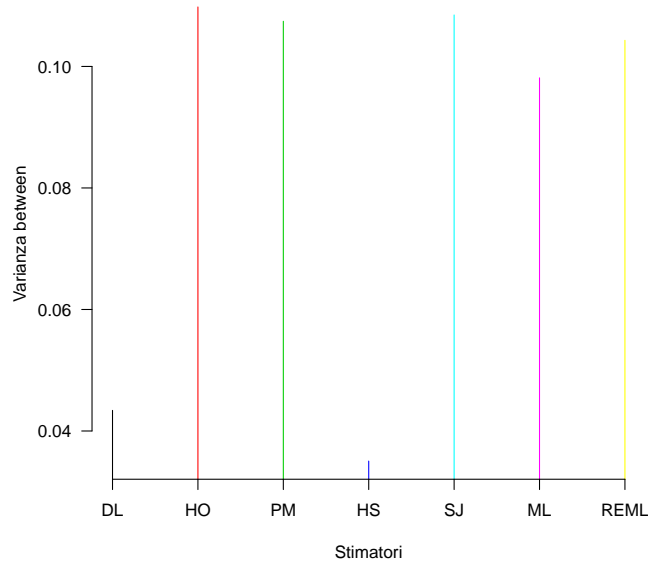


Figura 5.1: Varianza *between* calcolata con i diversi stimatori

Dal grafico in Figura 5.1 emergono due gruppi: il primo contiene gli stimatori DerSimonian-Laird e Hunter-Schmidt mentre il secondo contiene gli stimatori Hedges-Olkin, Paule-Mandel, Sidik-Jonkman, massima verosimiglianza e massima verosimiglianza ristretta.

Il primo gruppo fornisce una stima della varianza Δ^2 di molto più bassa del secondo, infatti $\hat{\Delta}_{DL}^2 = 0.043$ e $\hat{\Delta}_{HS}^2 = 0.035$. Per quanto riguarda invece il secondo gruppo, la varianza *between* è in media 0.106. In particolare, $\hat{\Delta}_{HO}^2 = 0.110$, $\hat{\Delta}_{PM}^2 = 0.107$, $\hat{\Delta}_{SJ}^2 = 0.108$, $\hat{\Delta}_{ML}^2 = 0.098$ e $\hat{\Delta}_{REML}^2 = 0.104$

Il basso valore di Δ^2 calcolato tramite lo stimatore DerSimonian-Laird può essere dovuto al fatto che tale stimatore sottostimi la varianza *between* quando la vera varianza Δ^2 è alta, come suggerisce la forte eterogeneità rilevata tramite il test Q di Cochran. Anche lo stimatore di Hunter-Schmidt porta a risultati distorti in senso negativo, e questo spiega perché $\hat{\Delta}_{HS}^2$ assuma il valore più basso rispetto a tutte le altre stime.

Viene confermato quanto era stato detto nella sezione 1.2.4: lo stimatore di Hedges-Olkin produce la stima della varianza tra studi più alta tra quelle considerate. Assumendo sempre che Δ^2 sia alta, si associa allo stimatore di Sidik-Jonkman un *bias* basso. Per quanto riguarda lo stimatore di massima

verosimiglianza, si può notare come produca la stima più bassa all'interno di questo secondo gruppo di stimatori, ma non viene riscontrata una forte distorsione negativa come ci si aspettava. In questo caso, lo stimatore migliore sembra essere quello di verosimiglianza ristretta: non sottostima la varianza tra studi e inoltre è quello che fornisce risultati migliori se gli studi che entrano a far parte della meta-analisi sono studi di grandi dimensioni. In questo caso, la numerosità campionaria media è 5107, quindi gli studi sono studi di grandi dimensioni.

Si riportano ora gli effetti medi $\hat{\mu}$ per il *risk ratio* calcolati con i diversi stimatori,

$$\hat{\mu} = \frac{\sum_{i=1}^k w_{i,re} y_i}{\sum_{i=1}^k w_{i,re}}$$

con $w_{i,re} = (v_i + \hat{\Delta}^2)^{-1}$ e $\text{var}(\hat{\mu}) = (\sum_{i=1}^k w_{i,re})^{-1}$.

Le diverse stime sono riportate nel grafico in Figura 5.2.

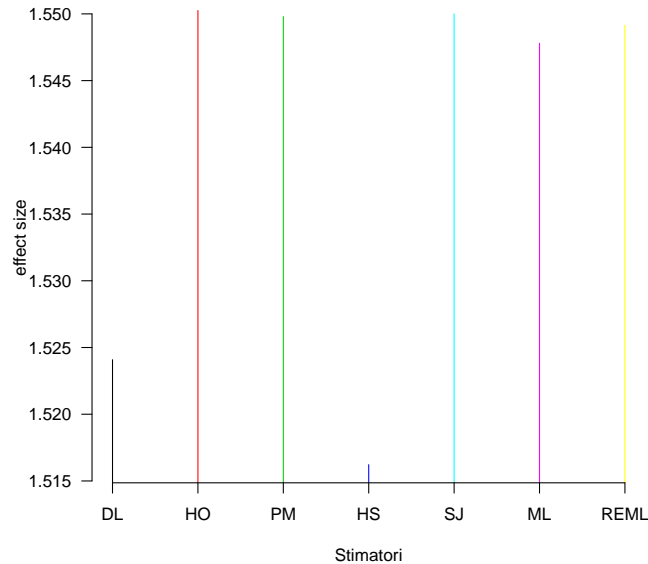


Figura 5.2: Effetto medio calcolato con i diversi stimatori

Si nota lo stesso raggruppamento osservato per le varianze stimate. Agli stimatori DerSimonian-Laird e Hunter-Schmidt corrispondono infatti stime

più basse: in particolare, sfruttando il metodo DerSimonian-Laird si ottiene un effetto medio pari a $\hat{\mu}_{DL} = 1.524$ mentre avvalendosi dello stimatore Hunter-Schmidt si ottiene un effetto medio pari a $\hat{\mu}_{HS} = 1.516$. Per gli altri stimatori, si ottengono invece $\hat{\mu}_{HO} = 1.550$, $\hat{\mu}_{PM} = 1.550$, $\hat{\mu}_{SJ} = 1.550$, $\hat{\mu}_{ML} = 1.548$ e $\hat{\mu}_{REML} = 1.549$.

Si riportano ora gli adattamenti dei modelli a effetti casuali, stimati tramite il software R, ottenuti prima sfruttando il metodo DerSimonian-Laird e poi lo stimatore di massima verosimiglianza ristretta.

Il modello a effetti casuali DerSimonian-Laird ha prodotto il risultato riportato in Tabella 5.2:

Tabella 5.2: Stima del modello a effetti casuali DerSimonian-Laird

	Stima	Std error	z-value	p-value
$\log(\hat{\mu})$	0.4214	0.0516	8.1608	<0.001
$\hat{\Delta}^2$	0.0434	0.0281		
$\ell(\hat{\mu}, \hat{\Delta}^2)$:	-9.4892			
AIC:	22.9783			

Il modello a effetti casuali ottenuto utilizzando lo stimatore di massima verosimiglianza ristretta per la stima della varianza *between* ha prodotto il risultato riportato in Tabella 5.3:

Tabella 5.3: Stima del modello a effetti casuali di massima verosimiglianza ristretta

	Stima	Std error	z-value	p-value
$\log(\hat{\mu})$	0.4377	0.0761	5.7543	<0.001
$\hat{\Delta}^2$	0.1043	0.0373		
$\ell(\hat{\mu}, \hat{\Delta}^2)$:	-6.7923			
AIC:	17.5846			

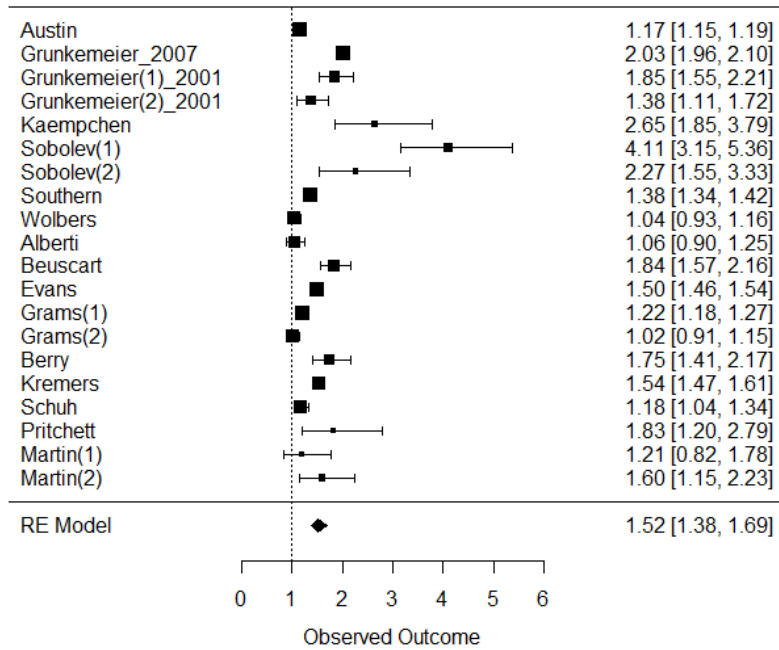
Come già detto, il modello DerSimonian-Laird sottostima la varianza e di conseguenza l'effetto medio rispetto al modello di massima verosimiglianza ristretta.

Per questi due modelli, I^2 , che rappresenta la percentuale di variabilità totale da imputare alla varianza *between*, assume rispettivamente il valore $I_{DL}^2 = 0.98$ e $I_{REML}^2 = 0.99$. La percentuale di variabilità totale dovuta

all'eterogeneità è quindi del 98% se si considera il modello DerSimonian-Laird e del 99% se si considera il modello di massima verosimiglianza ristretta. In entrambi i casi è molto alta. Si riportano in Figura 5.3 i due *forest plot*.

I *forest plot* sono grafici ampiamente usati nel presentare i risultati di una meta-analisi. Riportano l'effetto osservato con il relativo intervallo di confidenza per ogni studio che entra nella meta-analisi e l'effetto comune stimato con il relativo intervallo di confidenza. I quadrati in corrispondenza dei singoli studi rappresentano il contributo dello studio, ossia la precisione dello studio, dato che vengono aggregati in maniera da dare più importanza agli studi cui è associata una maggiore precisione: maggiore è l'area del quadrato, maggiore è la precisione dello studio e quindi maggiore è il contributo nella stima dell'effetto medio. L'effetto comune è rappresentato con un rombo.

Forest plot per modello DerSimonian-Laird



Forest plot per modello Massima Verosimiglianza Ristretta

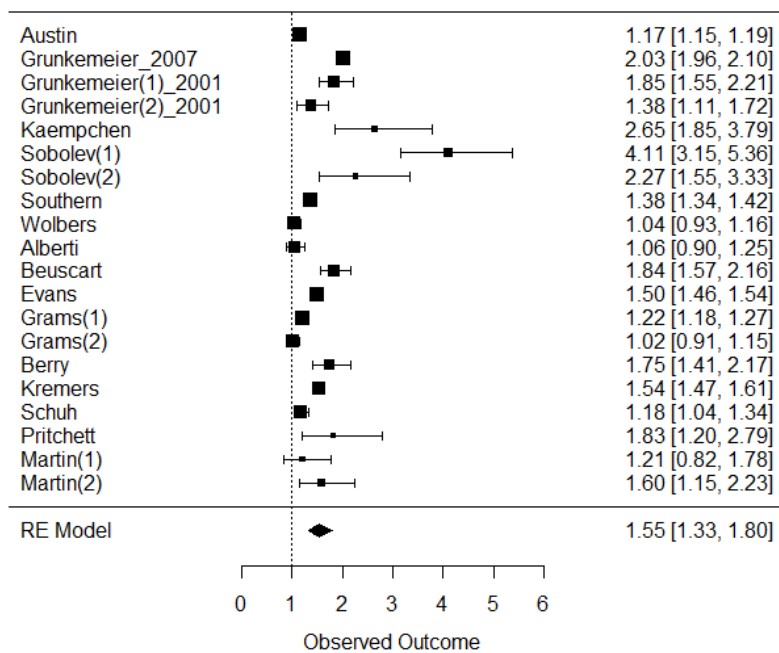


Figura 5.3: Forest plot dei due modelli

Il modello ottenuto con lo stimatore di massima verosimiglianza ristretta produce un valore di log-verosimiglianza maggiore rispetto al modello DerSimonian-Laird, quindi risulta migliore.

Dalla Figura 5.3 risulta come per entrambi i modelli l'intervallo di confidenza dell'effetto osservato non contenga il valore neutro 1 (intervallo di confidenza per DerSimonian-Laird [1.38, 1.69], per REML [1.33, 1.80]). Inoltre, entrambi i modelli forniscono un'indicazione simile riguardo la sovrastima dell'incidenza cumulata calcolata tramite il complemento a 1 dello stimatore di Kaplan-Meier: Kaplan-Meier fornisce stime superiori di più del 50% rispetto ai metodi che si basano sulla stima della funzione di incidenza cumulata.

Stratificando ora l'analisi dividendo gli studi per area clinica si ottengono gli effetti medi e i relativi intervalli di confidenza riportati in Tabella 5.4. Gli intervalli di confidenza sono stati calcolati tramite l'esponentiale dell'intervallo di confidenza alla Wald ottenuto per il logaritmo dell'effetto medio.

Tabella 5.4: Effetto medio calcolato stratificando per area clinica

area clinica	DL	REML
ossa e legamenti	1.473 [1.269, 1.710]	1.474 [1.268, 1.712]
cardiologia	1.749 [1.448, 2.112]	1.785 [1.344, 2.369]
nefrologia	1.295 [1.111, 1.510]	1.295 [1.049, 1.569]

L'area clinica cardiologia è quella che presenta un problema di sovrastima maggiore, pari al 75% se si usa il modello DerSimonian-Laird e pari al 79% se ci si avvale dello stimatore di massima verosimiglianza ristretta. Segue l'area clinica ossa e legamenti, per la quale lo stimatore di Kaplan-Meier sovrastima l'incidenza cumulata del 47% sia per il modello DerSimonian-Laird che per il modello di massima verosimiglianza ristretta. L'area clinica tra quelle considerate per la quale è stata riscontrata una sovrastima minore è la nefrologia, per la quale si ottengono stime del 30% superiori se si usa lo stimatore di Kaplan-Meier.

Anche stratificando per area clinica tutti i *risk ratio* risultano significativamente diversi da 1 a un livello di confidenza del 5%.

Capitolo 6

Conclusioni

È stata condotta una meta-analisi per quantificare di quanto l'analisi di sopravvivenza condotta con Kaplan-Meier sovrastimi l'incidenza cumulata nell'ambito di rischi competitivi. A questo scopo sono stati selezionati 16 studi, per un totale di 20 coorti, condotti in presenza di rischi competitivi, per i quali era riportata la stima dell'incidenza cumulata calcolata sia tramite Kaplan-Meier sia tramite la funzione di incidenza cumulata, e sono state messe a rapporto le due quantità.

Tutti i 20 *risk ratio* sono risultati maggiori di 1, e solo per quattro studi la sovrastima è risultata non significativa, con l'intervallo di confidenza che conteneva il valore neutro 1 (Wolbers *et al.*, 2014; Alberti *et al.*, 2003; Grams *et al.* 2012; Martin *et al.*, 2015).

Non sono state rilevate correlazioni significative tra la portata della sovrastima e la numerosità campionaria e tra la portata della sovrastima e il tempo di follow-up, ossia il modo in cui lo stimatore di Kaplan-Meier sovrastima l'incidenza cumulata non è influenzato dalla dimensione e dalla durata dello studio.

Nello svolgimento della meta-analisi sono stati messi a confronto i risultati ottenuti con gli stimatori DerSimonian-Laird, Hedges-Olkin, Paule-Mandel, Hunter-Schmidt, Sidik-Jonkman, massima verosimiglianza e massima verosimiglianza ristretta. Sono emersi due gruppi: è risultato infatti come gli stimatori DerSimonian-Laird e Hunter-Schmidt producano risultati inferiori rispetto agli altri nella stima della varianza *between* e di conseguenza nella stima dell'effetto medio. Nonostante ciò, tutti gli stimatori hanno portato a una stima di $\hat{\mu}$ superiore ad 1, ad indicare come Kaplan-Meier sovrastimi l'in-

cidenza cumulata in maniera significativa. Sono stati poi scelti due stimatori che fossero rappresentativi dei due gruppi. Per il gruppo che portava a una varianza inferiore è stato scelto lo stimatore DerSimonian-Laird perché è quello più famoso e più usato, mentre per il secondo gruppo è stato scelto lo stimatore di massima verosimiglianza ristretta perché è quello che in generale gode di proprietà migliori. Sono stati adattati i due modelli e sono stati confrontati. Le stime ottenute con lo stimatore $\hat{\Delta}_{REML}^2$ portano a una log-verosimiglianza maggiore, e ciò è un'indicazione a favore di tale modello. Un'altra indicazione che in questo caso porta a preferire $\hat{\Delta}_{REML}^2$ a $\hat{\Delta}_{DL}^2$ è il fatto che il secondo stimatore produca risultati distorti in senso negativo quando la vera varianza tra studi Δ^2 è alta. In questa applicazione il test Q di Cochran per l'omogeneità ha portato a un rifiuto dell'ipotesi nulla con un p -value inferiore allo 0.0001, quindi l'eterogeneità è molto presente e si può assumere che la varianza *between* sia alta. In conclusione quindi, è ritenuto più affidabile il risultato ottenuto con lo stimatore di massima verosimiglianza ristretta.

L'effetto medio risultante è $\hat{\mu}_{DL} = 1.52$ per il modello DerSimonian-Laird e $\hat{\mu}_{REML} = 1.55$ per il modello di massima verosimiglianza ristretta. Entrambi i modelli suggeriscono quindi una sovrastima superiore al 50% per l'incidenza cumulata da parte dello stimatore di Kaplan-Meier in presenza di rischi competitivi. All'effetto calcolato tramite $\hat{\Delta}_{REML}^2$ è associata un'incertezza maggiore evidenziata da un intervallo di confidenza più ampio a causa del fatto che la varianza *between* stimata è maggiore. I due valori ottenuti sono tuttavia molto simili e gli intervalli di confidenza si intersecano, quindi i due modelli non portano a risultati significativamente diversi.

Stratificando gli studi per area clinica trattata, risulta che l'area clinica che porta a una distorsione positiva maggiore è la cardiologia, con un effetto medio stimato pari a $\hat{\mu}_{card} = 1.75$ per lo stimatore DerSimonian-Laird e $\hat{\mu}_{card} = 1.79$ per lo stimatore di massima verosimiglianza ristretta. Segue l'area clinica ossa e legamenti con un effetto medio stimato pari a $\hat{\mu}_{ossa} = 1.47$ per entrambi i modelli, e per ultima la nefrologia, per la quale è stata stimata una sovrastima minore, pari a $\hat{\mu}_{nef} = 1.30$ per entrambi i modelli. L'andamento della sovrastima nelle aree cliniche ricalca quello della numerosità campionaria: infatti all'area cardiologia sono associate una maggiore numerosità campionaria e una maggiore sovrastima, mentre in nefrologia si riscontrano una minore numerosità e un minor errore.

Per concludere, da questa meta-analisi è risultato come, in presenza di rischi competitivi, non sia opportuno procedere ad analisi di sopravvivenza tramite lo stimatore di Kaplan-Meier. Esso infatti considera gli individui che sperimentano il rischio competitivo come censurati, e di conseguenza li tratta come se potessero ancora andare incontro all'evento di interesse, nonostante un evento competitivo sia un evento che per definizione preclude il verificarsi dell'evento di interesse. Ciò porta a una sovrastima dell'incidenza cumulata, intesa come il complemento a 1 della funzione di sopravvivenza. In particolare l'incidenza cumulata viene sovrastimata di più del 50%.

Appendice A

Codice R

A.1 Codice R per le analisi esplorative

A.1.1 Analisi univariate

```
summary(dati$rr)
sd(dati$rr)
boxplot(dati$rr)
points(mean(dati$rr), pch=4, col=2)
mean(dati$rr[dati$area_clinica=="card"])
mean(dati$rr[dati$area_clinica=="nef"])
mean(dati$rr[dati$area_clinica=="ossa"])
qqnorm(dati$rr)
qqline(dati$rr)
qqnorm(dati$rr[-6])
qqline(dati$rr[-6])
shapiro.test(dati$rr)
shapiro.test(dati$rr[-6])

summary(dati$n_rischio)
```

```

sd(dati$n_rischio)
boxplot(dati$n_rischio)
points(mean(dati$n_rischio), pch=4, col=2)
mean(dati$n_rischio[dati$area_clinica=="card"])
mean(dati$n_rischio[dati$area_clinica=="nef"])
mean(dati$n_rischio[dati$area_clinica=="ossa"])

summary(dati$f_up)
sd(dati$f_up)
boxplot(dati$f_up)
points(mean(dati$f_up), pch=4, col=2)
mean(dati$f_up[dati$area_clinica=="card"])
mean(dati$f_up[dati$area_clinica=="nef"])
mean(dati$f_up[dati$area_clinica=="ossa"])

barplot(t(c(15,4,0,1)), ylim=c(0,15), names.arg=c(1,2,3,4),
ylab="Frequenza", xlab="Numero di rischi competitivi", col="bisque")

```

A.1.2 Analisi bivariate

```

plot(dati$n_rischio, dati$rr, ylab="risk ratio", xlab="numerosità
campionaria")
cor(dati$n_rischio, dati$rr)
cor.test(dati$n_rischio, dati$rr)

plot(dati$f_up, dati$rr, ylab="risk ratio", xlab="tempo di follow-up")
cor(dati$f_up, dati$rr)

```

```
cor.test(dati$f_up, dati$rr)
```

A.2 Codice R per la meta-analisi

```
library(metafor)
```

```
dati <- read.table(file.choose(), header=T)
```

```
V <- 1/dati$n_eventi - 1/dati$n_rischio
```

STIMA MODELLO A EFFETTO FISSO

```
fm <- rma(yi=log(dati$rr), vi=V, method="FE")
```

```
summary(fm)
```

```
fm$b
```

```
exp(fm$b)
```

```
cbind(exp(fm$ci.lb), exp(fm$ci.ub))
```

STIMA DELLA VARIANZA BETWEEN

```
stim <- c("DL", "HE", "PM", "HS", "SJ", "ML", "REML")
```

```
Delta2 <- sapply(stim, function(method) {
```

```
  rma(yi=log(dati$rr), vi=V, method = method)$tau2
```

```
})
```

```
plot(y=Delta2, x=1:length(Delta2), type="h", pch=19, axes=F,
```

```
col=1:length(Delta2), xlab="Stimatori", ylab="Varianza between")
```

```
axis(2, las=1)
```

```
axis(1, at = 1:length(Delta2), lab=c("DL", "H0", "PM", "HS", "SJ",
```

```
"ML", "REML"))
```

STIMA EFFECT SIZE

```
muhat <- sapply(stim, function(method) {
  rma(yi=log(dati$rr), vi=V ,method = method)$b
})
exp(muhat)
plot(y=exp(muhat), x=1:length(muhat), type="h", pch=19, axes=F,
col=1:length(muhat), xlab="Stimatori", ylab="effect size")
axis(2, las=1)
axis(1, at = 1:length(muhat), lab=c("DL", "HO", "PM", "HS", "SJ",
"ML", "REML"))
```

ADATTAMENTO MODELLO DERSIMONIAN-LAIRD E MASSIMA VEROSIMIGLIANZA RISTRETTA

```
rm.dl <- rma(yi=log(dati$rr), vi=V, method="DL")
rm.reml <- rma(yi=log(dati$rr), vi=V, method="REML")
summary(rm.dl)
summary(rm.reml)
forest(rm.dl, transf=exp, slab=dati$autore, reflate=1)
title("Forest plot per modello DerSimonian-Laird")
forest(rm.reml, transf=exp, slab=dati$autore, reflate=1)
title("Forest plot per modello Massima Verosimiglianza Ristretta")
```

ANALISI STRATIFICATA PER AREA CLINICA

```
V_c <- 1/dati$n_eventi[dati$area_clinica=="card"] -
1/dati$n_rischio[dati$area_clinica=="card"]
rm.dl_c <- rma(yi=log(dati$rr[dati$area_clinica=="card"]), vi=V_c,
method="DL")
```

```

rm.reml_c <- rma(yi=log(dati$rr)[dati$area_clinica=="card"], vi=V_c,
method="REML")
summary(rm.dl_c)
summary(rm.reml_c)
V_n <- 1/dati$n_eventi[dati$area_clinica=="nef"] -
1/dati$n_rischio[dati$area_clinica=="nef"]
rm.dl_n <- rma(yi=log(dati$rr[dati$area_clinica=="nef"]), vi=V_n,
method="DL")
rm.reml_n <- rma(yi=log(dati$rr)[dati$area_clinica=="nef"], vi=V_n,
method="REML")
summary(rm.dl_n)
summary(rm.reml_n)
V_o <- 1/dati$n_eventi[dati$area_clinica=="ossa"] -
1/dati$n_rischio[dati$area_clinica=="ossa"]
rm.dl_o <- rma(yi=log(dati$rr[dati$area_clinica=="ossa"]), vi=V_o,
method="DL")
rm.reml_o <- rma(yi=log(dati$rr)[dati$area_clinica=="ossa"], vi=V_o,
method="REML")
summary(rm.dl_o)
summary(rm.reml_o)

```


Bibliografia

- [1] Glass, *Primary, Secondary and Meta-Analysis of Research*, Educational Researcher 1976; 5(10):3–8.
- [2] Lacny, Wilson, Clement, Roberts, Faris, Ghali, Marshall, *Kaplan-Meier survival analysis overestimates cumulative incidence of health-related events in competing risk setting: a meta-analysis*, Journal of Clinical Epidemiology 2018; 93:25-35
- [3] Hardy, Thompson, *Detecting and describing heterogeneity in meta-analysis*, Statistics in Medicine 1998; 17:841–856
- [4] Higgins, Thompson, *Quantifying heterogeneity in a meta-analysis*, Statistics in Medicine 2002; 21:1539-1558
- [5] Veroniki, Jackson, Viechtbauer, Bender, Bowden, Knapp, Kuss, Higgins, Langan, Salanti, *Methods to estimate the between-study variance and its uncertainty in meta-analysis*, Research Synthesis Methods 2016; 7:55-79
- [6] DerSimonian, Laird, *Meta-analysis in clinical trials*, Controlled Clinical Trials 1986; 7:177-188
- [7] Jackson, Bowden, Baker, *How does the DerSimonian and Laird procedure for random effects meta-analysis compare with its more efficient but harder*

- to compute counterparts?*, Journal of Statistical Planning and Inference 2010; 140:961-970
- [8] Hedges, Olkin, *Statistical Methods for Meta-Analysis*, Orlando: Academic Press 1985
- [9] Thorlund, Wetterslev, Awad, Thabane, Gluud, *Comparison of statistical inferences from the DerSimonian-Laird and alternative random-effects model meta-analyses-and empirical assessment of 920 Cochrane primary outcome meta-analyses.* , Research Synthesis Methods 2011; 2:238-253
- [10] Sidik, Jonkman, *A comparison of heterogeneity variance estimators in combining results of studies*, Statistics in Medicine 2007; 26:1964-1981
- [11] Paule, Mandel, *Consensus values and weighting factors*, National Institute of Standards and Technology 1982
- [12] Hunter, Schmidt, *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, Thousand Oaks, Calif:Sage 2004
- [13] Viechtbauer, *Bias and efficiency of meta-analytic variance estimators in the random-effects model*, Journal of Educational and Behavioral Statistics 2005; 30:261-293
- [14] Sidik, Jonkman, *Simple heterogeneity variance estimation for meta-analysis*, Journal of the Royal Statistical Society: Seres C: Applied Statistics 2005; 54:367-384
- [15] Hardy, Thompson, *A likelihood approach to meta-analysis with random effects*, Statistics in Medicine 1996; 15:619-629
- [16] Thompson, Sharp, *Explaining heterogeneity in meta-analysis: a comparison of methods*, Statistics in Medicine 1999; 18:2693-2708

- [17] Raudenbush, *Analyzing effect sizes: random-effects models*, The Handbook of Research Synthesis and Meta-Analysis, New York: Russel sage Foundation, 2009; pp 295-315
- [18] Klein, Moeschberger, *Survival analysis, Techniques for Censored and Truncated Data*, Springer, 2003; pp 21-32, 50-57, 92-104, 127-133
- [19] Williamson, Smith, Hutton, Marson, *Aggregate data meta-analysis with time-to-event outcomes*, Statistics in Medicine 2002; 21:3337-3351