

UNIVERSITÀ DEGLI STUDI DI PADOVA  
DIPARTIMENTO DI SCIENZE STATISTICHE  
CORSO DI LAUREA TRIENNALE IN  
STATISTICA PER L'ECONOMIA E L'IMPRESA



RELAZIONE FINALE

**Selezione tra modello ad effetti fissi e modello ad effetti casuali nella meta-analisi**

Relatore Prof.ssa Annamaria Guolo  
Dipartimento di Scienze Statistiche

Laureando: Riccardo Andres Bernabei  
Matricola 2009243

Anno Accademico 2022/2023



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Meta-analisi</b>	<b>3</b>
1.1 Modello ad effetti fissi . . . . .	4
1.2 Modello ad effetti casuali . . . . .	6
1.3 Stima di $\tau^2$ . . . . .	8
<b>2 Metodi di selezione del modello</b>	<b>13</b>
2.1 La statistica $Q$ . . . . .	13
2.2 La statistica $H^2$ . . . . .	14
2.3 La statistica $I^2$ . . . . .	15
2.4 Metodo <i>bootstrap</i> . . . . .	17
2.5 Criterio d'informazione di Akaike . . . . .	17
2.5.1 $AIC_{CEM}$ e $AIC_{REM}$ . . . . .	18
2.5.2 $GAIC_{REM}$ . . . . .	19
<b>3 Studio sperimentale</b>	<b>21</b>
3.1 Risultati . . . . .	23
<b>4 Applicazione ad un dataset reale</b>	<b>35</b>
<b>Conclusioni</b>	<b>38</b>
<b>Appendice</b>	<b>43</b>
.1 Comandi R per la simulazione dei dati . . . . .	43
.2 Comandi R per il calcolo di $\widehat{AIC}_{REM}$ , $\widehat{GAIC}_{REM}$ , $AIC_{CEM}$ . . . . .	46
<b>Bibliografia</b>	<b>53</b>



# Introduzione

La meta-analisi è una metodologia statistica che combina i risultati di diversi studi indipendenti su un determinato argomento per ottenere una stima più accurata di un certo parametro di interesse, al fine di avere una visione più completa e generale dell'argomento in questione.

Questa stima aggregata può essere ottenuta utilizzando diversi modelli statistici, tra cui il modello ad effetti fissi e il modello ad effetti casuali. Nel modello ad effetti fissi, si assume che gli effetti osservati negli studi siano tutti simili e che le differenze tra gli studi siano dovute solo a errori casuali. Nel modello ad effetti casuali, invece, si ammette che gli effetti osservati possano variare e che le differenze tra gli studi siano dovute sia a errori casuali che a differenze sistematiche tra gli studi stessi (Borenstein et al., 2009). Pertanto ciò che distingue questi due modelli è la presenza o meno di una certa eterogeneità tra i veri valori del parametro nei vari studi. Di fondamentale importanza risulta quindi riuscire ad identificare e valutare la significatività di questa eterogeneità, in modo da adattare il modello più corretto tra i due proposti. A questo scopo sono state proposte numerose statistiche test, come la statistica  $Q$ ,  $H^2$ ,  $I^2$  (Higgins & Thompson, 2002; DerSimonian & Laird, 1986) e i più recenti metodi *bootstrap* (Du et al., 2023). Le prime tre dipendono molto dal numero di studi  $k$  inclusi nella meta-analisi, portando ad elevati errori di I° tipo in presenza di valori alti di  $k$  (Higgins & Thompson, 2002; Botella et al., 2006; Borenstein et al., 2009), mentre il metodo *bootstrap* presenta il problema di un elevato costo computazionale e non riesce a controllare l'errore di I° tipo quando la numerosità campionaria degli studi è bassa (Du et al., 2023).

Si propone quindi in questa tesi un nuovo metodo, basato sul criterio d'informazione di Akaike (Akaike, 1974) calcolato per i modelli ad effetti fissi e ad effetti casuali, in modo da poter selezionare quello che presenterà l' $AIC$  più basso come proposto in Yang & H.Y. Kwan (2020). Se per il modello ad effetti fissi questo calcolo risulta essere facilmente derivabile dalla funzione di log-verosimiglianza, per quello ad effetti casuali se ne

deve calcolare una stima, data la presenza del parametro ignoto  $\tau^2$  riferito alla variabilità tra i vari studi, cioè la nostra eterogeneità. Questo parametro viene sostituito nel calcolo da una sua stima  $\hat{\tau}^2$ , che può essere ottenuta in varie maniere ma che in ogni caso presenta una certa distorsione rispetto al valore effettivo. Pertanto ci si concentra sul calcolo di una stima del valore atteso dell'*AIC* per questo modello, che indicheremo con *GAIC*, in modo da ridurre l'impatto della distorsione di  $\hat{\tau}^2$  (Yang & H.Y. Kwan, 2020)

Si proseguirà poi con uno studio sperimentale, effettuato tramite il software R (R Core Team, 2023), al fine di valutare questo metodo con quelli già esistenti e valutarne la capacità di identificare il modello corretto. A partire poi da dei dati reali si calcoleranno delle stime di  $\theta$  e  $\tau^2$  tramite la libreria *metafor* (Viechtbauer, 2010) e si applicheranno i metodi descritti in questa tesi per selezionare il modello corretto nel contesto di un caso reale.

# Capitolo 1

## Meta-analisi

La meta-analisi è diventata sempre più popolare nelle scienze sociali, mediche e nelle scienze naturali, in quanto consente di superare le limitazioni dei singoli studi, spesso caratterizzati da dimensioni campionarie limitate, fluttuazioni casuali o variazioni nel design dello studio. Attraverso la combinazione dei dati provenienti da studi indipendenti, la meta-analisi fornisce una stima aggregata più robusta del parametro di interesse e consente di ottenere una migliore comprensione del fenomeno in oggetto.

Oltre alla stima degli effetti, la meta-analisi può essere utilizzata per esplorare l'eterogeneità tra gli studi, identificare i fattori che contribuiscono a questa eterogeneità tramite una meta-regressione (Cinar et al., 2021) e condurre analisi di sottogruppi al fine di comprendere meglio le differenze tra le popolazioni o le condizioni di studio. Inoltre, la meta-analisi può essere utilizzata per valutare la presenza di bias di pubblicazione o altre forme di bias che possono influenzare i risultati degli studi inclusi (Borenstein et al., 2009).

La scelta tra il modello ad effetti fissi e ad effetti casuali dipende dalla presenza o meno di eterogeneità tra gli studi inclusi nella meta-analisi. Se gli studi non presentano variabili che possono influire sul risultato finale, cioè se gli effetti osservati sono simili tra gli studi, il modello ad effetti fissi può essere appropriato. Se, invece, gli studi presentano eterogeneità, il modello ad effetti casuali tiene conto di queste differenze sistematiche tra gli studi e fornisce stime più adeguate del parametro di interesse.

L'effetto di interesse altro non è che un parametro su cui si vuole trarre delle conclusioni tramite queste analisi. Per ogni studio si calcola una stima di tale parametro, al fine di utilizzare questi valori e tramite una media pesata calcolare una stima basata su tutti i dati in nostro possesso.

Si assegna ad ogni studio un peso tramite criteri oggettivi e ben definiti, come la numerosità del campione, la precisione della stima, ma che dipendono anche dal modello adottato, e questo peso determinerà quanto il singolo studio influisce sul risultato finale. Questi parametri sono basati su varie statistiche di riepilogo che possono essere fornite da uno studio. Per approfondimenti sui diversi tipi di parametri e sul calcolo delle loro stime, varianze e intervalli di confidenza si rimanda a Borenstein et al. (2009); Botella et al. (2006)

Indicheremo con  $y_i$ ,  $i = 1, \dots, k$ , il valore osservato del parametro nell' $i$ -esimo studio, mentre con  $\theta_i$ ,  $i = 1, \dots, k$  il vero valore all'interno dell' $i$ -esimo gruppo.

## 1.1 Modello ad effetti fissi

Quando si effettua una meta-analisi, si assume un modello statistico per descrivere come si distribuisce  $\theta$  all'interno degli studi in esame.

Scegliere un modello anziché un altro porta a delle differenze nel modo in cui si calcola la stima di  $\theta$ , in particolare nel modo in cui si calcolano i pesi da assegnare ad ogni studio.

Un primo modello, il più semplice, è il cosiddetto modello con effetti fissi (abbreviato spesso con CEM, Common Effect Model, o FEM, Fixed Effect Model).

Questo modello semplicemente prevede che il vero valore del parametro di interesse sia sempre lo stesso all'interno di ogni studio (Borenstein et al., 2009, 2010). Pertanto

$$\theta_1 = \theta_2 = \dots = \theta_k = \theta$$

con  $k$  che è il numero di studi considerati nella meta-analisi. Eventuali differenze tra i valori che osservo sono dovuti unicamente ad un errore casuale dovuto dal campione utilizzato.

Quindi ogni valore osservato del parametro è dato da

$$Y_i = \theta_i + \epsilon_i$$

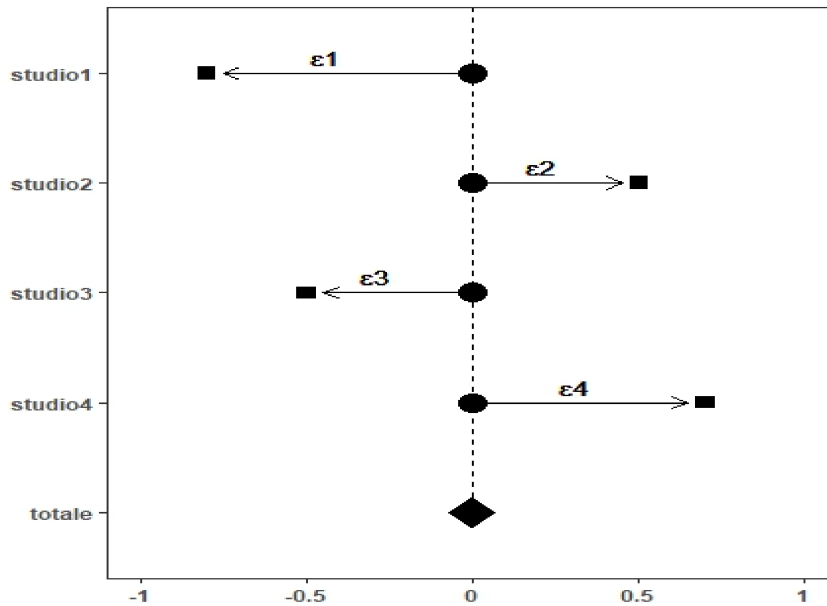
con  $\epsilon_i$  termine di errore distribuito come una  $N(0, \sigma_i^2)$ , indipendenti.

La varianza  $\sigma_i^2$  è la varianza osservata all'interno del singolo studio, unica fonte di variabilità quando si considera questo modello. Può essere stimata con buona precisione,



in particolare quando il campione per il singolo studio è molto numeroso, quindi viene spesso considerata come nota (Borenstein et al., 2009). E' sempre presente dal momento che ogni studio utilizza campioni differenti.

FIGURA 1.1: Distribuzione del vero valore del parametro di interesse in un modello CEM. Il punto tondo indica il vero valore per ogni studio, quello quadrato il valore osservato. Il punto quadrato in basso indica il vero valore di tutto l'insieme dei soggetti degli studi.



Ovviamente noi non conosciamo  $\theta$ , pertanto è di nostro interesse calcolarne una stima

$$\hat{\theta} = \sum_{i=1}^k W_i y_i / \sum_{i=1}^k W_i,$$

con  $W_i$  che è il peso associato ad ogni studio, dato da  $1/\sigma_i^2$ .

Data l'assunzione di normalità dei termini di errore  $\epsilon_i$ , lo stimatore  $\hat{\theta}$  si distribuisce come

$$\hat{\theta} \sim N(\theta, 1/\sum_{i=1}^k W_i).$$

Gli intervalli di confidenza di livello  $(1-\alpha)$  per  $\theta$  si possono calcolare a partire dalla distribuzione di  $\hat{\theta}$ , tramite una procedura alla Wald, come

$$\hat{\theta} \pm z_{1-\alpha/2} \sqrt{V(\hat{\theta})},$$

dove  $z_{1-\alpha/2}$  quantile  $1 - \alpha/2$  di una normale standard e  $V(\hat{\theta}) = 1/\sum_{i=1}^k W_i$ .

Per verificare l'ipotesi  $H_0 : \theta = \theta_{H0}$  contro l'ipotesi alternativa  $H_1 : \theta \neq \theta_{H0}$  si può utilizzare la statistica test

$$Z^{oss} = (\hat{\theta} - \theta_{H0})/\sqrt{V(\hat{\theta})},$$

che sotto l'ipotesi  $H_0$  ha una distribuzione approssimabile a quella di una Normale standard.

Il modello CEM è ragionevole nel momento in cui si può assumere omogeneità degli studi inseriti nella meta-analisi. D'altro canto il limite maggiore di questo modello risiede nel fatto che le conclusioni non possono essere estese a popolazioni differenti (Borenstein et al., 2009), che rende necessaria una modellazione diversa che tenga conto dell'eterogeneità tra gli studi. A questo proposito si fa riferimento ad un modello ad effetti casuali, illustrato nel successivo paragrafo.

## 1.2 Modello ad effetti casuali

A differenza del modello ad effetti fissi, con questo modello si ipotizza che il vero valore del parametro di interesse vari tra i  $k$  studi nella meta-analisi (Borenstein et al., 2009, 2010), e avremo quindi  $\theta_1, \theta_2, \dots, \theta_k$  valori diversi. Pertanto in questo modello non c'è un unico vero valore per il parametro di interesse, ma se ne dovrà calcolare una media a partire da tutti i differenti valori che ottengo dagli studi. La presenza di questa eterogeneità è dovuta a caratteristiche degli studi considerati nella meta-analisi, ad esempio, il metodo di selezione del campione, le caratteristiche dei soggetti inclusi nell'analisi, il tipo di trattamento applicato, il tipo di risposta misurata, ecc...

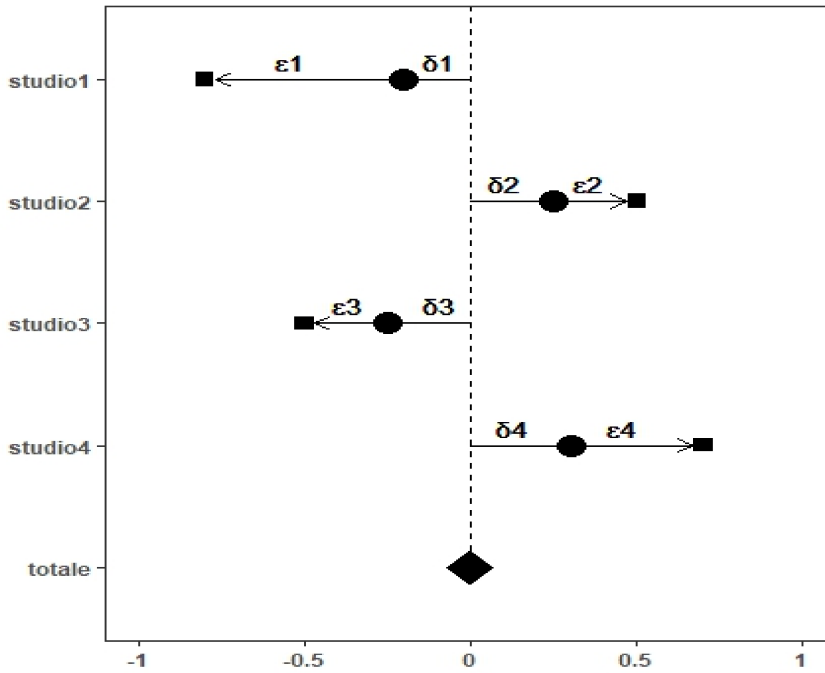
Sotto questo modello ogni valore osservato del parametro è ottenuto dalla variabile

$$Y_i = \mu + \delta_i + \epsilon_i \tag{1.1}$$

dove  $\delta_i \sim N(0, \tau^2)$  è un termine di errore che descrive le differenze dei veri valori del parametro  $\theta_i$  dal valore medio, che indicheremo in questo modello con  $\mu$ . Il termine  $\tau^2$  altro non è che la varianza *tra* i nostri studi, spesso indicata con il termine di eterogeneità.

Il termine  $\epsilon_i$  rimane invece l'errore casuale di ogni studio che spiega la differenza di  $y_i$  da  $\theta_i$ , sempre distribuito come una  $N(0, \sigma_i^2)$ . Il termine  $\sigma_i^2$  può essere inteso come la nostra varianza *dentro* gli studi che, come nel modello CEM, viene assunta nota e pari alla stima della varianza fornita da ogni studio.

FIGURA 1.2: Distribuzione del vero valore del parametro di interesse in un modello REM. Il punto tondo indica il vero valore per ogni studio, quello quadrato il valore osservato. Il punto quadrato in bassa indica il valore medio di tutto l'insieme dei soggetti degli studi.



Nella pratica però, il parametro  $\tau^2$  non è noto e va stimato. Sono stati proposti vari stimatori di  $\tau^2$ , che analizzeremo nel dettaglio più avanti. Intanto indicheremo con  $\hat{\tau}^2$  una generica stima di questo valore. Se  $\hat{\tau}^2 = 0$ , il modello si ridurrà ad un CEM. Se, invece, è presente una certa eterogeneità tra gli studi, si può comunque calcolare una stima di  $\mu$  tramite la formula

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

dove  $W_i^*$  sono i pesi associati ad ogni studio, pari a  $1/(\hat{\tau}^2 + \sigma_i^2)$ .

Dato che le variabili  $Y_i$  sono normalmente distribuite, lo stimatore è anch'esso distribuito normalmente secondo

$$\hat{\mu} \sim N(\mu, 1/\sum_{i=1}^k W_i^*).$$

Gli intervalli di confidenza di livello  $(1-\alpha)$  per  $\mu$  si possono calcolare a partire dalla distribuzione di  $\hat{\mu}$ , tramite una procedura alla Wald, come

$$\hat{\mu} \pm Z_{1-\alpha/2} \sqrt{V(\hat{\mu})},$$

dove  $V(\hat{\mu}) = 1/\sum_{i=1}^k W_i^*$ . Includere anche  $\hat{\tau}^2$  nel calcolo degli intervalli di confidenza porta ad ottenere intervalli più ampi rispetto a quelli del semplice modello CEM.

Per la verifica d'ipotesi  $H_0 : \mu = \mu_{H0}$  contro  $H_1 : \mu \neq \mu_{H0}$  si può utilizzare la statistica test

$$Z^{oss} = (\hat{\mu} - \mu_{H0})/\sqrt{V(\hat{\mu})},$$

che sotto l'ipotesi  $H_0$  ha una distribuzione approssimabile a quella di una Normale standard.

Si utilizza quindi il modello REM quando ipotizzo che nella meta-analisi includo studi simili, ma che comunque presentano certe differenze tra di loro dovute alla presenza di alcune variabili che influiscono sul risultato che ottengo da ogni studio.

Con il modello REM i risultati a cui si giunge possono essere estesi a soggetti non appartenenti al campione utilizzato, aumentandone di molto l'utilità (Borenstein et al., 2009).

### 1.3 Stima di $\tau^2$

La varianza  $\tau^2$  rappresenta tutta la variabilità in più di  $Y_i$  rispetto a quella che ci si aspetterebbe dal semplice errore campionario  $\sigma_i^2$ . Sono state proposte numerose alternative per stimare tale varianza (Lopez-Lopez et al., 2014; Sidik & Jonkman, 2005), data l'elevata influenza della numerosità degli studi sulla precisione di tale stima. L'obiettivo è quello di ottenere uno stimatore che abbia la minor distorsione possibile rispetto al vero valore di  $\tau^2$  e che sia il più efficiente possibile, cioè che abbia un Errore Quadratico Medio (EQM) basso (Sidik & Jonkman, 2007; Viechtbauer, 2005).

- **DerSimonian & Laird (DL)**: Proposto per la prima volta in DerSimonian & Laird (1986), questo approccio per la stima di  $\tau^2$  è basato sul metodo dei momenti, e pertanto presenta il vantaggio di non dover fare assunzioni sulla distribuzione dell'errore  $\delta_i$ .

La stima è ottenuta tramite

$$\hat{\tau}_{DL}^2 = \frac{q - (k - 1)}{\sum_{i=1}^n W_i - \frac{\sum_{i=1}^n W_i^2}{\sum_{i=1}^n W_i}}.$$

Se il valore ottenuto è negativo, si troncherà il valore di  $\hat{\tau}_{DL}^2$  a zero. Troncare il valore dello stimatore se minore di 0 introduce una certa quantità di distorsione positiva, specialmente se  $k$  è piccolo e la varianza dentro gli studi  $\sigma_i^2$  è grande (Viechtbauer, 2005). D'altro canto, più  $\tau^2$  si allontana da zero, più questa distorsione si riduce, dal momento che risulterà meno probabile ottenere delle stime negative. Questo aspetto non influisce solo sulla distorsione dello stimatore DL, ma anche su tutti gli altri metodi che presentano possibili valori negativi di  $\hat{\tau}^2$ .

Questo stimatore presenta il vantaggio di essere facilmente calcolabile e può essere facilmente implementato in qualunque software, caratteristiche importanti negli anni in cui fu sviluppato. Tuttavia,  $\hat{\tau}_{DL}^2$  presenta una distorsione negativa che aumenta all'aumentare di  $\tau^2$ , ma che con elevati valori di  $k$  e della numerosità campionaria degli studi si riduce (Sidik & Jonkman, 2007; Viechtbauer, 2005). L'EQM si riduce all'aumentare di  $k$  e all'aumentare della numerosità campionaria di ogni studio, e per elevati valori di questi parametri risulta essere tra i migliori stimatori che si possono adottare, viceversa, in particolar modo se  $k$  è ridotto,  $\hat{\tau}_{DL}^2$  risulta essere molto inefficiente e comporta un elevato MSE rispetto ad altri metodi.

- **Hedges (HE)**: Stimatore introdotta per la prima volta nel 1954 da Cochran e discusso da Edges 1983, è anch'esso basato sul metodo dei momenti. La stima è ottenuta tramite la formula di Lopez-Lopez et al. (2014)

$$\hat{\tau}_{HE}^2 = \frac{1}{k-1} \sum_{i=1}^k (y_i - \bar{y})^2 - \frac{1}{k} \sum_{i=1}^k \sigma_i^2,$$

con  $\bar{y}$  a indicare la media campionaria, ottenuta tramite

$$\bar{y} = \frac{\sum_{i=1}^k y_i}{k}.$$

Anche per questo stimatore, se il valore ottenuto è negativo si troncherà  $\hat{\tau}_{HE}^2$  a zero.

Le stime ottenute tramite questo metodo presentano una distorsione positiva rispetto al vero valore  $\tau^2$  in quasi tutte le circostanze. Ha i valori più elevati di EQM in quasi tutti i casi, eccetto che quando  $\tau^2$  e  $k$  sono grandi come dimostrato in Sidik & Jonkman (2007).

- **Hunter-Schmidt (HS)**: simile allo stimatore DL, presenta una distorsione negativa ma a livello di EQM presenta buone proprietà, simili a quelle dello stimatore ottenuto tramite la massimizzazione della funzione di verosimiglianza. La stima si ottiene tramite la formula in Viechtbauer (2005)

$$\hat{\tau}_{HS}^2 = \frac{q - k}{\sum_{i=1}^k w_i}.$$

con  $q$  a indicare il valore osservato della statistica  $Q$ , che illustreremo nel capitolo successivo. Se il valore ottenuto è negativo, si troncherà  $\hat{\tau}_{HS}^2$  a zero.

Come detto, presenta una distorsione negativa, che aumenta all'aumentare di  $\tau^2$  ma che diminuisce all'aumentare di  $k$ . Ha un MSE minore di  $\hat{\tau}_{DL}^2$  e di  $\hat{\tau}_{HE}^2$ , specialmente se  $\tau^2$  è grande (Viechtbauer, 2005).

- **Sidik-Jonkman (SJ)**: Introdotto per la prima volta in Sidik & Jonkman (2005), è basato sulla somma pesata del quadrato dei residui, utilizzando come pesi il rapporto  $\sigma_i^2/\tau_0^2$ . Sia  $\tau_0^2$  una stima iniziale per  $\tau^2$ , ottenuta tramite

$$\tau_0^2 = \frac{1}{k} \sum_{i=1}^k (y_i - \bar{y})^2,$$

che utilizzeremo per calcolare i pesi tramite

$$\hat{q}_i = \frac{\sigma_i^2}{\tau_0^2} + 1,$$

da cui ricavo il valore della stima di  $\tau^2$  con la formula

$$\hat{\tau}_{SJ}^2 = \frac{1}{k-1} \sum_{i=1}^k \hat{q}_i^{-1} (y_i - \hat{\mu}_{\hat{q}_i})^2, \quad (1.2)$$

dove  $\hat{\mu}_{\hat{q}_i}$  è la media pesata per  $\hat{q}_i^{-1}$  dei valori osservati  $y_i$ . Non è necessario troncare a 0 il valore della stima ottenuta con la formula (1.2) dato che questa fornisce sempre una stima positiva di  $\tau^2$ .

Presenta una distorsione positiva in tutti i possibili casi, che diminuisce all'aumentare di  $\tau^2$ . In generale risulta lo stimatore con la minor distorsione, ed uno dei migliori in termini di MSE se  $\tau^2$  è grande (Sidik & Jonkman, 2005, 2007).

- **Massima Verosimiglianza (ML):** Gli stimatori precedenti si basavano sul metodo dei momenti, pertanto non si assumevano ipotesi sulla distribuzione sul parametro di interesse  $Y_i$ .

Questo stimatore invece poggia sull'assunzione che

$$\begin{aligned} Y_i &\sim N(\theta_i, \sigma_i^2), \\ \theta_i &\sim N(\mu, \tau^2). \end{aligned}$$

Con questo metodo non si fa altro che massimizzare la funzione di log-verosimiglianza di una variabile casuale Normale di media  $\mu$  e varianza pari a  $\sigma_i^2 + \tau^2$  (Raudenbush, 2009)

$$l(\mu, \tau^2) = -\frac{k}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k \log(\sigma_i^2 + \tau^2) - \frac{1}{2} \sum_{i=1}^k \frac{(y_i - \mu)^2}{(\sigma_i^2 + \tau^2)},$$

con  $\mu \in \mathbb{R}$  e  $\tau^2 > 0$ .

Si ottiene quindi

$$\hat{\mu}(\tau_{ML}^2) = \frac{\sum_{i=1}^k W_i^* y_i}{\sum_{i=1}^k W_i^*}, \quad (1.3)$$

$$\hat{\tau}_{ML}^2 = \max \left\{ 0, \frac{\sum_{i=1}^k W_i^{*2} ((y_i - \hat{\mu}(\tau_{ML}^2))^2 - \sigma_i^2)}{\sum_{i=1}^k W_i^{*2}} \right\}.$$

Il valore della stima  $\hat{\tau}_{ML}^2$  si ottiene tramite un processo iterativo. Si assegna un valore iniziale plausibile per  $\tau^2$ , e a partire da questo si ottiene una prima stima per  $\hat{\mu}(\tau_{ML}^2)$ , che poi si sostituisce nella formula di  $\hat{\tau}_{ML}^2$  aggiornandone il valore. Si itera così finché non si ottiene la convergenza alle stime finali di  $\hat{\tau}_{ML}^2$  e  $\hat{\mu}(\tau_{ML}^2)$ .

La stima  $\hat{\tau}_{ML}^2$  presenta una distorsione negativa in ogni circostanza, distorsione che aumenta in valore all'aumentare di  $\tau^2$ .

Presenta un MSE basso rispetto agli altri metodi di stima, specialmente se il numero di studi  $k$  è molto piccolo (Sidik & Jonkman, 2007). Un problema potrebbe essere il fatto che la convergenza non è sempre garantita, specialmente se la funzione di log-verosimiglianza presenta dei massimi locali.

Nonostante questo metodo sia basato sull'assunzione di normalità di  $\theta_i$ , anche in assenza di questa condizione comunque  $\hat{\tau}_{ML}^2$  mantiene le sue caratteristiche, specialmente in confronto agli altri stimatori di  $\tau^2$  (Kontopantelis & Reeves, 2012b).

- **Log-verosimiglianza ristretta (REML):** Le stime di  $\tau^2$  con questo metodo sono ottenute massimizzando la funzione di log-verosimiglianza ristretta, cioè la log-verosimiglianza marginale dei residui  $r_i = y_i - \hat{\mu}$

$$l_{RE}(\tau^2) = -\frac{k}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^k \log(\sigma_i^2 + \tau^2) - \frac{1}{2} \sum_{i=1}^k \frac{(y_i - \hat{\mu})^2}{\sigma_i^2 + \tau^2} - \frac{1}{2} \log \left( \sum_{i=1}^k \frac{1}{\sigma_i^2 + \tau^2} \right), \quad (1.4)$$

con  $\mu \in \mathbb{R}$  e  $\tau^2 > 0$ , da cui si ottiene (Raudenbush (2009))

$$\hat{\tau}_{REML}^2 = \max \left\{ 0, \frac{\sum_{i=1}^k W_i^{*2} ((y_i - \hat{\mu}(\tau_{ML}^2))^2 + \sigma_i^2)}{\sum_{i=1}^k W_i^{*2}} + \frac{1}{\sum_{i=1}^k W_i^{*2}} \right\}.$$

Anche per questo metodo si ricorre ad una procedura iterativa come nello stimatore basato sulla massima verosimiglianza, assegnando un valore iniziale plausibile a  $\hat{\tau}_{REML}^2$  e aggiornando le stime di  $\hat{\mu}(\hat{\tau}_{REML}^2)$  tramite la formula (1.3) e  $\hat{\tau}_{REML}^2$  ad ogni passo fino alla convergenza.

Questo metodo si utilizza per ovviare alla distorsione dello stimatore di massima verosimiglianza, difatti presenta comunque una distorsione negativa che aumenta all'aumentare di  $\tau^2$ , ma che nel complesso risulta essere minore di quello ottenuto con la stima ML o DL. Si perde parte del vantaggio del metodo ML nel ridurre l'EQM, anche se questo rimane comunque inferiore rispetto al metodo DL, soprattutto nel caso di piccoli valori di  $k$  (Sidik & Jonkman, 2007).

Anche per  $\hat{\tau}_{REML}^2$ , sebbene la distribuzione di  $\theta_i$  presenta un qualche impatto, comunque non altera in modo significativo le performance di questo stimatore, anche quando l'assunzione di normalità di  $\theta_i$  non è rispettata (Kontopantelis & Reeves, 2012a).

Il metodo REML, ML sono preferibili a livello di MSE, ma sono anche quelli che presentano un maggiore distorsione negativa. Questa distorsione porta ad ottenere valori della varianza di  $\hat{\mu}$  inferiori, andando ad aumentare la probabilità di errore di I° tipo nei test di significatività su questo parametro.

Come proposto in Sidik & Jonkman (2007), l'MSE potrebbe essere un criterio non adeguato per la selezione dello stimatore di  $\tau^2$ , e si dovrebbe porre più attenzione sulla distorsione ottenuta. E' consigliabile usare il metodo REML ed ML solamente quando  $\tau^2$  è piccolo, negli altri casi è preferibile usare gli stimatori  $\hat{\tau}_{HE}^2$  e  $\hat{\tau}_{S,J}^2$  che risultano essere quelli con le caratteristiche migliori a livello di distorsione.



# Capitolo 2

## Metodi di selezione del modello

Nella scelta tra modelli ad effetti fissi e modelli ad effetti casuali si possono utilizzare diversi strumenti, di seguito descritti. Questi strumenti sono atti a verificare la presenza o meno di una significativa eterogeneità, generalmente tramite un test di ipotesi che valuta  $H_0 : \tau^2 = 0$  contro  $H_1 : \tau^2 \neq 0$ . Se si accetta  $H_0$  non sembrano esserci differenze significative tra gli studi.

### 2.1 La statistica $Q$

Questa statistica, introdotta in Cochran (1954), fornisce un'indicazione sulla presenza o meno di una varianza tra gli studi, ma non da informazioni su quanto questa sia elevata e su quanto influenzi i risultati.

Questa statistica si ottiene tramite

$$Q = \sum_{i=1}^k W_i (y_i - \hat{\theta}), \quad (2.1)$$

che altro non è che una somma pesata del quadrato degli errori tra il valore osservato nei vari studi  $y_i$  e la stima del vero valore ottenuta sulla base di un modello CEM  $\hat{\theta}$ , con i pesi assegnati in base all'inverso della varianza  $\sigma_i^2$  di ogni studio .

Sotto  $H_0$ , la statistica  $Q$  ha una distribuzione

$$Q \sim \chi_{k-1}^2.$$

Indicando con  $q$  il valore osservato della statistica  $Q$  ottenuta tramite la formula (2.1), possiamo rifiutare l'ipotesi  $H_0$  ad un livello di significatività  $\alpha$  nel momento in cui  $q$  cade nella regione di rifiuto

$$R_\alpha = \{q : q > \chi_{k-1; 1-\alpha}^2\}$$

con  $\chi_{k-1;1-\alpha}^2$  quantile  $1 - \alpha$  di un Chi-quadro con  $k - 1$  gradi di libertà.

Le conclusioni che si traggono tramite questa statistica dipendono molto dal numero di studi inclusi nella meta-analisi. Se  $k$  è grande,  $Q$  è in grado di identificare ogni minima eterogeneità tra i gruppi, eterogeneità che spesso sono così piccole da essere poco influenti. Viceversa, se  $k$  è molto piccolo,  $Q$  porta ad accettare  $H_0$  anche quando in realtà è presente una certa varianza tra i gruppi.

Inoltre se il parametro  $\theta$  non è distribuito come una Normale, l'approssimazione di  $Q$  ad una distribuzione  $\chi^2$  con  $k - 1$  gradi di libertà non è accurata.

## 2.2 La statistica $H^2$

Per ridurre la dipendenza di  $Q$  dal numero di studi, Higgins & Thompson (2002) propose di usare un indice basato su  $\tau^2$  e  $\sigma_i^2$  ed indipendente da  $k$ .

La statistica  $H^2$  descrive in proporzione quanto il valore ottenuto per la statistica  $Q$  è maggiore rispetto al suo valore atteso sotto  $H_0$ . Nello specifico, si ottiene  $H^2$  con la formula

$$H^2 = \frac{Q}{k-1}.$$

Un valore di  $H^2$  pari ad 1 indica l'assenza di eterogeneità tra gli studi. Viceversa, un valore  $> 1$  denota la presenza di una variabilità in eccesso non spiegabile tramite il solo errore campionario, e pertanto da attribuire all'eterogeneità tra i gruppi.

Per questa statistica si possono ottenere anche degli intervalli di confidenza di livello  $(1 - \alpha)$  considerando la trasformata  $\ln(H)$  tramite una procedura alla Wald

$$\ln(H) \pm z_{1-\alpha/2} \sqrt{V(\ln(H))},$$

dove  $\sqrt{V(\ln(H))}$  si ottiene tramite:

- se  $Q > k$ :

$$\sqrt{V(\ln(H))} = \frac{1 \ln(Q) - \ln(k-1)}{2 \sqrt{2Q} - \sqrt{2k-3}},$$

- se  $Q < k$ :

$$\sqrt{V(\ln(H))} = \sqrt{\frac{1}{2(k-2)} \left(1 - \frac{1}{3(k-2)^2}\right)}.$$

L'intervallo di confidenza che otteniamo è per la trasformata  $\ln(H)$ , applicando la funzione inversa ai due estremi ottenuti, possiamo ricavare un intervallo di confidenza per la nostra statistica di interesse  $H^2$ . Se l'intervallo di confidenza per  $H^2$  così calcolato non contiene il valore 1 indica che è presente una certa eterogeneità tra gli studi ad un livello di significatività  $\alpha$ . La varianza di  $\ln(H)$  dipende dal numero di studi e cresce al diminuire di  $k$  (Higgins & Thompson, 2002). Pertanto valutare la presenza o meno di eterogeneità considerando l'intervallo di confidenza di questa statistica presenta gli stessi limiti di  $Q$ , portando ad accettare  $H_0$  anche in presenza di un'elevata eterogeneità se  $k$  è piccolo oppure a rifiutare  $H_0$  anche se questa risulta essere minima e poco influente se  $k$  è grande.

Invece un altro vantaggio di questo indice è che posso confrontare i valori di  $H^2$  ottenuti da meta-analisi che considerano studi con parametri di interesse differenti, anche se l'interpretazione degli intervalli di confidenza ottenuti deve essere svolta con cautela, dal momento che potrebbero non essere direttamente confrontabili tra di loro (Botella et al., 2006; Higgins & Thompson, 2002).

## 2.3 La statistica $I^2$

Questa statistica risulta essere una trasformata di  $H^2$  (Higgins & Thompson, 2002), presenta però una differente ma più semplice interpretazione. La statistica  $I^2$  infatti indica quanto della varianza totale è imputabile alla varianza tra i gruppi  $\tau^2$  e assume valori compresi nell'intervallo  $[0, 1]$ . Si calcola tramite la formula

$$I^2 = \max \left\{ 0, \frac{q - df}{q} \right\}, \quad (2.2)$$

con  $df$  che sono i gradi di libertà della variabile  $\chi^2$  con cui approssimo la distribuzione della statistica  $Q$ , pari a  $k - 1$ . Dato che il numeratore  $q - df$  può assumere valori negativi, si tronca in questi casi il valore di  $I^2$  a 0. Se  $I^2$  ha valori bassi, allora le differenze tra i valori osservati del parametro di interesse sono per lo più dovuti al caso come nel modello CEM. Viceversa, valori alti e prossimi ad 1 di  $I^2$  indicano la presenza di una certa varianza tra gli studi, pertanto è preferibile adottare il modello REM. In Higgins & Thompson (2002) si propongono i valori di 25%, 50%, 75% come soglie di  $I^2$  per indicare la presenza di una bassa, media ed alta eterogeneità.

Posso calcolare tramite una procedura alla Wald un intervallo di confidenza di livello  $(1 - \alpha)$  per questo parametro tramite le seguenti formule (Borenstein et al., 2009):

se  $q > (df + 1)$

$$B = 0.5 \times \frac{\ln(q) - \ln(df)}{\sqrt{2q} - \sqrt{2df - 1}},$$

se  $q < (df + 1)$

$$B = \sqrt{\frac{1}{2(df - 1) \left( 1 - \left( \frac{1}{3(df - 1)^2} \right) \right)}},$$

calcolo poi i valori

$$L = \exp \left[ 0.5 \times \ln\left(\frac{q}{df}\right) - Z_{1-\alpha/2} \times B \right],$$

$$U = \exp \left[ 0.5 \times \ln\left(\frac{q}{df}\right) + Z_{1-\alpha/2} \times B \right],$$

$$C = \sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i},$$

ed infine

$$a = \frac{df(L^2 - 1)}{C},$$

$$b = \frac{df(U^2 - 1)}{C},$$

con  $a$  che risulta essere l'estremo inferiore mentre  $b$  l'estremo superiore di questo intervallo. Si può usare questo intervallo di confidenza per valutare la presenza di eterogeneità o meno, infatti se l'intervallo  $(a, b)$  non contiene il valore 0 vuol dire che è presente una certa eterogeneità tra gli studi con un livello di significatività  $\alpha$ .

L'utilizzo dell'intervallo di confidenza di  $I^2$  per valutare la presenza o meno di eterogeneità soffre degli stessi problemi della statistica  $Q$  e  $H^2$ , cioè è in grado di identificare come significativa ogni minima eterogeneità se  $k$  ha valori grandi e invece porta ad accettare  $H_0$  anche se risulta esserci una elevata eterogeneità se  $k$  è piccolo (Botella et al., 2006; Higgins & Thompson, 2002).

Anche tramite questo indice posso confrontare i valori di  $I^2$  ottenuti da meta-analisi che considerano studi con parametri di interesse differenti, ma di nuovo l'interpretazione degli intervalli di confidenza ottenuti deve essere svolta con cautela, dal momento che potrebbero non essere direttamente confrontabili tra di loro (Botella et al., 2006; Higgins & Thompson, 2002).

## 2.4 Metodo *bootstrap*

Il metodo *bootstrap* è un metodo computazionalmente intensivo basato sulla simulazione che permette di condurre inferenza anche per strutture complesse di dati (Efron & Tibshirani, 1986). Nello specifico si utilizza una verifica d'ipotesi del tipo  $H_0 : \tau^2 = \lambda$  contro l'ipotesi alternativa  $H_1 : \tau^2 > \lambda$  (Du et al., 2023), dove  $\lambda$  è una soglia fissata a piacere.

Sotto l'ipotesi  $H_0$  vengono creati  $B$  campioni *bootstrap* (Du et al., 2023) con reinserimento a partire dai dati osservati  $y_1, \dots, y_k$ . Sia  $y_1^b, \dots, y_k^b$ ,  $b = 1, \dots, B$  un campione *bootstrap*, per ognuno di questi calcolo il valore del log-rapporto di verosimiglianza  $LR^b$ , calcolato con la formula (D.O. Stram, 1994)

$$LR = -2 [l_{RE}(\tau^2 = 0) - l_{RE}(\hat{\tau}^2)],$$

dove  $l_{RE}(\tau^2 = 0)$  è la log-verosimiglianza ristretta descritta in (2.5) calcolata quando  $\tau^2 = 0$ . Invece  $l_{RE}(\hat{\tau}^2)$  è sempre la log-verosimiglianza ristretta calcolata utilizzando una stima di  $\tau^2$  ottenuta a partire dai dati simulati.

Si calcola poi il quantile  $1 - \alpha$  della distribuzione di tutti gli  $LR^b$  ottenuti e il valore  $LR^0$  ottenuto a partire dai dati osservati. Se  $LR^0 > LR_{1-\alpha}$  si rifiuta  $H_0$  ad un livello di significatività  $\alpha$  (Du et al., 2023). Si può calcolare anche un livello di significatività osservato  $\alpha^{oss}$  calcolando il quantile di  $LR^0$  se incluso nella distribuzione degli  $LR^b$  simulati.

Questo metodo risulta essere troppo conservativo quando il numero di studi  $k$  è piccolo e quando si utilizzando alcuni parametri in particolare, per esempio il log-odds ratio, portando ad un errore di I° tipo minore dei livelli nominali (Du et al., 2023).

## 2.5 Criterio d'informazione di Akaike

Il criterio d'informazione di Akaike, che abbrevieremo d'ora in avanti con *AIC*, è stato fin dalla sua introduzione in Akaike (1974) ampiamente utilizzato come strumento per la selezione tra modelli statistici. La sua particolarità è data dal fatto che a differenza di classici test d'ipotesi, permette di confrontare tra loro anche modelli non annidati. L'idea alla base di *AIC* è quella di massimizzare la funzione di log-verosimiglianza attesa, espressa tramite

$$E_{X, \hat{\beta}}(\ln f(X|\hat{\beta})) = E_{\hat{\beta}}(\int f(x|\beta) \ln f(x|\hat{\beta}) dx),$$

con  $X$  variabile casuale avente la funzione di probabilità  $f(x|\beta)$  e con  $\hat{\beta}$  stima di massima verosimiglianza del vettore di parametri  $\beta$ .

In Akaike (1974) si propone di calcolare il valore

$$AIC = 2p - 2 \sum_{i=1}^n \ln f(x_i|\hat{\beta}), \quad (2.3)$$

con  $n$  ad indicare il numero di osservazioni della variabile  $X$ , mentre  $p$  è il numero di parametri inclusi nel modello. Il valore di  $AIC$  risulta facilmente calcolabile e non utilizza parametri ignoti, e sotto condizioni di regolarità è stato dimostrato in Akaike (1974) che la quantità  $-AIC/2n$  è un'approssimazione accurata di  $E_{X,\hat{\beta}}(\ln f(X|\hat{\beta}))$ .

Si calcola questo valore per tutti i vari modelli proposti, e poi si seleziona quello che presenta un  $AIC$  minore.

### 2.5.1 $AIC_{CEM}$ e $AIC_{REM}$

In Yang & H.Y. Kwan (2020) si propone di utilizzare il criterio di Akaike anche per la selezione tra modello CEM e REM nell'ambito di una meta-analisi. In particolare si conosce la funzione di log-verosimiglianza del modello ad effetti fissi, data da

$$l(\theta) = -\frac{1}{2} \sum_{i=1}^k \ln(2\pi\sigma_i^2) - \frac{1}{2} \sum_{i=1}^k \frac{(y_i - \theta)^2}{\sigma_i^2}, \quad (2.4)$$

con  $\theta \in \mathbb{R}$ .

Mentre per un modello ad effetti casuali si ha

$$l(\mu, \tau^2) = -\frac{1}{2} \sum_{i=1}^k \ln\{2\pi(\sigma_i^2 + \tau^2)\} - \frac{1}{2} \sum_{i=1}^k \frac{(y_i - \mu)^2}{\sigma_i^2 + \tau^2}, \quad (2.5)$$

con  $\mu \in \mathbb{R}$ , e  $\tau^2 > 0$ .

utilizzando la formula (2.3) possiamo calcolare i valori  $AIC_{REM}$  e  $AIC_{CEM}$ .

Partendo dal modello CEM, dal momento che si considera  $\sigma_i^2$  noto, stimiamo unicamente  $\hat{\theta}$  e quindi avremo che  $p = 1$ . utilizzando la log-verosimiglianza in (2.4) otteniamo

$$AIC_{CEM} = 2 + \sum_{i=1}^k \left\{ \ln(2\pi\sigma_i^2) + \frac{(y_i - \hat{\theta})^2}{\sigma_i^2} \right\}. \quad (2.6)$$

Per  $AIC_{REM}$  invece, partendo dall'ipotesi che  $\tau^2$  sia noto e che pertanto l'unico parametro da stimare sia  $\hat{\mu}$ , otteniamo utilizzando la log-verosiglianza (2.5) la formula

$$AIC_{REM} = 2 + \sum_{i=1}^k \ln\{2\pi(\sigma_i^2 + \tau^2)\} + \sum_{i=1}^k \frac{(y_i - \hat{\mu})^2}{\sigma_i^2 + \tau^2},$$

se si utilizza invece una stima  $\hat{\tau}^2$ , si può riscrivere la formula come

$$\widehat{AIC}_{REM} = 2 + \sum_{i=1}^k \ln\{2\pi(\sigma_i^2 + \hat{\tau}^2)\} + \sum_{i=1}^k \frac{(y_i - \hat{\mu})^2}{\sigma_i^2 + \hat{\tau}^2}, \quad (2.7)$$

con  $\widehat{AIC}_{REM}$  che è il valore osservato di  $AIC_{REM}$ .

Si selezionerà il modello CEM se  $AIC_{CEM}$  risulta minore di  $AIC_{REM}$ , al contrario, si selezionerà il modello REM se  $AIC_{REM}$  risulterà essere inferiore ad  $AIC_{CEM}$ .

### 2.5.2 $\widehat{GAIC}_{REM}$

Le stime di  $\tau^2$  presentano un certo errore di stima, che varia in base al tipo di stimatore, il numero di studi  $k$  e il valore di  $\tau^2$  ma che comunque risulta essere sempre presente.

Si propone in Yang & H.Y. Kwan (2020) di calcolare a partire dai valori osservati il valore atteso  $E(\widehat{AIC}_{REM})$ , che indicheremo con  $\widehat{GAIC}_{REM}$ , per ridurre l'effetto dell'errore dovuto all'uso della stima  $\hat{\tau}^2$ .

Nello specifico, possiamo notare come  $\sum_{i=1}^k \frac{(y_i - \hat{\mu})^2}{\sigma_i^2 + \tau^2}$  segue una distribuzione  $\chi^2$  con  $k - 1$  gradi di libertà (Hartung, 1999), pertanto si ottiene la formula

$$\widehat{GAIC}_{REM} = E(\widehat{AIC}_{REM}) = k + 1 + \sum_{i=1}^k \ln\{2\pi(\sigma_i^2 + \hat{\tau}^2)\}. \quad (2.8)$$

Questo valore fornisce una stima più stabile e precisa di  $AIC_{REM}$  rispetto ad  $\widehat{AIC}_{REM}$ , come dimostrato in Yang & H.Y. Kwan (2020), specialmente se si hanno pochi studi  $k$  e se la numerosità campionaria degli studi è elevata. Inoltre  $\widehat{GAIC}_{REM}$  fornisce una stima più precisa di  $AIC_{REM}$  anche quando le varianze  $\sigma_i^2$  sono grandi e  $\tau^2$  presenta valori non troppo elevati.





## Capitolo 3

# Studio sperimentale

Per valutare l'efficacia del nuovo metodo proposto basato sul calcolo di  $AIC_{CEM}$  e  $AIC_{REM}$  rispetto ai metodi più tradizionali, come  $Q$ ,  $I^2$ , presentiamo uno studio di simulazione in cui applichiamo a dei dati simulati a partire da valori fissati dei parametri  $k$ ,  $\tau^2$  e  $\theta$  le procedure precedentemente descritte.

I dati saranno simulati a partire dal modello (1.1), considerando differenti scenari sulla base di griglie per i valori di  $k$ ,  $\tau^2$  e  $\theta$ . I metodi di selezione saranno confrontati tra loro in base alla performance nei vari scenari, valutando la proporzione di identificazioni corrette del modello generatore dei dati, sulla base di  $B$  replicazioni dello studio sperimentale.

Utilizzeremo come parametro di interesse la differenza standardizzata tra le medie di un gruppo di trattamento e un gruppo di controllo, data da

$$\delta = \frac{\mu_1 - \mu_2}{\sigma},$$

con  $\mu_1$ ,  $\mu_2$  le medie del gruppo a cui applico il trattamento e quello di controllo, e  $\sigma_1 = \sigma_2 = \sigma$  la deviazione standard delle osservazioni nei due gruppi.

Una stima non distorta di  $\delta$  si può ottenere come (Hedges & Olkin, 1985)

$$d = \frac{\bar{y}_1 - \bar{y}_2}{\hat{S}} * J, \tag{3.1}$$

dove

$$J = 1 - \frac{3}{4(n_1 + n_2) - 9},$$
$$\hat{S} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}.$$

La varianza è data da

$$\sigma_d^2 = \left( \frac{n_1 + n_2}{n_1 * n_2} + \frac{g^2}{2(n_1 + n_2)} \right) * J^2. \quad (3.2)$$

La simulazione considera i seguenti valori dei parametri, ispirati a studi meta-analisi su trattamenti clinici in psicologia, come riportato in Rubio-Aparicio (2018):

- dimensione campionaria  $k$ ,  $k \in \{10, 15, 25, 40, 70\}$
- varianza tra gli studi  $\tau^2$ ,  $\tau^2 \in \{0, 0.05, 0.1, 0.15, 0.2, 1\}$
- parametro di interesse  $\theta$ ,  $\theta \in \{0, 0.25, 0.5, 0.75, 1\}$ .

Per ogni possibile combinazione dei parametri, si considerano  $B = 1000$  replicazioni dal modello (1.1), secondo il seguente algoritmo. Nello specifico, simuliamo prima  $\theta_{i,1}$  e  $\theta_{i,2}$ ,  $i = 1, \dots, k$  da un  $N(\theta, \tau^2)$  i primi e da una  $N(0, \tau^2)$  i secondi, realizzazioni delle variabili casuali date dagli effetti casuali. Si simulano inoltre i valori delle varianze dentro i due gruppi,  $\sigma_{i,1}^2$ ,  $\sigma_{i,2}^2$ , da una  $U(0.005, 1)$ , e della numerosità campionaria  $n_{i,1}$ ,  $n_{i,2}$  da una  $U(15, 90)$ , con 15, 90 che sono i valori minimi e massimi della mediana della numerosità campionaria calcolata per ognuna delle 54 meta-analisi considerate in Rubio-Aparicio (2018). In seguito si simulano per ognuno dei  $k$  studi  $n_{i,1}$  osservazioni del gruppo dei trattati e  $n_{i,2}$  osservazioni di quello di controllo da una  $N(\theta_{i,1}, \sigma_{i,1}^2)$  e da una  $N(\theta_{i,2}, \sigma_{i,2}^2)$  rispettivamente, e tramite questi dati simulati calcolo prima per ogni studio il valore di  $\bar{y}_1$ ,  $\bar{y}_2$ ,  $S_1^2$  e  $S_2^2$  che userò per calcolare  $d$  e  $\sigma_d^2$  tramite le formule (3.1), (3.2). Per ognuna delle  $b$  simulazioni calcolo il valore della statistica  $Q$ ,  $I^2$ ,  $AIC_{CEM}$ ,  $\widehat{AIC}_{REM}$  e  $\widehat{GAIC}_{REM}$ , dividendo quest'ultimi due metodi in base allo stimatore usato per  $\tau^2$ . Infine per ognuno dei possibili scenari riporto per ogni test la proporzione tra le  $b$  simulazioni che hanno portato ad accettare l'ipotesi di un modello ad effetti fissi per il parametro di interesse rispetto al totale delle simulazioni, oltre che il valore medio per la statistica  $Q$ ,  $I^2$ ,  $AIC_{CEM}$ ,  $\widehat{AIC}_{REM}$  e  $\widehat{GAIC}_{REM}$ .

### 3.1 Risultati

Per una maggiore comprensione, indicheremo con  $\widehat{GAIC}_{REM,DL}$ ,  $\widehat{GAIC}_{REM,HE}$ ,  $\widehat{GAIC}_{REM,SJ}$  e  $\widehat{GAIC}_{REM,REML}$  le stime per  $\widehat{GAIC}_{REM}$  ottenute tramite  $\hat{\tau}_{DL}^2$ ,  $\hat{\tau}_{HE}^2$ ,  $\hat{\tau}_{SJ}^2$  e  $\hat{\tau}_{REML}^2$  rispettivamente. Allo stesso modo, indicheremo con  $\widehat{AIC}_{REM,DL}$ ,  $\widehat{AIC}_{REM,HE}$ ,  $\widehat{AIC}_{REM,SJ}$  e  $\widehat{AIC}_{REM,REML}$  le stime per  $\widehat{AIC}_{REM}$  ottenute a partire da  $\hat{\tau}_{DL}^2$ ,  $\hat{\tau}_{HE}^2$ ,  $\hat{\tau}_{SJ}^2$  e  $\hat{\tau}_{REML}^2$ . Ci concentriamo inizialmente nel caso in cui  $\tau^2 = 0$ , cioè quando i valori del parametro di interesse nei diversi studi sono generati da un modello ad effetti fissi. Nella Figura 3.1 si confrontano le proporzioni tra il numero di volte in cui il confronto tra  $AIC_{CEM}$  e  $\widehat{GAIC}_{REM}$  ha portato ad accettare l'ipotesi  $H_0$  di assenza di eterogeneità rispetto al totale delle  $b$  simulazioni, con i valori di  $\widehat{GAIC}_{REM}$  ottenuti tramite l'utilizzo di diversi stimatori per  $\tau^2$ .

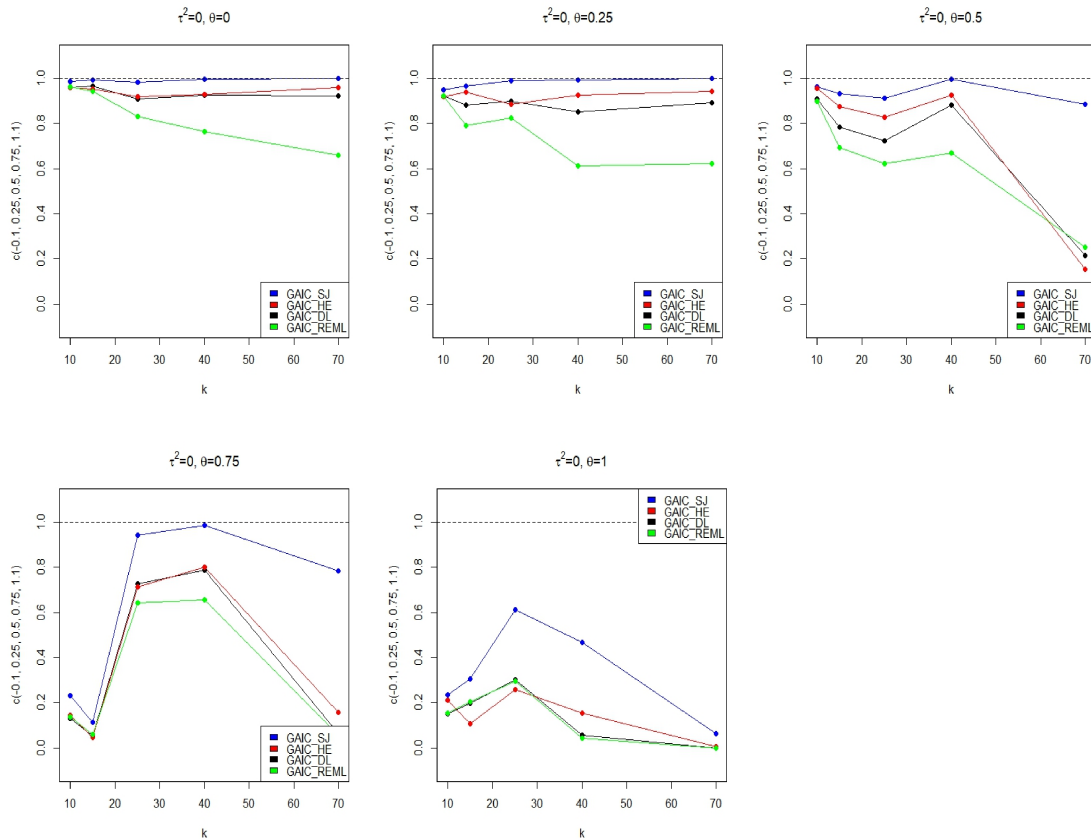


FIGURA 3.1: In ordinata le proporzioni tra le  $B$  simulazioni dove il metodo di selezione del modello tramite l'utilizzo dell' $AIC$  ha portato ad accettare l'ipotesi di assenza di eterogeneità, in ascissa il valore del numero di studi inclusi nella meta-analisi  $k$ . I diversi colori indicano i diversi stimatori per  $\tau^2$  utilizzati. I dati sono stati simulati ponendo  $\tau^2 = 0$ .

Come si può osservare dalla Figura 3.1, l'utilizzo della stima  $\hat{\tau}_{S,J}^2$  nel calcolo di  $\widehat{GAIC}_{REM}$  porta a selezionare il modello corretto un numero maggiore di volte rispetto agli stessi valori ottenuti utilizzando le stime  $\hat{\tau}_{DL}^2$ ,  $\hat{\tau}_{HE}^2$ ,  $\hat{\tau}_{REML}^2$  a prescindere dallo scenario analizzato. L'utilizzo di  $\widehat{GAIC}_{REM,SJ}$  porta a selezionare correttamente in quasi tutti i casi il modello ad effetti fissi quando il parametro di interesse  $\theta = 0$  o presenta valori prossimi a questa soglia, mentre tutti e quattro i metodi peggiorano quando  $\theta$  tende ad 1. Un aumento di  $k$  porta a migliori performance, ma a partire da  $\theta = 0.5$  queste all'aumentare di  $k$  inizialmente migliorano, per poi peggiorare per  $k = 70$ , portandoci a preferire un numero maggiore di volte il modello ad effetti casuali.

Nella Figura 3.2 si confrontano le proporzioni tra il numero di modelli ad effetti fissi correttamente individuati rispetto alle  $B$  simulazioni totali, per  $\widehat{GAIC}_{REM,SJ}$  e la statistica  $Q$ .

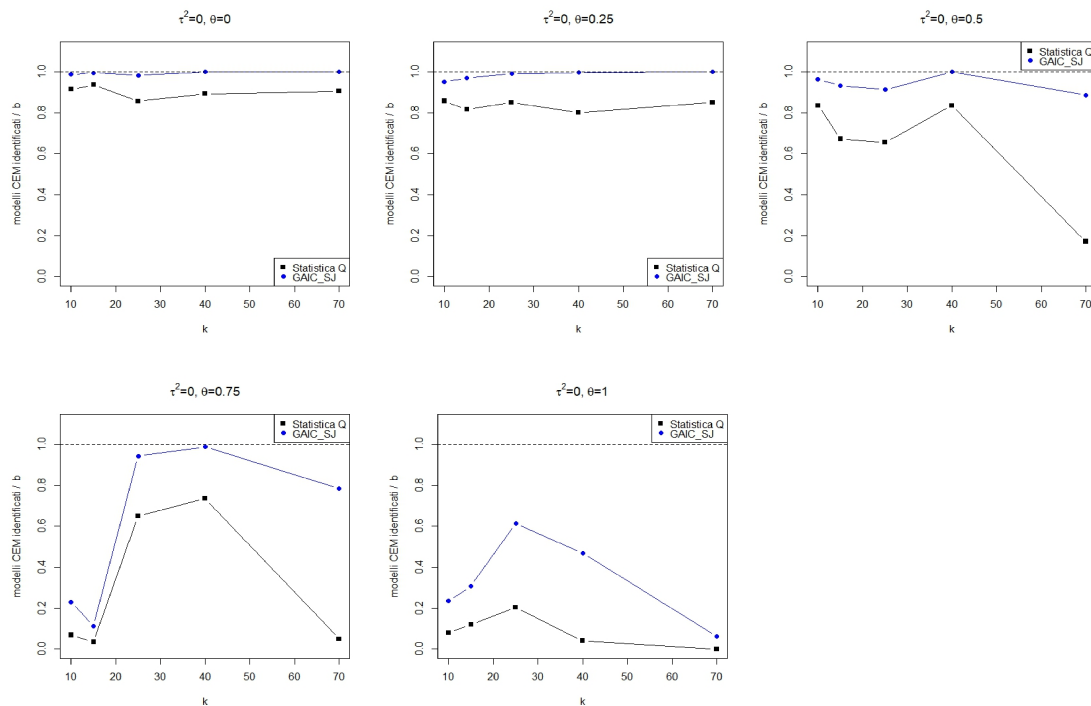


FIGURA 3.2: In ordinata le proporzioni tra le  $B$  simulazioni dove il metodo di selezione del modello tramite il calcolo dell'  $AIC$ , utilizzando la stima  $\hat{\tau}_{S,J}^2$ , e della statistica  $Q$  ha portato ad accettare l'ipotesi di assenza di eterogeneità, in ascissa il numero di studi inclusi nella meta-analisi  $k$ . I dati sono stati simulati ponendo  $\tau^2 = 0$ .

Si nota subito dalla Figura 3.2 come il metodo basato sul calcolo dell' $AIC$  che utilizza la stima  $\widehat{GAIC}_{REM,SJ}$  abbia performance migliori rispetto alla statistica  $Q$  in tutti gli scenari. Si nota ancora come per valori prossimi a 0 di  $\theta$ ,  $\widehat{GAIC}_{REM,SJ}$  porti a identificare il modello corretto ad effetti fissi in quasi tutte le simulazioni, con performance che aumentano all'aumentare di  $k$  per lo meno fino a  $\theta = 0.5$ , infatti per valori più elevati di  $\theta$  un aumento di  $k$  porta inizialmente a migliori performance, ma per  $k = 70$  si ha un peggioramento delle stesse. In generale, un aumento di  $\theta$  porta ad una diminuzione delle performance di questi due metodi, con  $\widehat{GAIC}_{REM,SJ}$  che risulta comunque sempre preferibile tra i due. Utilizzando  $\widehat{GAIC}_{REM,DL}$ ,  $\widehat{GAIC}_{REM,HE}$  e  $\widehat{GAIC}_{REM,REML}$  si ottengono performance nell'identificare il modello ad effetti fissi inferiori alla statistica  $Q$  in ogni scenario analizzato in Figura 3.2

Nella Tabella 3.1 sono riportati i rapporti tra il numero di volte in cui i differenti metodi di selezione del modello analizzati hanno portato a preferire il modello ad effetti fissi CEM rispetto al totale delle  $B$  simulazioni, per differenti scenari ottenuti facendo variare  $k$ ,  $\theta$  e  $\tau^2$ . Tutti gli scenari analizzati in questa tabella utilizzavano dati simulati a partire da un modello ad effetti casuali, dato che si utilizzavano valori di  $\tau^2 > 0$ . Pertanto più questo rapporto è basso, più le performance sono alte e quindi più il metodo di selezione analizzato è in grado di identificare correttamente la presenza di questa eterogeneità e favorire l'utilizzo di un modello ad effetti casuali.

In generale, all'aumentare dei valori di  $k$ ,  $\theta$  e  $\tau^2$  si riduce il numero di modelli ad effetti fissi identificati, a favore di un modello ad effetti casuali. La statistica  $Q$  sembra essere la migliore nell'identificare questo modello, mentre l'utilizzo di  $\widehat{GAIC}_{REM,SJ}$  porta alle performance peggiori in tutti gli scenari analizzati nella tabella. Le stime  $\widehat{GAIC}_{REM,REML}$ ,  $\widehat{GAIC}_{REM,HE}$ ,  $\widehat{GAIC}_{REM,DL}$  presentano performance molto simili tra loro, ma comunque inferiori a quelle della statistica  $Q$ . Nelle condizioni di bassi valori di  $k$  e di  $\tau^2$  tutti i metodi analizzati presentano il problema di portare a selezionare il modello ad effetti fissi quando effettivamente una certa eterogeneità è presente. In queste condizioni la statistica  $Q$  risulta quella con le performance migliori, mentre i nuovi metodi proposti risultano tutti avere performance nettamente inferiori. Il numero di modelli ad effetti casuali correttamente identificati in generale aumentano con l'aumentare di  $\theta$ , ma aumenti del valore di  $k$  e  $\tau^2$  rimangono quelli che portano le maggiori variazioni nelle performance dei metodi di selezione del modello.

		$\theta = 0$				
		$Q$	$\widehat{GAIC}_{REM,SJ}$	$\widehat{GAIC}_{REM,DL}$	$\widehat{GAIC}_{REM,HE}$	$\widehat{GAIC}_{REM,REML}$
$\tau^2 = 0.05$	$k = 10$	0.229	0.402	0.325	0.338	0.327
	$k = 15$	0.300	0.553	0.403	0.370	0.404
	$k = 25$	0.050	0.228	0.078	0.083	0.078
	$k = 40$	0.072	0.444	0.094	0.131	0.092
	$k = 70$	0.001	0.281	0.004	0.056	0.006
$\tau^2 = 0.1$	$k = 10$	0.200	0.355	0.270	0.293	0.272
	$k = 15$	0.080	0.234	0.129	0.151	0.124
	$k = 25$	0.027	0.108	0.036	0.055	0.032
	$k = 40$	0.000	0.008	0.000	0.004	0.000
	$k = 70$	0.000	0.030	0.000	0.002	0.000
$\tau^2 = 0.15$	$k = 10$	0.087	0.169	0.131	0.132	0.131
	$k = 15$	0.002	0.023	0.007	0.011	0.007
	$k = 25$	0.000	0.006	0.001	0.003	0.000
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 0.2$	$k = 10$	0.025	0.084	0.053	0.066	0.055
	$k = 15$	0.011	0.034	0.020	0.030	0.018
	$k = 25$	0.003	0.012	0.003	0.008	0.003
	$k = 40$	0.000	0.001	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 1$	$k = 10$	0.000	0.004	0.000	0.003	0.000
	$k = 15$	0.000	0.000	0.000	0.000	0.000
	$k = 25$	0.000	0.000	0.000	0.000	0.000
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
		$\theta = 0.25$				
		$Q$	$\widehat{GAIC}_{REM,SJ}$	$\widehat{GAIC}_{REM,DL}$	$\widehat{GAIC}_{REM,HE}$	$\widehat{GAIC}_{REM,REML}$
$\tau^2 = 0.05$	$k = 10$	0.536	0.749	0.674	0.677	0.677
	$k = 15$	0.319	0.574	0.420	0.432	0.394
	$k = 25$	0.151	0.448	0.199	0.227	0.186
	$k = 40$	0.030	0.242	0.033	0.066	0.026
	$k = 70$	0.000	0.067	0.000	0.009	0.001
$\tau^2 = 0.1$	$k = 10$	0.154	0.296	0.222	0.248	0.219
	$k = 15$	0.089	0.212	0.127	0.126	0.123
	$k = 25$	0.011	0.101	0.020	0.051	0.016
	$k = 40$	0.000	0.002	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000

		$Q$	$\widehat{GAIC}_{REM,SJ}$	$\widehat{GAIC}_{REM,DL}$	$\widehat{GAIC}_{REM,HE}$	$\widehat{GAIC}_{REM,REML}$
$\tau^2 = 0.15$	$k = 10$	0.140	0.269	0.202	0.219	0.206
	$k = 15$	0.027	0.092	0.040	0.058	0.039
	$k = 25$	0.001	0.013	0.001	0.004	0.000
	$k = 40$	0.000	0.001	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 0.2$	$k = 10$	0.026	0.065	0.040	0.049	0.044
	$k = 15$	0.013	0.047	0.020	0.027	0.020
	$k = 25$	0.000	0.002	0.000	0.002	0.000
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 1$	$k = 10$	0.001	0.002	0.001	0.001	0.001
	$k = 15$	0.000	0.001	0.000	0.000	0.000
	$k = 25$	0.000	0.000	0.000	0.000	0.000
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\theta = 0.5$						
		$Q$	$\widehat{GAIC}_{REM,SJ}$	$\widehat{GAIC}_{REM,DL}$	$\widehat{GAIC}_{REM,HE}$	$\widehat{GAIC}_{REM,REML}$
$\tau^2 = 0.05$	$k = 10$	0.327	0.502	0.441	0.420	0.441
	$k = 15$	0.339	0.588	0.428	0.432	0.423
	$k = 25$	0.047	0.223	0.074	0.120	0.071
	$k = 40$	0.043	0.298	0.054	0.086	0.043
	$k = 70$	0.000	0.039	0.000	0.006	0.000
$\tau^2 = 0.1$	$k = 10$	0.344	0.512	0.439	0.452	0.438
	$k = 15$	0.059	0.170	0.097	0.114	0.087
	$k = 25$	0.028	0.111	0.039	0.052	0.035
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 0.15$	$k = 10$	0.135	0.278	0.196	0.224	0.198
	$k = 15$	0.047	0.127	0.073	0.095	0.067
	$k = 25$	0.004	0.035	0.009	0.012	0.008
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 0.2$	$k = 10$	0.049	0.096	0.074	0.075	0.071
	$k = 15$	0.006	0.014	0.007	0.010	0.007
	$k = 25$	0.000	0.005	0.001	0.000	0.001
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 1$	$k = 10$	0.000	0.003	0.003	0.003	0.003
	$k = 15$	0.000	0.000	0.000	0.000	0.000
	$k = 25$	0.000	0.000	0.000	0.000	0.000
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000

		$\theta = 0.75$				
		$Q$	$\widehat{GAIC}_{REM,SJ}$	$\widehat{GAIC}_{REM,DL}$	$\widehat{GAIC}_{REM,HE}$	$\widehat{GAIC}_{REM,REML}$
$\tau^2 = 0.05$	$k = 10$	0.245	0.442	0.345	0.404	0.351
	$k = 15$	0.372	0.610	0.471	0.452	0.463
	$k = 25$	0.088	0.291	0.121	0.142	0.112
	$k = 40$	0.000	0.027	0.002	0.003	0.002
	$k = 70$	0.000	0.020	0.000	0.002	0.000
$\tau^2 = 0.1$	$k = 10$	0.297	0.507	0.412	0.425	0.420
	$k = 15$	0.041	0.089	0.054	0.058	0.054
	$k = 25$	0.006	0.051	0.010	0.017	0.008
	$k = 40$	0.000	0.002	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 0.15$	$k = 10$	0.062	0.142	0.111	0.118	0.111
	$k = 15$	0.008	0.035	0.018	0.023	0.018
	$k = 25$	0.000	0.020	0.000	0.002	0.000
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 0.2$	$k = 10$	0.017	0.044	0.030	0.037	0.031
	$k = 15$	0.010	0.059	0.019	0.027	0.017
	$k = 25$	0.000	0.001	0.000	0.000	0.000
	$k = 40$	0.000	0.020	0.000	0.002	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 1$	$k = 10$	0.000	0.000	0.000	0.000	0.000
	$k = 15$	0.000	0.000	0.000	0.000	0.000
	$k = 25$	0.000	0.000	0.000	0.000	0.000
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
		$\theta = 1$				
		$Q$	$\widehat{GAIC}_{REM,SJ}$	$\widehat{GAIC}_{REM,DL}$	$\widehat{GAIC}_{REM,HE}$	$\widehat{GAIC}_{REM,REML}$
$\tau^2 = 0.05$	$k = 10$	0.051	0.113	0.089	0.078	0.089
	$k = 15$	0.003	0.009	0.005	0.005	0.004
	$k = 25$	0.037	0.156	0.051	0.056	0.047
	$k = 40$	0.001	0.014	0.001	0.003	0.001
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 0.1$	$k = 10$	0.065	0.142	0.108	0.110	0.108
	$k = 15$	0.077	0.185	0.111	0.125	0.110
	$k = 25$	0.001	0.013	0.003	0.000	0.004
	$k = 40$	0.000	0.008	0.000	0.001	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000



	$Q$	$\widehat{GAIC}_{REM,SJ}$	$\widehat{GAIC}_{REM,DL}$	$\widehat{GAIC}_{REM,HE}$	$\widehat{GAIC}_{REM,REML}$	
$\tau^2 = 0.15$	$k = 10$	0.044	0.123	0.084	0.098	0.088
	$k = 15$	0.030	0.109	0.050	0.079	0.048
	$k = 25$	0.001	0.015	0.003	0.005	0.004
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 0.2$	$k = 10$	0.054	0.105	0.081	0.080	0.081
	$k = 15$	0.013	0.039	0.018	0.030	0.014
	$k = 25$	0.001	0.006	0.001	0.003	0.002
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000
$\tau^2 = 1$	$k = 10$	0.000	0.000	0.000	0.000	0.000
	$k = 15$	0.000	0.000	0.000	0.000	0.000
	$k = 25$	0.000	0.000	0.000	0.000	0.000
	$k = 40$	0.000	0.000	0.000	0.000	0.000
	$k = 70$	0.000	0.000	0.000	0.000	0.000

TABELLA 3.1: Proporzione di volte rispetto alle  $B$  simulazioni in cui i diversi metodi di selezione del modello hanno portato alla scelta di un modello ad effetti fissi, per diversi valori di  $\tau^2$ ,  $\theta$  e  $k$  utilizzati per la simulazione dei dati.

In generale, l'utilizzo della stima  $\hat{\tau}_{REML}^2$  nel calcolo di  $\widehat{GAIC}_{REM}$  porta, rispetto alle altre stime di  $\hat{\tau}^2$ , a identificare un numero di volte inferiore rispetto alle altre stime di  $\widehat{GAIC}_{REM}$  il modello ad effetti fissi quando questo è effettivamente il modello utilizzato per simulare i dati, e porta sempre a identificare un numero inferiore di volte il modello ad effetti fissi quando il modello che ha simulato i dati è un modello ad effetti casuali. Viceversa, l'utilizzo della stima  $\hat{\tau}_{SJ}^2$  porta a preferire il modello ad effetti fissi un numero di volte maggiore, sia quando i dati sono effettivamente simulati a partire da questo modello, sia quando invece il modello corretto è quello ad effetti casuali. Nello scenario di bassi valori di  $k$  e  $\tau^2$ , ma comunque mantenendo  $\tau^2 > 0$ , la statistica  $Q$  risulta la migliore rispetto a  $\widehat{GAIC}_{REM,DL}$ ,  $\widehat{GAIC}_{REM,SJ}$ ,  $\widehat{GAIC}_{REM,HE}$  e  $\widehat{GAIC}_{REM,REML}$  nell'identificare correttamente il modello ad effetti casuali.

Se invece si confronta la stima  $\widehat{AIC}_{REM}$  con  $AIC_{CEM}$ , come proposto in Figura 3.3, per i valori più bassi di  $k$  la statistica  $Q$  è la migliore nell'identificare correttamente il modello ad effetti fissi come quello utilizzato per simulare le osservazioni, mentre a partire da  $k = 25$  le performance migliori nel caso di  $\tau^2 = 0$  le presenta  $\widehat{AIC}_{REM,SJ}$ . Anche utilizzando la stima  $\widehat{AIC}_{REM}$ , se inizialmente un aumento di  $k$  porta a migliori

performance, per  $k = 70$  tutti i metodi presentano un chiaro peggioramento nel riuscire ad identificare correttamente il modello ad effetti fissi come quello utilizzato per simulare i dati. Per  $\tau^2 > 0$ ,  $\widehat{AIC}_{REM,SJ}$ ,  $\widehat{AIC}_{REM,DL}$ ,  $\widehat{AIC}_{REM,HE}$  e  $\widehat{AIC}_{REM,REML}$  presentano performance migliori della statistica  $Q$ , performance che migliorano all'aumentare del numero di studi inclusi nella meta-analisi, e per i valori più elevati di  $k$  tutti i metodi portano a preferire correttamente il modello ad effetti casuali in tutte le  $B$  simulazioni. Questi risultati sono stati ottenuti ponendo  $\theta = 0.5$ , per valori differenti di  $\theta$  non cambiano le conclusioni. Per valori più elevati di  $\theta$  si osserva come i metodi di selezione considerati portino a preferire più spesso il modello ad effetti casuali, riducendo in generale le performance quando il modello utilizzato per simulare i dati è quello ad effetti fissi, mentre migliorano le performance nel caso in cui il modello utilizzato sia quello ad effetti casuali. Viceversa, valori inferiori di  $\theta$  portano a preferire più spesso il modello ad effetti fissi, con le performance che migliorano pertanto quando questo è il modello che effettivamente è stato usato per simulare i dati.

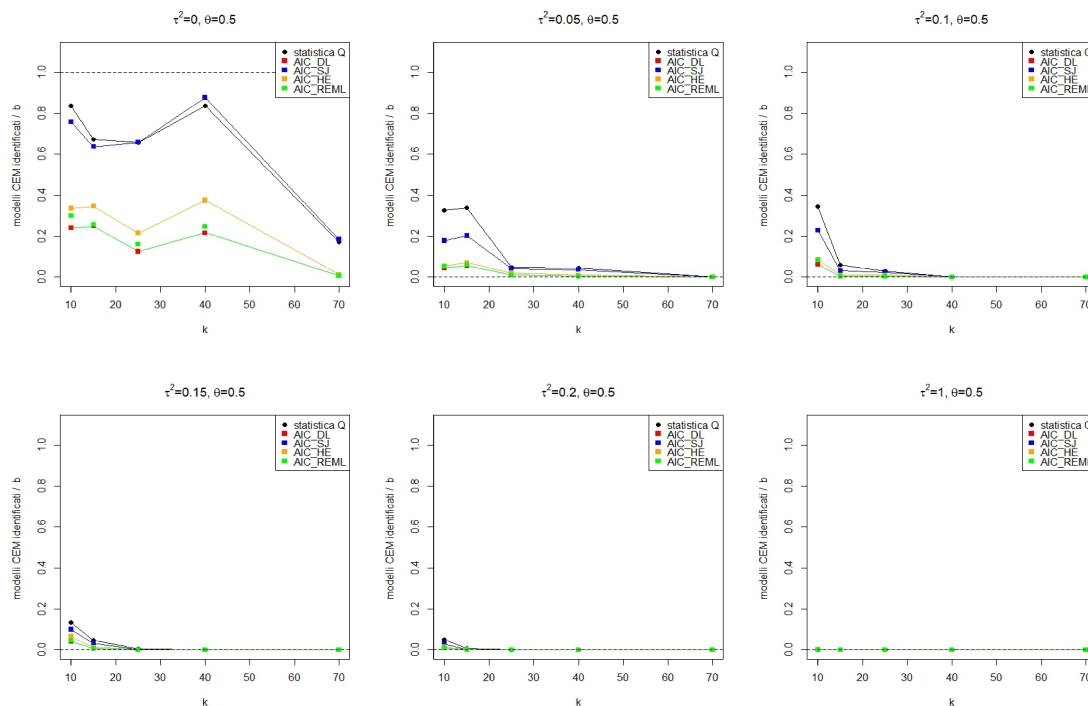


FIGURA 3.3: In ordinata le proporzioni tra le  $B$  simulazioni dove il metodo di selezione ha portato a preferire il modello ad effetti fissi, in ascissa il numero di studi inclusi nella meta-analisi  $k$ . I dati sono stati simulati per differenti valori di  $\tau^2$  e fissando  $\theta = 0.5$ , e vengono confrontate le performance della statistica  $Q$  con quelle del metodo di selezione basato sul calcolo dell' $AIC$  usando  $\widehat{AIC}_{REM}$ , calcolato con differenti stime di  $\tau^2$ , per il confronto con  $AIC_{CEM}$ .

Nella Tabella 3.2 si confrontano i valori medi per  $AIC_{CEM}$  e  $\widehat{AIC}_{REM}$ ,  $\widehat{GAIC}_{REM}$  calcolati utilizzando differenti stime per  $\tau^2$  e simulando i dati da cui si ricavano questi valori a partire da un modello ad effetti fissi con  $\tau^2 = 0$ ,  $k = 10$  e facendo variare i valori di  $\theta$ . L'utilizzo di  $\widehat{AIC}_{REM,DL}$ ,  $\widehat{AIC}_{REM,HE}$  e  $\widehat{AIC}_{REM,REML}$  porta a preferire erroneamente il modello ad effetti casuali per ogni valore di  $\theta$ . Anche  $\widehat{AIC}_{REM,SJ}$  a partire da  $\theta = 0.5$  porta a preferire questo modello. L'utilizzo di  $\widehat{GAIC}_{REM}$  porta a preferire indipendentemente dalla stima  $\hat{\tau}^2$  il modello ad effetti casuali a partire da  $\theta = 0.75$ . Per altri valori di  $\theta$  invece,  $\widehat{GAIC}_{REM,DL}$ ,  $\widehat{GAIC}_{REM,HE}$ ,  $\widehat{GAIC}_{REM,SJ}$  e  $\widehat{GAIC}_{REM,REML}$  portano a preferire correttamente il modello ad effetti fissi. I valori di  $\widehat{AIC}_{REM,DL}$ ,  $\widehat{AIC}_{REM,HE}$  e  $\widehat{AIC}_{REM,REML}$  sono molto simili tra loro, come anche quelli di  $\widehat{GAIC}_{REM,DL}$ ,  $\widehat{GAIC}_{REM,HE}$  e  $\widehat{GAIC}_{REM,REML}$ . Invece  $\widehat{AIC}_{REM,SJ}$  e  $\widehat{GAIC}_{REM,SJ}$  presentano valori più elevati rispetto a quelli ottenuti utilizzando differenti stime di  $\tau^2$ , anche se queste differenze diminuiscono all'aumentare di  $\theta$ .

Analizzando invece la Tabella 3.3, dove sono riportati gli stessi valori della Tabella 3.2 calcolati per  $k = 70$ , si nota come  $\widehat{AIC}_{REM,DL}$ ,  $\widehat{AIC}_{REM,HE}$  e  $\widehat{AIC}_{REM,REML}$  portino a preferire anche in questo caso per ogni valore di  $\theta$  il modello ad effetti casuali. Nuovamente,  $\widehat{AIC}_{REM,SJ}$  porta a preferire il modello ad effetti casuali solamente a partire da  $\theta = 0.25$ , e per valori superiori di  $\theta$  porta sempre a preferire questo modello. Utilizzando  $\widehat{GAIC}_{REM}$ , a partire da  $\theta = 0.5$  si preferisce il modello ad effetti casuali a prescindere dalla stima di  $\tau^2$  utilizzata. Anche per  $k = 70$ ,  $\widehat{AIC}_{REM,SJ}$  e  $\widehat{GAIC}_{REM,SJ}$  presentano valori più elevati rispetto a quelli ottenuti utilizzando altre stime di  $\tau^2$ , ma queste differenze tendono a diminuire con l'aumentare di  $\theta$ .

Per altri valori di  $k$  si ottengono scenari intermedi tra quelli riportati nelle Tabelle 3.2 e 3.3, che rappresentano gli scenari che abbiamo utilizzato per simulare i dati con meno e più studi inclusi nella meta-analisi.

	$\theta = 0$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$	$\theta = 1$
$AIC_{CEM}$	-1.680	-0.615	-0.741	17.044	19.792
$\widehat{GAIC}_{REM,DL}$	2.215	2.715	2.411	10.433	11.998
$\widehat{GAIC}_{REM,HE}$	2.510	2.919	3.272	11.121	13.744
$\widehat{GAIC}_{REM,SJ}$	4.927	5.309	5.382	11.812	14.055
$\widehat{GAIC}_{REM,REML}$	2.125	2.719	2.223	10.883	12.803
$\widehat{AIC}_{REM,DL}$	-2.016	-1.265	-1.276	8.005	10.220
$\widehat{AIC}_{REM,HE}$	-1.902	-1.158	-1.029	8.079	10.396
$\widehat{AIC}_{REM,SJ}$	-1.141	-0.486	-0.330	8.177	10.459
$\widehat{AIC}_{REM,REML}$	-2.043	-1.279	-1.321	8.028	10.248

TABELLA 3.2: Confronto tra i valori medi di  $AIC_{CEM}$  e  $\widehat{GAIC}_{REM}$ ,  $\widehat{AIC}_{REM}$  ottenuti tramite differenti stime di  $\tau^2$ . I valori sono ottenuti a partire da dati simulati ponendo  $\tau^2 = 0$  e  $k = 10$  e facendo variare  $\theta$ .

	$\theta = 0$	$\theta = 0.25$	$\theta = 0.5$	$\theta = 0.75$	$\theta = 1$
$AIC_{CEM}$	-11.475	-17.727	17.724	29.115	87.955
$\widehat{GAIC}_{REM,DL}$	-5.546	-12.929	12.388	19.734	49.686
$\widehat{GAIC}_{REM,HE}$	-1.859	-8.540	11.359	22.177	61.970
$\widehat{GAIC}_{REM,SJ}$	21.320	13.698	29.755	36.551	66.459
$\widehat{GAIC}_{REM,REML}$	-7.378	-14.965	12.872	19.420	51.702
$\widehat{AIC}_{REM,DL}$	-11.747	-18.186	9.314	18.045	53.911
$\widehat{AIC}_{REM,HE}$	-11.100	-17.395	9.550	18.291	54.738
$\widehat{AIC}_{REM,SJ}$	-4.886	-11.736	11.579	20.225	55.425
$\widehat{AIC}_{REM,REML}$	-11.881	-18.331	9.302	18.031	53.896

TABELLA 3.3: Confronto tra i valori medi di  $AIC_{CEM}$  e  $\widehat{GAIC}_{REM}$ ,  $\widehat{AIC}_{REM}$  ottenuti tramite differenti stime di  $\tau^2$ . I valori sono ottenuti a partire da dati simulati ponendo  $\tau^2 = 0$  e  $k = 70$  e facendo variare  $\theta$ .

Nella Tabella 3.4 si confrontano i valori medi di  $AIC_{CEM}$  e  $\widehat{AIC}_{REM}$ ,  $\widehat{GAIC}_{REM}$  calcolati utilizzando differenti stime per  $\tau^2$  e simulando i dati da cui si ricavano questi valori a partire da un modello ad effetti casuali con  $\tau^2 > 0$  e ponendo  $k = 10$ ,  $\theta = 0.5$ . Per ogni possibile scenario e per ogni possibile stima con cui confrontare  $AIC_{CEM}$ , si preferisce sempre il modello ad effetti casuali. I valori di  $\widehat{GAIC}_{REM,DL}$ ,  $\widehat{GAIC}_{REM,HE}$  e  $\widehat{GAIC}_{REM,REML}$  sono molto simili tra loro, come anche quelli di  $\widehat{AIC}_{REM,DL}$ ,  $\widehat{AIC}_{REM,HE}$  e  $\widehat{AIC}_{REM,REML}$ . I valori di  $\widehat{AIC}_{REM,SJ}$  e  $\widehat{GAIC}_{REM,SJ}$  risultano essere più elevati, anche se con l'aumentare di  $\tau^2$  queste differenze tendono a diminuire. Facendo variare  $\theta$ , le conclusioni a cui si giunge rimangono le stesse, solamente per  $\theta = 0$  e  $\tau^2 = 0.05$ ,  $\widehat{GAIC}_{REM,SJ}$  porterebbe a preferire erroneamente il modello ad effetti fissi.

	$\tau^2 = 0.05$	$\tau^2 = 0.1$	$\tau^2 = 0.15$	$\tau^2 = 0.2$	$\tau^2 = 1$
$AIC_{CEM}$	9.963	12.494	20.009	35.708	123.080
$\widehat{GAIC}_{REM,DL}$	6.987	9.590	10.774	13.481	27.558
$\widehat{GAIC}_{REM,HE}$	6.809	9.618	11.687	13.711	29.425
$\widehat{GAIC}_{REM,SJ}$	8.169	10.995	12.325	14.163	29.292
$\widehat{GAIC}_{REM,REML}$	6.941	9.533	11.152	13.690	29.091
$\widehat{AIC}_{REM,DL}$	3.753	6.437	8.375	10.734	26.444
$\widehat{AIC}_{REM,HE}$	3.773	6.518	8.512	10.763	26.294
$\widehat{AIC}_{REM,SJ}$	4.006	6.751	8.610	10.801	26.271
$\widehat{AIC}_{REM,REML}$	3.741	6.410	8.375	10.716	26.244

TABELLA 3.4: Confronto tra i valori medi di  $AIC_{CEM}$  e  $\widehat{GAIC}_{REM}$ ,  $\widehat{AIC}_{REM}$  ottenuti tramite differenti stime di  $\tau^2$ . I valori sono ottenuti a partire da dati simulati ponendo  $\theta = 0.5$  e  $k = 10$  e facendo variare  $\tau^2$ .

Nella Tabella 3.5 si confrontano nuovamente i valori medi di  $AIC_{CEM}$  e  $\widehat{AIC}_{REM}$ ,  $\widehat{GAIC}_{REM}$  ponendo sempre  $\theta = 0.5$  ma con  $k = 70$ , e facendo variare  $\tau^2$ . Qualunque sia la stima che uso per  $AIC_{REM}$ , questa comunque risulterà inferiore ad  $AIC_{CEM}$ , portandoci sempre a preferire correttamente il modello ad effetti casuali. Anche in questo caso i valori che posso calcolare per  $\widehat{AIC}_{REM}$  e  $\widehat{GAIC}_{REM}$  sono molto simili tra loro, eccetto  $\widehat{AIC}_{REM,SJ}$  e  $\widehat{GAIC}_{REM,SJ}$  che per i valori più bassi di  $\tau^2$  presentano valori maggiori rispetto agli altri, e rispetto alla Tabella 3.4 tutti i valori sono più grandi. Per altri valori di  $\theta$  le conclusioni a cui si giunge rimangono analoghe al caso  $\theta = 0.5$ .

	$\tau^2 = 0.05$	$\tau^2 = 0.1$	$\tau^2 = 0.15$	$\tau^2 = 0.2$	$\tau^2 = 1$
$AIC_{CEM}$	100.825	148.350	216.511	348.606	1195.796
$\widehat{GAIC}_{REM,DL}$	51.108	64.818	83.023	108.097	196.253
$\widehat{GAIC}_{REM,HE}$	61.805	66.282	90.215	113.888	219.200
$\widehat{GAIC}_{REM,SJ}$	65.776	71.174	92.040	114.881	217.717
$\widehat{GAIC}_{REM,REML}$	53.211	65.613	86.775	112.182	215.570
$\widehat{AIC}_{REM,DL}$	54.392	62.997	85.565	110.210	217.281
$\widehat{AIC}_{REM,HE}$	55.093	63.054	85.638	110.112	214.517
$\widehat{AIC}_{REM,SJ}$	55.573	63.308	85.742	110.135	214.416
$\widehat{AIC}_{REM,REML}$	54.365	62.985	85.463	110.036	214.343

TABELLA 3.5: Confronto tra i valori medi di  $AIC_{CEM}$  e  $\widehat{GAIC}_{REM}$ ,  $\widehat{AIC}_{REM}$  ottenuti tramite differenti stime di  $\tau^2$ . I valori sono ottenuti a partire da dati simulati ponendo  $\theta = 0.5$  e  $k = 70$  e facendo variare  $\tau^2$ .

## Capitolo 4

# Applicazione ad un dataset reale

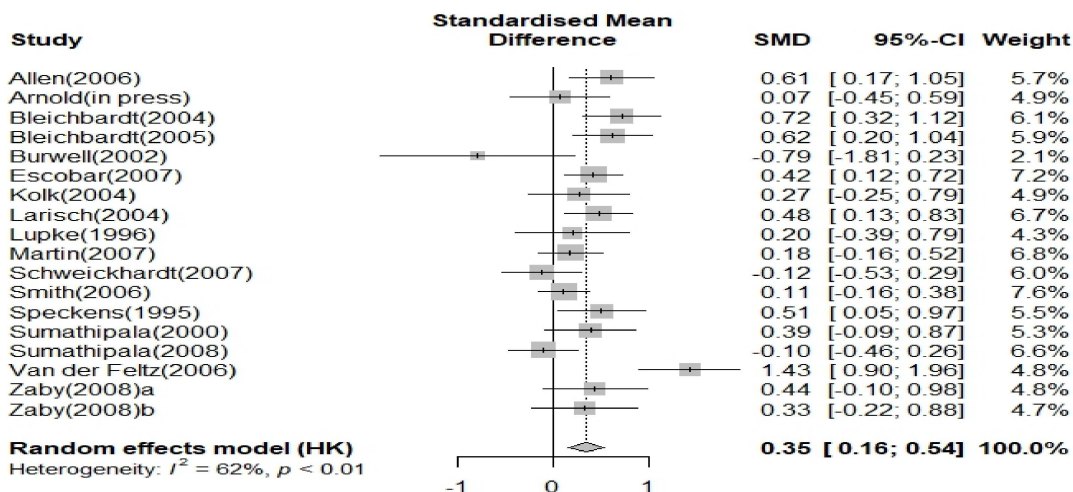
I sintomi fisici di causa sconosciuta, abbreviati con la sigla *MUPS*, sono sintomi a livello fisico presentati dai pazienti ma che non sembrano essere causati da una qualche patologia organica identificabile. Questi sintomi influenzano negativamente la vita lavorativa e privata degli individui, e possono condurre a situazioni di stress e disordini di natura psichica. Per questo oltre agli esami e alle cure mediche tradizionali, per questi pazienti è indispensabile analizzare gli aspetti psicologici della malattia, anche per ridurre l'enorme peso che hanno sulle spese sanitarie, spesso senza che queste portino a miglioramenti concreti.

Nello studio di Kleinstäuber et al. (2011) si valuta l'effetto della psicoterapia sui sintomi fisici sperimentati dai pazienti tramite una meta-analisi di precedenti studi in cui si confrontava lo stesso parametro tra un gruppo di controllo e uno di trattamento tramite il calcolo di una differenza standardizzata tra le medie dei due gruppi, calcolata tramite la formula (3.1). Al gruppo dei trattati venivano prescritti degli incontri con degli specialisti, con le sedute psicoterapeutiche che avevano lo scopo di ricostruire la cognizione del paziente rispetto alla sua malattia, ridurre lo stress e l'ansia e anche quello di sviluppare tecniche di rilassamento. I dati degli studi sono riportati nella Tabella 4.1.

Studio	$d$	$\sigma_d$	$n_{controllo}$	$n_{trattati}$
Allen et al. (2006)	0.61	0.224	43	41
Arnold et al. (2009)	0.07	0.265	28	28
Bleichhardt et al. (2004)	0.72	0.204	89	34
Bleichhardt et al. (2005)	0.62	0.214	68	34
Walsh (2002)	-0.79	0.52	9	7
Escobar et al. (2007)	0.42	0.153	87	85
Kolk et al. (2004)	0.27	0.265	80	18
Larisch et al. (2004)	0.48	0.179	73	54
Lupke et al. (1996)	0.20	0.301	27	18
Martin et al. (2007)	0.18	0.173	70	70
Schweickhardt et al. (2007)	-0.12	0.209	49	42
Smith et al. (2006)	0.11	0.138	98	102
Speckens et al. (1995)	0.51	0.235	39	38
Sumathipala et al. (2000)	0.39	0.245	34	34
Sumathipala et al. (2008)	-0.10	0.184	64	60
van der Feltz-Cornelis et al. (2006)	1.43	0.27	58	23
Zaby et al. (2008)a	0.44	0.276	22	35
Zaby et al. (2008)b	0.33	0.281	20	35

TABELLA 4.1: Caratteristiche degli studi inclusi nella meta-analisi in Kleinstäuber et al. (2011).

FIGURA 4.1: Forest-plot degli studi inclusi nella meta-analisi. Si riporta per ogni studio il valore del parametro di interesse, il relativo intervallo di confidenza di livello 95% e il relativo peso dello studio quando si applica un modello ad effetti casuali.





Tramite il modello ad effetti casuali (1.1), applicato ai dati con il comando [rma](#) della libreria *metafor*, otteniamo una stima del parametro di interesse  $\hat{\mu} = 0.3481$ , con una varianza di 0.007, e applicando un test di significatività di livello  $\alpha = 0.05$ , possiamo concludere che il trattamento psicoterapeutico porta a ridurre i sintomi sperimentati dai pazienti. E' fornita anche la stima  $\hat{\tau}_{REML}^2 = 0.078$ , e un valore per la statistica  $Q$  di 44.6811 che porta a rifiutare l'ipotesi  $H_0$  di assenza di eterogeneità ad un livello di significatività  $\alpha = 0.05$ . Applicando le formule (2.8), (2.6), utilizzando la stima  $\hat{\tau}_{REML}^2$  otteniamo  $AIC_{CEM} = 24.867$  e  $\widehat{GAIC}_{REM,REML} = 17.175$ , confrontando questi due valori si è portati a preferire il modello ad effetti casuali. Utilizzando la formula (2.6) otteniamo  $\widehat{AIC}_{REM,REML} = 16.56$  che se confrontato con  $AIC_{CEM}$  ci porta a preferire anche con questa stima il modello ad effetti casuali.

Per valutare il metodo di selezione basato sul calcolo dell' $AIC$ , eseguiamo una simulazione con in questo caso  $B = 4000$  ripetizioni, in cui ad ogni simulazione a partire dal modello (1.1) simuliamo i dati ponendo  $k = 18$ ,  $\tau^2 = 0.078$  e  $\theta = 0.3481$  e le numerosità campionarie per ogni studio  $n_{i1}$  e  $n_{i2}$  uguali per ogni  $b$ -esima simulazione e pari a quelle osservate nei dati reali. Le osservazioni per il gruppo di controllo e quello dei trattati sono simulate con lo stesso procedimento esposto nello studio sperimentale, e anche le quantità calcolate risultano essere le stesse.

Nella Tabella 4.2 sono riportate le proporzioni tra le volte in cui il metodo di selezione del modello corrispondente ha portato a preferire il modello ad effetti fissi, rispetto alle  $B$  simulazioni totali. Il numero di modelli ad effetti fissi erroneamente preferiti è il più basso se utilizzo il metodo basato sul calcolo dell' $AIC$  utilizzando la stima  $\widehat{AIC}_{REM,DL}$  e  $\widehat{AIC}_{REM,REML}$  per il confronto con  $AIC_{CEM}$ . In generale, comunque, usando la stima  $\widehat{AIC}_{REM}$  si preferisce un numero di volte maggiore il modello ad effetti casuali rispetto alla statistica  $Q$ , che in quasi l' 8% delle simulazioni porta a preferire il modello sbagliato, mentre l'uso di  $\widehat{GAIC}_{REM}$ , indipendentemente dal metodo che uso per calcolare  $\hat{\tau}^2$ , porta ad un numero minore di modelli ad effetti casuali preferiti rispetto a quello ad effetti fissi. Il metodo peggiore risulta essere quello che usa  $\widehat{GAIC}_{REM,SJ}$ , che in più del 25% delle simulazioni porta a preferire il modello errato.

---

Statistica $Q$			
0.079			
$\widehat{GAIC}_{REM,DL}$	$\widehat{GAIC}_{REM,HE}$	$\widehat{GAIC}_{REM,SJ}$	$\widehat{GAIC}_{REM,REML}$
0.118	0.155	0.251	0.106
$\widehat{AIC}_{REM,DL}$	$\widehat{AIC}_{REM,HE}$	$\widehat{AIC}_{REM,SJ}$	$\widehat{AIC}_{REM,REML}$
0.012	0.024	0.058	0.012

TABELLA 4.2: Rapporto tra il numero di volte in cui i differenti metodi di selezione del modello hanno portato a preferire il modello ad effetti fissi rispetto alle  $B$  simulazioni.

# Conclusioni

In questa tesi sono stati studiati alcuni metodi di selezione del modello ad effetti casuali rispetto al modello ad effetti fissi nella meta-analisi, che vanno oltre l'utilizzo delle statistiche di base quali  $Q$ ,  $I^2$ ,  $H^2$ . L'attenzione è stata posta sui metodi che si basano sul criterio  $AIC$ . Dato che  $AIC_{REM}$  non può essere calcolato direttamente, dato che  $\tau^2$  non è noto, si utilizza una sua stima  $\widehat{AIC}_{REM}$  oppure  $\widehat{GAIC}_{REM}$ , che riduce l'errore dovuto alla distorsione di  $\hat{\tau}^2$ . Nello studio sperimentale si sono simulati i dati facendo variare i valori di  $k$ ,  $\tau^2$  e  $\theta$ , e per ogni scenario si è calcolata la proporzione tra il numero di modelli ad effetti fissi che il metodo di selezione in esame portava a preferire, rispetto alle  $B$  simulazioni totali. Questo per la statistica  $Q$  e per il metodo basato sul calcolo dell' $AIC$ , distinguendo in base a quale stima di  $AIC_{REM}$  si usava e anche in base a quale stima di  $\tau^2$  veniva utilizzata. Considerando la stima  $\widehat{GAIC}_{REM}$ , quella ottenuta utilizzando lo stimatore  $\hat{\tau}_{S,J}^2$  risulta essere la migliore nell'identificare il modello ad effetti fissi quando questo è quello che effettivamente genera le osservazioni. Utilizzando la stima  $\widehat{GAIC}_{REM}$  in generale il numero di modelli ad effetti fissi correttamente identificati cala, in favore di un maggior numero di modelli ad effetti casuali, quando  $k$  aumenta, portandoci a concludere che, come la statistica  $Q$ , un aumento di  $k$  porti a considerare come significativa ogni minima eterogeneità. Questo non risulta essere del tutto vero per  $\widehat{GAIC}_{REM,SJ}$ , che per i valori più bassi di  $\theta$  considerati per le simulazioni, vede un aumento delle performance all'aumentare di  $k$ . Se invece è presente una certa eterogeneità, e quindi le osservazioni vengono generate utilizzando un modello ad effetti casuali, la stima  $\widehat{GAIC}_{REM}$  porta a performance inferiori anche alla stessa statistica  $Q$ , specialmente nelle condizioni peggiori quando  $k$  e  $\tau^2$  presentano i valori più bassi utilizzati per le simulazioni. Valori più elevati di  $\theta$  portano ad un miglioramento delle performance, ma solo se anche i valori di  $k$  e  $\tau^2$  aumentano si ottengono per i vari  $\widehat{GAIC}_{REM}$ , specialmente per  $\widehat{GAIC}_{REM,REML}$  e  $\widehat{GAIC}_{REM,DL}$ , performance simili a quelle della statistica  $Q$  nell'identificare correttamente il modello ad effetti casuali. Se invece considero la stima  $\widehat{AIC}_{REM}$  le performance della statistica  $Q$  nell'identificare correttamente il modello ad effetti fissi risultano essere maggiori rispetto a  $\widehat{AIC}_{REM,DL}$ ,

$\widehat{AIC}_{REM,HE}$ ,  $\widehat{AIC}_{REM,SJ}$  e  $\widehat{AIC}_{REM,REML}$  per bassi valori di  $k$ , mentre per valori più elevati risulta essere preferibile l'uso di  $\widehat{AIC}_{REM,SJ}$ . Anche in questo caso, si può notare come un aumento del numero di studi inclusi nella meta-analisi porti a migliori performance, ma per i valori più elevati di  $k$  queste presentano un peggioramento. Per quanto riguarda le performance nell'identificare il modello ad effetti casuali,  $\widehat{AIC}_{REM,DL}$ ,  $\widehat{AIC}_{REM,HE}$ ,  $\widehat{AIC}_{REM,SJ}$  e  $\widehat{AIC}_{REM,REML}$  hanno performance migliori della statistica  $Q$ , anche se all'aumentare di  $k$  queste differenze tendono a diminuire. Da notare come  $\widehat{AIC}_{REM,SJ}$  risulti essere migliore della statistica  $Q$  per i valori più alti di  $k$  nell'identificare il modello corretto, a prescindere dal fatto che questo sia quello ad effetti casuali o ad effetti fissi.

Utilizzando i valori medi che ottengo dalle  $B$  simulazioni per  $\widehat{AIC}_{REM}$  e  $\widehat{GAIC}_{REM}$ , si nota l'influenza di un aumento di  $\theta$ , che per i valori più elevati porta a considerare come significativa ogni minima eterogeneità. Questo comporta minori performance quando i dati sono simulati a partire da un modello ad effetti fissi. Per quanto riguarda invece le performance nel caso in cui i dati siano simulati utilizzando un modello ad effetti casuali, variazioni di  $\theta$  non sembrano influire sulle performance. I valori medi di  $\widehat{GAIC}_{REM,DL}$ ,  $\widehat{GAIC}_{REM,SJ}$  e  $\widehat{GAIC}_{REM,REML}$  risultano essere molto simili tra di loro, come quelli di  $\widehat{AIC}_{REM,DL}$ ,  $\widehat{AIC}_{REM,HE}$  e  $\widehat{AIC}_{REM,REML}$ , mentre i valori di  $\widehat{GAIC}_{REM,SJ}$  e  $\widehat{AIC}_{REM,SJ}$  risultano essere più elevati anche se queste differenze diminuiscono all'aumentare di  $k$  e  $\theta$ .

Alla luce di quanto emerso dallo studio sperimentale, per i valori più elevati di  $k$  il metodo che confronta  $AIC_{CEM}$  con  $\widehat{AIC}_{REM,SJ}$  è in grado di identificare il modello corretto utilizzato per simulare i dati più spesso della statistica  $Q$ . Per gli scenari in cui si hanno pochi studi inclusi nella meta-analisi  $k$ , e per valori piccoli di  $\tau^2$ , non si è riusciti a identificare una valida alternativa alla statistica  $Q$ , dato che tutti i metodi proposti presentavano performance inferiori a questa statistica in queste condizioni. Gli altri metodi proposti portavano a preferire troppo spesso il modello ad effetti fissi quando i dati erano simulati da uno ad effetti casuali o viceversa a preferire troppo spesso il modello ad effetti casuali quando i dati invece erano generati da uno ad effetti fissi. Inoltre  $\widehat{AIC}_{REM}$  risulta essere una stima meno stabile e precisa di  $AIC_{REM}$  rispetto a  $\widehat{GAIC}_{REM}$ , come dimostrato in Yang & H.Y. Kwan (2020).

Sviluppi futuri potrebbero concentrarsi sull'utilizzo di altri criteri d'informazione, come il  $BIC$  o l' $AIC_c$ . Inoltre nello studio di simulazione abbiamo estratto le varianze dentro gli studi  $\sigma_{i,1}^2$ ,  $\sigma_{i,1}^2$  da una  $U(0.005, 1)$ , ma si potrebbe considerare altre distribuzioni e soprattutto far variare l'intervallo dei valori che questo parametro può assumere, magari considerando anche disegni particolari come, ad esempio, scenari in cui si mantengono

dei rapporti fissi tra i valori di  $\sigma^2$  nel gruppo di controllo e di trattamento. Si potrebbe anche valutare le performance di questi metodi in scenari che utilizzano valori di  $k$ ,  $\theta$  e  $\tau^2$  diversi da quelli considerati in questa tesi, e parametri di interesse che non siano la sola differenza standardizzata tra medie. Un ultimo spunto di riflessione potrebbe essere il comportamento del metodo basato sul calcolo dell'*AIC* nel caso in cui alcuni studi presentino valori anomali, e magari confrontare le performance con il metodo *bootstrap*.



# Appendice

## .1 Comandi R per la simulazione dei dati

---

```
library(mvtnorm)

set.seed(1234)

k = c(10, 15, 25, 40, 70)
theta = c(0, 0.25, 0.5, 0.75, 1)
tau2 = c(0, 0.05, 0.1, 0.15, 0.2, 1)
n.rep = 1000
dati.all = list(NULL)

#inizio ciclo j
for(j in 1:length(k)){

  this.k = k[j]

  #inizio ciclo l
  for(l in 1:length(tau2)){

    #inizio ciclo h
    for(h in 1:length(theta)){

      #creo delle liste vuote in cui metterò i valori
      mean.y1 = mean.y2 = var.y1 = var.y2 = d = sigma2.d =
        rep(NA, this.k)
```

```
#assegno i valori di theta e tau2 a delle nuove
#variabili
this.tau2 = tau2[1]
this.theta = theta[h]

#simulo i valori delle variane nei gruppi controllo
#e trattamento
sigma2i.1 = runif(this.k, 0.005, 5)
sigma2i.2 = runif(this.k, 0.005, 5)

#simulo i valori per il numero di soggetti nel gruppo
#trattamento e controllo
ni1 = round(runif(this.k, 15, 90))
ni2 = round(runif(this.k, 15, 90))

## inizia ciclo b
for(b in 1:n.rep){

  ## dentro ogni studio
  for(i in 1:this.k){

    #simulo i valori di theta_i
    thetai.1 = rnorm(this.k, mean=this.theta,
                    sd=sqrt(this.tau2))
    thetai.2 = rnorm(this.k, mean=0,
                    sd=sqrt(this.tau2))

    #simulo le osservazioni dentro i gruppi
    yi1 = rnorm(ni1[i], thetai.1[i],
               sd=sqrt(sigma2i.1[i]))
    yi2 = rnorm(ni2[i], thetai.2[i],
               sd=sqrt(sigma2i.2[i]))

    #calcolo medie e varianza delle osservazioni
    mean.y1[i] = mean(yi1)
    var.y1[i] = var(yi1)
```





## .2 Comandi R per il calcolo di $\widehat{AIC}_{REM}$ , $\widehat{GAIC}_{REM}$ , $AIC_{CEM}$

---

```

library(metafor)

## definire i valori
k = c(10, 15, 25, 40,70)
theta = c(0, 0.25, 0.5, 0.75, 1)
tau2 = c(0.0, 0.05, 0.1, 0.15, 0.2, 1)
n.rep = 1000

#creo la funzione per il calcolo di AIC
metaAIC = function(g, sigma2d, k, model , tau2){
  sum1 = 0
  sum2 = 0
  #inizio ciclo i
  for (i in 1:k){
    sum1 = sum1+log(2*pi*(sigma2d[i]+tau2))
    sum2 = sum2+((d[i]-model$beta)^2)/(sigma2d[i]+tau2)
  }
  #chiusura ciclo i
  AIC = sum1+sum2
  return(AIC)
}

#chiusura funzione

#creo la funzione per il calcolo di GAIC
metaGAIC = function(g, sigma2d, k, model , tau2){
  sum1 = 0
  #inizio ciclo i
  for (i in 1:k){
    sum1 = sum1+log(2*pi*(sigma2d[i]+tau2))
  }
  #chiusura ciclo i

```

```
    GAIC = sum1+2+k
    return(GAIC)
}
#chiusura funzione

#inizio ciclo j
for(j in 1:length(k)){

    this.k=k[j]

    #inizio ciclo l
    for(l in 1:length(tau2)){

        #inizio ciclo h
        for(h in 1:length(theta)){

            this.tau2 = tau2[l]
            this.theta = theta[h]

            #carico lo scenario di interesse
            load(paste('scenario', this.k, '_tau', this.tau2,
                '_theta', this.theta, '.RData', sep=''))

            #creo le varie liste in cui inseriro i valori
            Q = rep(NA, n.rep)
            I2 = rep(NA, n.rep)

            GAIC_REM_reml = rep(NA, n.rep)
            AIC_REM_reml = rep(NA, n.rep)
            diff_GA_reml = rep(NA, n.rep)
            diff_AA_reml = rep(NA, n.rep)

            GAIC_REM_dl = rep(NA, n.rep)
            AIC_REM_dl = rep(NA, n.rep)
            diff_GA_dl = rep(NA, n.rep)
```

```
diff_AA_dl = rep(NA, n.rep)

GAIC_REM_sj = rep(NA, n.rep)
AIC_REM_sj = rep(NA, n.rep)
diff_GA_sj = rep(NA, n.rep)
diff_AA_sj = rep(NA, n.rep)

GAIC_REM_he = rep(NA, n.rep)
AIC_REM_he = rep(NA, n.rep)
diff_GA_he = rep(NA, n.rep)
diff_AA_he = rep(NA, n.rep)

AIC_CEM = rep(NA, n.rep)

#inizio ciclo b
for(b in 1: n.rep){

  this.data = dati.all[[b]]

  #modelli con le diverse modalita di stima
  #di tau^2
  REM_reml = rma(this.data$d,
                 this.data$sigma2.d)
  REM_dl = rma(this.data$d, this.data$sigma2.d,
               method="DL")
  REM_sj = rma(this.data$d, this.data$sigma2.d,
               method="SJ")
  REM_he = rma(this.data$d, this.data$sigma2.d,
               method="HE")

  #modello ad effetti fissi
  CEM = rma(this.data$d, this.data$sigma2.d,
            method="FE")

  #calcolo i valori di QB
```

```
Q[b] = REM_reml$QEp

#e i valori di I^2
I2[b] = REM_reml$I2

#calcolo i valori di AIC e GAIC.
AIC_REM_reml[b] = metaAIC(this.data$d,
                           this.data$sigma2.d,
                           this.k, REM_reml,
                           REM_reml$tau2)
GAIC_REM_reml[b] = metaGAIC(this.data$d,
                              this.data$sigma2.d,
                              this.k, REM_reml,
                              REM_reml$tau2)

AIC_REM_dl[b] = metaAIC(this.data$d,
                         this.data$sigma2.d,
                         this.k, REM_dl,
                         REM_dl$tau2)
GAIC_REM_dl[b] = metaGAIC(this.data$d,
                            this.data$sigma2.d,
                            this.k, REM_dl,
                            REM_dl$tau2)

AIC_REM_sj[b] = metaAIC(this.data$d,
                         this.data$sigma2.d,
                         this.k, REM_sj,
                         REM_sj$tau2)
GAIC_REM_sj[b] = metaGAIC(this.data$d,
                            this.data$sigma2.d,
                            this.k, REM_sj,
                            REM_sj$tau2)

AIC_REM_he[b] = metaAIC(this.data$d,
                         this.data$sigma2.d,
```

```

        this.k, REM_he,
        REM_he$tau2)
GAIC_REM_he[b] = metaGAIC(this.data$d,
        this.data$sigma2.d,
        this.k, REM_he,
        REM_he$tau2)

#calcolo AICCEM
AIC_CEM[b]= metaAIC(this.data$d,
        this.data$sigma2.d,
        this.k, CEM, 0)

#calcolo le differenze dei vari
#AIC e GAIC da AICCEM
diff_GA_reml[b]=GAIC_REM_reml[b]-AIC_CEM[b]
diff_GA_dl[b]=GAIC_REM_dl[b]-AIC_CEM[b]
diff_GA_sj[b]=GAIC_REM_sj[b]-AIC_CEM[b]
diff_GA_he[b]=GAIC_REM_he[b]-AIC_CEM[b]

diff_AA_reml[b]=AIC_REM_reml[b]-AIC_CEM[b]
diff_AA_dl[b]=AIC_REM_dl[b]-AIC_CEM[b]
diff_AA_sj[b]=AIC_REM_sj[b]-AIC_CEM[b]
diff_AA_he[b]=AIC_REM_he[b]-AIC_CEM[b]
}
#chiusura ciclo b

#statistica Q
p_Q = length(Q[Q > 0.05])/n.rep
mean.Q =mean(Q)

#REML
p_GAIC_reml = length(diff_GA_reml[diff_GA_reml > 0])
            /n.rep
p_AIC_reml = length(diff_AA_reml[diff_AA_reml > 0])
            /n.rep

```

```
#DL
p_GAIC_dl = length(diff_GA_dl[diff_GA_dl > 0])/n.rep
p_AIC_dl = length(diff_GAreml[diff_AA_dl > 0])/n.rep

#SJ
p_GAIC_sj = length(diff_GA_sj[diff_GA_sj > 0])/n.rep
p_AIC_sj = length(diff_AA_sj[diff_AA_sj > 0])/n.rep

#HE
p_GAIC_he = length(diff_GA_he[diff_GA_he > 0])/n.rep
p_AIC_he = length(diff_AA_he[diff_AA_he > 0])/n.rep

#calcolo ora le medie dei vari AIC e GAIC
#che ho ottenuto
mean.AIC_CEM = mean(AIC_CEM)

mean.AIC_REM_reml=mean(AIC_REM_reml)
mean.AIC_REM_dl = mean(AIC_REM_dl)
mean.AIC_REM_sj = mean(AIC_REM_sj)
mean.AIC_REM_he = mean(AIC_REM_he)

mean.GAIC_REM_reml = mean(GAIC_REM_reml)
mean.GAIC_REM_dl = mean(GAIC_REM_dl)
mean.GAIC_REM_sj = mean(GAIC_REM_sj)
mean.GAIC_REM_he = mean(GAIC_REM_he)

#calcolo la media degli I^2 ottenuti
mean.I2 = mean(I2)

stat<- data.frame(p_Q, mean.Q, mean.I2, p_GAIC_reml,
                  p_GAIC_dl, p_GAIC_sj, p_GAIC_he,
                  p_AIC_reml, p_AIC_dl, p_AIC_sj,
```

```
    p_AIC_he, mean.AIC_CEM,
    mean.AIC_REM_reml, mean.AIC_REM_dl,
    mean.AIC_REM_sj, mean.AIC_REM_he,
    mean.GAIC_REM_reml, mean.GAIC_REM_dl,
    mean.GAIC_REM_sj, mean.GAIC_REM_he)

    save(stat, file=paste('statistiche', this.k, '_tau',
                        this.tau2, '_theta', this.theta,
                        '.RData', sep=''))

}
#chiusura ciclo h

}
#chiusura ciclo h

}
#chiusura ciclo j
```

---



# Bibliografía

- AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control* **19**, 716–723.
- ALLEN, L. A., WOOLFOLK, R. L., ESCOBAR, J. I., GARA, M. A. & HAMER, R. M. (2006). Cognitive-behavioral therapy for somatization disorder: A randomized controlled trial. *Archives of Internal Medicine* **166**, 1512–1518.
- ARNOLD, I. A., DE WAAL, M. W., EEKHOF, J. A., ASSENDELFT, W. J., SPINHOVEN, P. & VAN HEMERT, A. M. (2009). Medically unexplained physical symptoms in primary care: A controlled study on the effectiveness of cognitive-behavioral treatment by the family physician. *Psychosomatics* **50**, 515–524.
- BLEICHHARDT, G., TIMMER, B. & RIEF, W. (2004). Cognitive-behavioural therapy for patients with multiple somatoform symptoms—a randomised controlled trial in tertiary care. *Journal of Psychosomatic Research* **56**, 449–454.
- BLEICHHARDT, G., TIMMER, B. & RIEF, W. (2005). Hypochondriasis among patients with multiple somatoform symptoms—psychopathology and outcome of a cognitive-behavioral therapy. *Journal of Contemporary Psychotherapy* **35**, 239–249.
- BORENSTEIN, M., HEDGES, L., HIGGINS, J. & ROTHSTEIN, H. (2009). *Introduction to Meta-Analysis*. Wiley.
- BORENSTEIN, M., HEDGES, L. V., HIGGINS, J. P. T. & ROTHSTEIN, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* , 97–111.
- BOTELLA, J., HUEDO-MEDINA, T., SANCHEZ-MECA, J. & MARIN-MARTINEZ, F. (2006). Assessing heterogeneity in meta-analysis: Q statistic or  $i^2$  index? *Psychological Methods* **11**, 191–206.

- CINAR, O., J.UMBANHOWAR, J.D.HOEKSEMA & W.VIECHTBAUER. (2021). Using information-theoretic approaches for model selection in meta-analysis. *Research Synthesis Methods* , 537–556.
- COCHRAN, W. G. (1954). The combination of estimates from different experiments. *Biometrics* **10**, 101–129.
- DERSIMONIAN, R. & LAIRD, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188.
- D.O. STRAM, J. L. (1994). Variance components testing in the longitudinal mixed effects model. *Multivariate Behav Res.* **50**, 1171–1177.
- DU, H., JIANG, G. & Z.KE (2023). Bootstrap-based between-study heterogeneity tests in meta-analysis. *Multivariate Behav Res.* **58**, 484–503.
- EFRON, B. & TIBSHIRANI, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1**, 54–75.
- ESCOBAR, J. I., GARA, M. A., DIAZ-MARTINEZ, A. M., INTERIAN, A., WARMAN, M., ALLEN, L. A., WOOLFOLK, R. L., JAHN, E. & RODGERS, D. (2007). Effectiveness of a time-limited cognitive behavior therapy-type intervention among primary care patients with medically unexplained symptoms. *The Annals of Family Medicine* **5**, 328–335.
- HARTUNG, J. (1999). An alternative method for meta-analysis. *Biometrical Journal* **41**, 901–916.
- HEDGES, L. V. & OLKIN, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando: Academic Press.
- HIGGINS, J. P. T. & THOMPSON, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statist. Med* **21**, 1539–1558.
- KLEINSTÄUBER, M., WITTHÖFT, M. & HILLER, W. (2011). Efficacy of short-term psychotherapy for multiple medically unexplained physical symptoms: A meta-analysis. *Clinical Psychology Review* , 146–160.
- KOLK, A., SCHAGEN, S. & HANEWALD, G. (2004). Multiple medically unexplained physical symptoms and health care utilization: outcome of psychological intervention

- and patient-related predictors of change. *Journal of psychosomatic research* **57**, 379–389.
- KONTOPANTELIS, E. & REEVES, D. (2012a). Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A comparison between dersimonian–laird and restricted maximum likelihood. *Statistical Methods in Medical Research* **21**, 657–659.
- KONTOPANTELIS, E. & REEVES, D. (2012b). Performance of statistical methods for meta-analysis when true study effects are non-normally distributed: A simulation study. *Statistical Methods in Medical Research* **21**, 409–426.
- LARISCH, A., SCHWEICKHARDT, A., WIRSCHING, M. & FRITZSCHE, K. (2004). Psychosocial interventions for somatizing patients by the general practitioner: A randomized controlled trial. *Journal of Psychosomatic Research* **57**, 507–514.
- LOPEZ-LOPEZ, J., MARIN-MARTINEZ, F., SANCHEZ-MECA, J., DEN NOORTGATE, W. V. & VIECHTBAUER, W. (2014). Estimation of the predicitive power of the model in mixed-effects meta-regression: A simulation study. *British journal of Mathematical and Statistical Psicology* **67**, 30–48.
- LUPKE, U., EHLERT, U. & HELLHAMMER, D. (1996). Verhaltensmedizin im allgemeinkrankenhaus: Verlaufsuntersuchung an patienten mit somatoformen störungen. *Verhaltenstherapie* **6**, 22–32.
- MARTIN, A., RAUH, E., FICHTER, M. & RIEF, W. (2007). A one-session treatment for patients suffering from medically unexplained symptoms in primary care: A randomized clinical trial. *Psychosomatics* **48**, 294–303.
- R CORE TEAM (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- RAUDENBUSH, S. (2009). Analyzing effect sizes: Random-effects models. *The Handbook of Research Synthesis and Meta-analysis* , 295–315.
- RUBIO-APARICIO, MARÍA MARÍN-MARTÍNEZ, F. S.-M. J. L.-L. J. A. (2018). A methodological review of meta-analyses of the effectiveness of clinical psychology treatments. *Behavior Research Methods* **50**, 2057–2073.
- SCHWEICKHARDT, A., LARISCH, A., WIRSCHING, M. & FRITZSCHE, K. (2007). Short-term psychotherapeutic interventions for somatizing patients in the general

- hospital: a randomized controlled study. *Psychotherapy and psychosomatics* **76**, 339–346.
- SIDIK, K. & JONKMAN, J. (2005). Simple heterogeneity variance estimation for meta-analysis. *Appl. Statist* , 367–384.
- SIDIK, K. & JONKMAN, J. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Stat Med* , 1964–81.
- SMITH, R. C., LYLES, J. S., GARDINER, J. C., SIRBU, C., HODGES, A., COLLINS, C., DWAMENA, F. C., LEIN, C., GIVEN, C. W., GIVEN, B. et al. (2006). Primary care clinicians treat patients with medically unexplained symptoms: a randomized controlled trial. *Journal of general internal medicine* **21**, 671–677.
- SPECKENS, A. E., VAN HEMERT, A. M., SPINHOVEN, P., HAWTON, K. E., BOLK, J. H. & ROOIJMANS, H. G. (1995). Cognitive behavioural therapy for medically unexplained physical symptoms: a randomised controlled trial. *Bmj* **311**, 1328–1332.
- SUMATHIPALA, A., HEWEGE, S., HANWELLA, R. & MANN, A. (2000). Randomized controlled trial of cognitive behaviour therapy for repeated consultations for medically unexplained complaints: a feasibility study in sri lanka. *Psychological medicine* **30**, 747–757.
- SUMATHIPALA, A., SIRIBADDANA, S., ABEYSINGHA, M., DE SILVA, P., DEWEY, M., PRINCE, M. & MANN, A. (2008). Cognitive-behavioural therapy v. structured care for medically unexplained symptoms: randomised controlled trial. *The British Journal of Psychiatry* **193**, 51–59.
- VAN DER FELTZ-CORNELIS, C. M., VAN OPPEN, P., ADÈR, H. J. & VAN DYCK, R. (2006). Randomised controlled trial of a collaborative care model with psychiatric consultation for persistent medically unexplained symptoms in general practice. *Psychotherapy and psychosomatics* **75**, 282–289.
- VIECHTBAUER, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* **30**, 261–293.
- VIECHTBAUER, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software* **36**, 1–48.

- 
- WALSH, S. R. B. (2002). *Emotion-focused couples therapy as a treatment of somatoform disorders: An outcome study*. Ph.D. thesis, Virginia Polytechnic Institute and State University.
- YANG, K. & H.Y. KWAN, Z. YU, T. T. (2020). Model selection between the fixed effects model and the random-effects model in meta-analysis. *Statistic and its interface* **13**, 501–510.
- ZABY, A., HEIDER, J. & SCHRÖDER, A. (2008). Warten, entspannung oder verhaltenstherapie. *Zeitschrift für Klinische Psychologie und Psychotherapie* **37**, 15–23.

