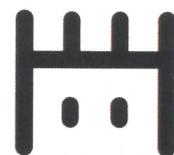




# Minicurso Uma introdução à Meta-Análise

Maria de Fátima Brilhante



# Uma Introdução à Meta-Análise

MARIA DE FÁTIMA BRILHANTE

Faculdade de Ciências e Tecnologia da Universidade dos Açores  
Centro de Estatística e Aplicações da Universidade de Lisboa

Edições SPE

Copyright © 2017 - Maria de Fátima Brilhante

Todos os direitos reservados

**FICHA TÉCNICA:**

**Livro:** Uma Introdução à Meta-Análise

**Autor:** Maria de Fátima Brilhante

Faculdade de Ciências e Tecnologia da Universidade dos Açores  
Centro de Estatística e Aplicações da Universidade de Lisboa

**Editora:** Sociedade Portuguesa de Estatística

**Impressão:** Instituto Nacional de Estatística

**Tiragem:** 300 exemplares

**ISBN:**

**Depósito Legal:**

# Prefácio

A primeira vez que ouvi falar sobre Meta-Análise foi em 2009 quando, a convite do Professor Dinis Pestana, participei no júri das provas de doutoramento do Professor Fernando Sequeira, que defendeu a tese intitulada *Meta-Análise: Harmonização de Testes usando os Valores de Prova*. Desde então comecei a interessar-me pelo assunto, pelo que foi com bastante agrado que aceitei, e agradeço, o convite da Sociedade Portuguesa de Estatística para lecionar um minicurso sobre Meta-Análise no seu XXIII Congresso.

A Meta-Análise, na sua formalização atual, existe desde 1976, embora a primeira tentativa que se conhece para resolver um problema através de uma meta-análise date do início do século XIX, com a aplicação do método dos mínimos quadrados pelo matemático Adrien-Marie Legendre. Nos últimos quarenta anos, a quantidade de artigos publicados com aplicações e propostas de métodos estatísticos para a Meta-Análise tem aumentado exponencialmente, como resposta à necessidade de combinar, harmonizar ou sintetizar toda a informação que se tem vindo a acumular sobre determinados problemas em diferentes áreas da Ciência.

Muito do trabalho que tenho feito na área tem sido em colaboração com outros colegas, e mais de perto com o Professor Dinis Pestana, a quem agradeço os comentários que fez a uma primeira versão deste texto. Também quero agradecer ao Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL) todo o apoio que me tem

dado ao longo destes anos para a minha formação contínua na área da Estatística.

Antes de começarmos esta viagem juntos, gostaria de chamar a atenção para o facto do texto aqui apresentado pretender ser apenas uma introdução ao tema, estando longe de ser exaustivo. O objetivo fundamental deste texto é dar a conhecer os principais procedimentos e ferramentas usados em Meta-Análise, numa perspetiva mais clássica.

Por fim quero dedicar este livro ao Professor Dinis Pestana porque tem contribuído de forma significativa para a investigação, o ensino e a divulgação da Estatística em Portugal.

Ponta Delgada, 17 de setembro de 2017

Maria de Fátima Brilhante

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Combinar Informação</b>	<b>7</b>
2.1	Exemplos de Tamanhos do Efeito . . . . .	8
2.1.1	Tamanhos do Efeito Baseados em Valores Médios	9
2.1.2	Tamanhos do Efeito Baseados em Correlações . .	12
2.1.3	Tamanhos do Efeito Baseados em Proporções . .	13
<b>3</b>	<b>Modelos em Meta-Análise</b>	<b>17</b>
3.1	Modelo de Efeitos Fixos . . . . .	17
3.2	Modelo de Efeitos Aleatórios . . . . .	20
3.2.1	Estimação do Parâmetro de Heterogeneidade . .	23
3.2.2	Impacto da Heterogeneidade numa Meta-Análise	30
3.3	Métodos Baseados em Valores- $p$ . . . . .	46
3.3.1	Valores- $p$ Generalizados e Aleatórios . . . . .	52
<b>4</b>	<b>Representações Gráficas para a Meta-Análise</b>	<b>59</b>
4.1	Gráfico Floresta . . . . .	59
4.2	Gráfico Radial . . . . .	61

4.3	Gráfico de Baujat . . . . .	64
4.4	Gráfico de L'Abbé . . . . .	67
<b>5</b>	<b>Outros Métodos em Meta-Análise</b>	<b>69</b>
5.1	Análise de Subgrupos . . . . .	69
5.2	Meta-Regressão . . . . .	81
5.3	Meta-Análise Cumulativa . . . . .	87
<b>6</b>	<b>Viés de Publicação</b>	<b>93</b>
6.1	Gráfico Funil . . . . .	94
6.2	Testes para o Viés de Publicação . . . . .	97
6.3	Número de Estudos na Gaveta . . . . .	103
6.4	Métodos para a Correção do Viés de Publicação . . . . .	108
<b>7</b>	<b>Calibrar e Combinar Evidência</b>	<b>123</b>
7.1	Combinar Evidência . . . . .	128
<b>8</b>	<b>Podemos Confiar numa Meta-Análise?</b>	<b>141</b>
	<b>Bibliografia</b>	<b>147</b>
	<b>Índice Remissivo</b>	<b>157</b>

# Capítulo 1

## Introdução

*Science is a cumulative process.*

Shoemaker *et al.* (2003)

Desde o início do século XX até 1990 alguns livros (e pediatras) aconselhavam os pais a deitarem os seus filhos pequenos de barriga para baixo. Durante esse período observaram-se várias mortes não explicadas de crianças pequenas, aparentemente saudáveis, durante o seu sono, fenómeno que passou a ser conhecido pela síndrome da morte súbita infantil (SIDS), ou morte no berço.

Autópsias realizadas em 1944 pelo patologista norte-americano Abramson revelaram que dois terços das crianças que morreram devido a mecanismos de sufocamento, foram encontradas deitadas no berço sobre a barriga. Como os resultados de Abramson foram corroborados por outros relatórios realizados no Reino Unido e Austrália, iniciou-se em 1945 uma campanha para informar as pessoas sobre o risco da SIDS associada à posição de dormir. Mas, as conclusões das experiências realizadas pelo pediatra Woolley (1945), entre as quais a rejeição da tese de sufocamento na SIDS, como consequência da posição de dormir, ditaram o cancelamento da referida campanha. Contudo, no início de 1990 investigadores observaram uma redução de pelo menos 50% no

risco da SIDS, altura em que muitos pais passaram a deitar os seus filhos pequenos sobre as costas.

As informações contraditórias sobre este assunto motivaram Gilbert *et al.* (2004) a realizar uma meta-análise, tendo esta sido baseada em 40 estudos comparativos sobre a associação entre a SIDS e a posição de dormir. Uma das conclusões surpreendentes dessa meta-análise é que já no início da década de 70 havia evidência acumulada no sentido de um maior risco para a SIDS para a posição de dormir sobre a barriga, quando comparado com o de outras posições. Nas conclusões do artigo as autoras afirmam que

*Advice to put infants to sleep on the front for nearly half a century was contrary to evidence available from 1970 that this was likely to be harmful. Systematic review of preventable risk factors for SIDS from 1970 would have led to earlier recognition of the risks of sleeping on the front and might have prevented over 10 000 infant deaths in the UK and at least 50 000 in Europe, the USA and Australasia.*

Este exemplo foi um de vários apresentados por Sir Ian Chalmers na sua conferência de 2005 intitulada “*The scandalous failure of scientists to cumulate scientifically*”.<sup>1</sup> Na sua comunicação, Chalmers procurou chamar a atenção para a importância da acumulação de evidência na construção do conhecimento científico atual, e para o eventual impacto negativo que uma ausência de acumulação, sobretudo na área da Saúde, pode ter na vida das pessoas comuns.

## O que é a Meta-Análise?

Um princípio fundamental em Ciência é a reprodutibilidade, ou seja, a possibilidade de uma experiência (ou estudo) ser repetida pelo mesmo investigador ou por outros de forma independente. Porém, é possível

---

<sup>1</sup>Sir Ian Chalmers é um investigador na área da Saúde, sendo um dos fundadores do *Cochrane Collaboration* e o coordenador do *James Lind Initiative*.

que no meio de um conjunto de experiências realizadas com o objetivo de responder a questões semelhantes, não se chegue às mesmas conclusões. Certamente que cada um de nós já foi confrontado pelos média com informações contraditórias sobre um determinado assunto. Nestes casos surge-nos a dúvida, em quem devemos acreditar?

Uma maneira de abordarmos este tipo de situações é procurarmos combinar, de forma adequada, toda a informação relevante disponível sobre o tópico de interesse, por forma a obtermos uma conclusão global que nos dê mais confiança. É precisamente neste contexto que a Meta-Análise nos pode ser útil.

Mas afinal, o que é a Meta-Análise? A Meta-Análise define-se basicamente como sendo a reanálise das análises estatísticas com o objetivo de combinar, harmonizar ou sintetizar a informação disponível sobre um determinado problema. O termo foi usado pela primeira vez por Gene Glass (1976) para designar este tipo de procedimento. Nas suas palavras,

*Meta-analysis refers to the analysis of analysis. I use it to refer to the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature.*

A Meta-Análise surge assim associada às ditas revisões sistemáticas que diferem das chamadas revisões narrativas. As revisões narrativas procuram fazer uma revisão da literatura com o objetivo de manter essencialmente a comunidade científica atualizada sobre um determinado tema, sem que haja necessariamente a preocupação de uma seleção criteriosa dos estudos, nem da integração dos resultados estatísticos existentes.

Como curiosidade histórica, o envolvimento de Glass em questões de meta-análise deveu-se inicialmente à sua irritação com os comentários do psicólogo Hans J. Eysenck (1916-1997) sobre a ineficácia da psicoterapia, quando o próprio Glass tinha beneficiado com o tratamento. Eysenck afirmava não existirem dados que suportassem a sua eficácia. Determinado a mostrar que o psicólogo estava errado, Glass conseguiu reunir um conjunto considerável de estudos que procuravam comparar o efeito da psicoterapia com o efeito do não tratamento (grupo de controlo), sendo que para cada estudo usou como medida uma diferença de médias estandardizada. As conclusões dessa meta-análise, apresentadas no encontro da *American Educational Research Association* de 1976, foram que havia uma vantagem média substancial com a psicoterapia e, por conseguinte, evidência a favor da eficácia do tratamento. Todavia, a resposta de Eysenck não se fez esperar muito quando publica em 1978 o artigo “*An Exercise in Mega-Silliness*”, onde ridiculariza o trabalho de Glass com o comentário “*garbage in-garbage out*”, face à qualidade duvidosa de alguns estudos usados.

Apesar da formalização da Meta-Análise se dever a Glass, há muito que se conheciam trabalhos que procuraram combinar ou sintetizar a informação espalhada por diferentes estudos. A primeira meta-análise conhecida surge em 1805 no livro “*Nouvelles méthodes pour la détermination des orbites des comètes*” de Adrien-Marie Legendre (1752-1833), onde surge pela primeira vez o método dos mínimos quadrados.<sup>2</sup> Como exemplo de aplicação do método, Legendre usa-o para estimar o comprimento do arco de meridiano de Paris entre Dunkirk e Barcelona, usando para o efeito a combinação dos dados recolhidos durante a expedição de 1795, encomendada a François-André Méchain e Jean-Baptiste Joseph Delambre pela assembleia constituinte francesa. Já no século XX, Karl Pearson realiza em 1904 uma síntese dos resultados sobre imunidade e mortalidade em 11 estudos, realizados na Índia e África de Sul, sobre a infeção por febre tifóide. Mais tarde, Tippett (1931), Fisher (1932),

---

<sup>2</sup>A publicação de Legendre provocou uma disputa com Carl Friedrich Gauss (1777-1855) sobre os direitos de autor, porque este último afirmava ter inventado o método em 1795.

Cochran (1937) e Yates e Cochran (1938) combinaram a informação de experiências agrícolas realizadas com o propósito de estimar o efeito tratamento e testar a sua significância.

A importância atual da Meta-Análise nas revisões sistemáticas ligadas, por exemplo, à área da Saúde é de tal ordem, que foi criada uma organização não governamental e não lucrativa chamada *Cochrane Collaboration*, que integra uma rede de investigadores, profissionais de saúde, pacientes, cuidadores e voluntários de vários países, inclusive de Portugal. O lema da organização é “*Trusted evidence. Informed decisions. Better health.*”. Associada a esta organização existe a *Cochrane Library*, que consiste numa coleção de bases de dados contendo informações sobre estudos considerados de elevada qualidade, sobre os mais variados temas, e que poderão ser usadas pelos investigadores em futuras meta-análises. A *Cochrane Collaboration* fornece também um conjunto de linhas orientadoras a ser seguido pelos investigadores que desejem fazer revisões sistemáticas, disponibilizando igualmente um *software* estatístico (RevMan5) para a realização de meta-análises. Outras organizações semelhantes têm aparecido, como a *Campbell Collaboration* que se preocupa com revisões sistemáticas nas áreas das Ciências Sociais, Ciências da Educação e Psicologia.

As principais etapas numa meta-análise podem ser resumidas em:

1. Formulação do problema;
2. Seleção dos estudos para a meta-análise (após o que deve ser uma pesquisa exaustiva em diferentes revistas de especialidade, bases de dados, etc.);
3. Escolha do tamanho do efeito adequado para descrever os resultados numéricos de cada estudo selecionado;
4. Determinação do efeito combinado, ou efeito síntese, e da sua precisão;
5. Divulgação dos resultados finais.

Nos capítulos seguintes daremos uma panorâmica geral sobre as etapas 3 e 4 mencionadas acima, e sobre alguns procedimentos comuns e problemas que podem surgir numa meta-análise, e como devemos lidar com estes. Convém, contudo, salientarmos que a etapa 2 é certamente uma das mais importantes em Meta-Análise, podendo nalguns casos se revelar uma tarefa espinhosa. Sobre este assunto diremos que a seleção dos estudos costuma ser da responsabilidade do investigador, pois é este que define, em função das questões que pretende responder com a análise, um conjunto sólido de critérios de inclusão e exclusão dos estudos. Porém, é fundamental que os critérios usados garantam que os estudos selecionados sejam minimamente comparáveis entre si. Por exemplo, para auxiliar os investigadores na etapa 2, a *Cochrane Collaboration* disponibiliza uma espécie de guião de critérios a ser tido em conta para a inclusão/exclusão dos estudos identificados. Para mais informações sobre a etapa 2 consultar, por exemplo, Cooper *et al.* (2009).

## Capítulo 2

# Combinar Informação

*Meta-analysis can produce strong evidence where at first sight there seems to be only weak evidence. It can turn long CIs into short ones (well sort of), find answers in what looks like a mess, and settle heated controversies.*

Geoff Cumming (2012)

A Meta-Análise, como análise estatística das sínteses de trabalhos de investigação, recorre a vários métodos estatísticos conhecidos, mas também a outros que têm sido propostos especificamente para a área. Trata-se de uma ferramenta de investigação útil em áreas como a Medicina, Psicologia, Ciências do Ambiente, Ciências Sociais e Ciências da Educação. Com uma meta-análise podemos, por exemplo, harmonizar respostas parciais contraditórias, ou encontrar uma significância global que amostras parciais de dimensão pequena não revelam, geralmente por falta de potência estatística dos testes usados.

Na maioria das meta-análises efetuadas procura-se, em primeiro lugar, descrever os resultados de cada estudo selecionado em termos de um índice numérico que estime o chamado *tamanho do efeito*. Por exemplo, em ensaios clínicos onde estão envolvidos um grupo de tratamento e um grupo de controlo, o tamanho do efeito traduz normalmente

a magnitude do efeito tratamento. Mas, mais geralmente, um tamanho do efeito quantificará a relação entre duas variáveis ou a diferença entre dois grupos. No entanto, poderão ser usados índices que não quantifiquem relações entre variáveis ou diferenças entre grupos, como uma simples proporção ou uma média.

Um dos objetivos da combinação ou síntese dos tamanhos do efeito numa meta-análise é procurar aumentar a precisão da estimativa do parâmetro de interesse.<sup>1</sup> Vejamos um exemplo simples.

Suponhamos que  $K$  estudos independentes foram realizados com o intuito de estimar uma média populacional  $\mu$  em cada estudo. No estudo  $k$ ,  $\mu$  foi estimado pela média amostral  $\bar{X}_k$ , calculada à base de uma amostra aleatória de dimensão  $n_k$ ,  $k = 1, \dots, K$ . Usando as diferentes estimativas fornecidas pelos estudos, podemos estimar  $\mu$  através da combinação

$$\hat{\mu}_{\text{comb}} = \frac{n_1 \bar{X}_1 + \dots + n_K \bar{X}_K}{N},$$

em que  $N = n_1 + \dots + n_K$ .<sup>2</sup> Se assumirmos que a variância populacional é  $\sigma^2$  para cada estudo,

$$\text{Var}(\hat{\mu}_{\text{comb}}) = \sigma_{\hat{\mu}_{\text{comb}}}^2 = \frac{\sigma^2}{N} < \frac{\sigma^2}{n_k} = \text{Var}(\bar{X}_k), \quad k = 1, \dots, K,$$

ou seja, verificamos que a precisão do estimador  $\hat{\mu}_{\text{comb}}$  é maior que a de qualquer estimador individual  $\bar{X}_k$ ,  $k = 1, \dots, K$ . No entanto, chamamos a atenção para o facto de poder não se observar um aumento na precisão em cenários mais complexos.

## 2.1 Exemplos de Tamanhos do Efeito

De entre os tamanhos do efeito mais usados em Meta-Análise destacam-se uma diferença de valores médios (simples ou standar-

---

<sup>1</sup>Note-se que se está a cometer um abuso de linguagem, porque as estimativas não têm uma precisão associada, mas sim os estimadores que as originam.

<sup>2</sup>Existem outras formas de combinar as diferentes estimativas de  $\mu$ .

dizada), um risco relativo, um rácio de vantagens, mais vulgarmente conhecido por *odds ratio*, e um coeficiente de correlação. Doravante, usaremos a letra grega  $\theta$  para representar um tamanho do efeito genérico.

### 2.1.1 Tamanhos do Efeito Baseados em Valores Médios

Consideremos duas variáveis aleatórias  $X_1$  e  $X_2$  contínuas com valores médios  $\mu_1$  e  $\mu_2$ , e variâncias  $\sigma_1^2$  e  $\sigma_2^2$ , respetivamente.<sup>3</sup> Para um tamanho do efeito  $\theta$  baseado em médias, podemos usar a diferença de valores médios (simples)

$$\theta = \text{MD} = \mu_1 - \mu_2, \quad (2.1)$$

ou a diferença de valores médios estandardizada

$$\theta = \text{SMD} = \frac{\mu_1 - \mu_2}{\sigma}, \quad (2.2)$$

onde  $\sigma$  representa o desvio padrão comum das populações.<sup>4</sup>

Faz sentido usarmos o tamanho do efeito (2.1) quando todos os estudos medem de forma igual os mesmos resultados, ou seja, quando a escala de medição usada é a mesma em cada estudo. Por sua vez, um tamanho do efeito do tipo (2.2) deverá ser usado quando todos os estudos medem os mesmos resultados, mas fazem-no recorrendo a escalas de medição diferentes. Como exemplo desta última situação, suponhamos que uma meta-análise incide sobre um conjunto de estudos cujos resultados são o grau de severidade da depressão, grau esse avaliado nalguns estudos com a escala de Hamilton e noutros com a escala de Montgomery-Åsberg.

---

<sup>3</sup>Apesar das notações  $X_1$  e  $X_2$  representar usualmente variáveis aleatórias, neste texto também as usaremos para representar populações. Também poderemos usar, em alternativa, a notação  $E(X)$  para representar o valor médio de uma variável aleatória  $X$ .

<sup>4</sup>Assume-se geralmente que as populações são homocedásticas, isto é,  $\sigma_1^2 = \sigma_2^2$ , embora esta condição possa ser relaxada.

Consideremos, em primeiro lugar, que as populações  $X_1$  e  $X_2$  são independentes. Sejam  $(X_{11}, \dots, X_{1n_1})$  e  $(X_{21}, \dots, X_{2n_2})$  amostras aleatórias de  $X_1$  e  $X_2$ , respetivamente, e denotemos por  $\bar{X}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} X_{kj}$  e  $S_k^2 = \frac{1}{n_k-1} \sum_{j=1}^{n_k} (X_{kj} - \bar{X}_k)^2$ ,  $k = 1, 2$ .

O estimador natural do parâmetro  $\theta$  definido em (2.1) é

$$\hat{\theta} = \bar{X}_1 - \bar{X}_2, \quad (2.3)$$

sendo um estimador não enviesado de  $\theta$ , com variância

$$\text{Var}(\hat{\theta}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Como, geralmente, as variâncias populacionais são desconhecidas, a variância de  $\hat{\theta}$  pode ser estimada por

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}.$$

No caso particular de  $\sigma_1^2 = \sigma_2^2$ ,  $\widehat{\text{Var}}(\hat{\theta}) = S_p^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)$ , em que  $S_p^2 = [(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2] / (n_1 + n_2 - 2)$ .

Em relação ao parâmetro  $\theta$  definido em (2.2), se assumirmos que as populações são homocedásticas, este pode ser estimado pela:

- **Medida de Cohen**

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_p}.$$

Esta medida (Cohen, 1969) tende a sobrestimar o valor de  $\theta$ , especialmente com amostras de dimensão pequena, sendo por isso recomendado em alternativa o uso da medida de Hedges (Hedges, 1981) que procura corrigir o seu viés;

- **Medida de Hedges**

$$g = \left( 1 - \frac{3}{4N - 9} \right) d,$$

em que  $N = n_1 + n_2$ ;<sup>5</sup>

- **Medida de Glass**

$$\Delta = \frac{\bar{X}_1 - \bar{X}_2}{S_2},$$

em que  $S_2$  é o desvio padrão do grupo de controle (Glass *et al.*, 1981).<sup>6</sup>

Para as três medidas acima tem-se que

$$\widehat{\text{Var}}(d) \approx \frac{N}{n_1 n_2} + \frac{d^2}{2N},$$

$$\widehat{\text{Var}}(g) \approx \left( \frac{N}{n_1 n_2} + \frac{g^2}{2N} \right) \left( 1 - \frac{3}{4N-9} \right)^2$$

e

$$\widehat{\text{Var}}(\Delta) \approx \frac{N}{n_1 n_2} + \frac{\Delta^2}{2(n_2 - 1)}.$$

Suponhamos agora que  $((X_{11}, X_{21}), \dots, (X_{1n}, X_{2n}))$  é uma amostra aleatória da população contínua  $(X_1, X_2)$ , em que  $\text{Var}(X_1) = \sigma_1^2$ ,  $\text{Var}(X_2) = \sigma_2^2$  e  $\rho = \text{Cor}(X_1, X_2)$ , em que  $\rho$  é o coeficiente de correlação populacional. Neste caso, o estimador (2.3) tem variância

$$\text{Var}(\hat{\theta}) = \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{n}.$$

Desconhecendo-se os valores de  $\sigma_1^2$  e  $\sigma_2^2$ , a variância de  $\hat{\theta}$  pode ser estimada por

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{S_1^2 + S_2^2 - 2rS_1S_2}{n},$$

---

<sup>5</sup>Para um outro estimador de  $\sigma^2$  diferente de  $S_p^2$  usado na medida de Cohen, considera-se para a medida de Hedges  $g = J \times d$ , onde  $J = \left(1 - \frac{3}{4df-1}\right)$  e  $df$  representa o número de graus de liberdade associado à estimação de  $\sigma^2$ .

<sup>6</sup>A justificação para  $\hat{\sigma} = S_2$  é porque se acredita que o grupo de controle fornece uma estimativa mais fiável do desvio padrão comum, do que aquela que é obtida a partir do grupo experimental. Por outro lado, se se assumir populações heterocedásticas ( $\sigma_1^2 \neq \sigma_2^2$ ), há quem prefira utilizar a medida de Glass para estimar a diferença de valores médios estandardizada.

onde  $r = \hat{\rho}$  representa o coeficiente de correlação de Pearson definido por

$$r = \frac{\sum_{j=1}^n (X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2)}{\sqrt{\sum_{j=1}^n (X_{1j} - \bar{X}_1)^2 \sum_{j=1}^n (X_{2j} - \bar{X}_2)^2}}.$$

Por outro lado, se atendermos a que a medida de dispersão mais reportada em estudos que recorrem a amostras emparelhadas é o desvio padrão das diferenças das medições, podemos considerar como estimador de  $\theta$  definido em (2.2)

$$\hat{\theta} = \frac{\bar{X}_1 - \bar{X}_2}{S},$$

onde  $S = S_D / \sqrt{2(1-r)}$  e  $S_D$  é o desvio padrão empírico das diferenças  $D_j = X_{1j} - X_{2j}$ ,  $j = 1, \dots, n$ .

Neste caso, a variância de  $\hat{\theta}$  será estimada por

$$\widehat{\text{Var}}(\hat{\theta}) = \left( \frac{1}{n} + \frac{\hat{\theta}^2}{2n} \right) 2(1-r).$$

### 2.1.2 Tamanhos do Efeito Baseados em Correlações

Para uma população contínua  $(X_1, X_2)$ , o tamanho do efeito baseado no coeficiente de correlação populacional poderá ser considerado, ou seja,

$$\theta = \rho.$$

Com base numa amostra aleatória  $((X_{11}, X_{21}), \dots, (X_{1n}, X_{2n}))$ , o parâmetro anterior é naturalmente estimado pelo coeficiente de correlação de Pearson. No entanto, para efeitos de inferência, é usual considerar-se em alternativa o parâmetro

$$\zeta = \frac{1}{2} \ln \left( \frac{1+\rho}{1-\rho} \right),$$

porque o estimador

$$Z = \hat{\zeta} = \frac{1}{2} \ln \left( \frac{1+r}{1-r} \right)$$

possui distribuição amostral assintótica normal, com variância estimada por

$$\widehat{\text{Var}}(Z) = \frac{1}{n-3}.$$

### 2.1.3 Tamanhos do Efeito Baseados em Proporções

Quando os estudos reportam resultados para duas variáveis binárias, os tamanhos do efeito mais usados são um risco relativo, um *odds ratio* e uma diferença de proporções. Para os exemplificarmos, suponhamos que uma meta-análise se baseará numa coleção de estudos que procuraram comparar as probabilidades de sucesso em duas populações independentes.<sup>7</sup> Para cada estudo a informação é organizada numa tabela de contingência do tipo

Grupo	Sucesso		Total
	Sim	Não	
1	$a$	$b$	$n_1 = a + b$
2	$c$	$d$	$n_2 = c + d$

Sejam  $p_1$  e  $p_2$  as probabilidades de sucesso nas populações subjacentes aos grupos (amostras) 1 e 2, respetivamente, com  $0 < p_1 < 1$  e  $0 < p_2 < 1$ .

- **Risco Relativo**

O quociente

$$\theta = \text{RR} = \frac{p_1}{p_2}$$

é chamado risco relativo e é estimado por

$$\widehat{\theta} = \frac{\widehat{p}_1}{\widehat{p}_2} = \frac{a/n_1}{c/n_2} \quad (c \neq 0).$$

Sabe-se que a distribuição amostral assintótica de  $\ln \widehat{\theta}$  é normal, com variância estimada por

$$\widehat{\text{Var}}(\ln \widehat{\theta}) = \frac{1}{a} - \frac{1}{n_1} + \frac{1}{c} - \frac{1}{n_2} \quad (a, c \neq 0).$$

<sup>7</sup>Entende-se aqui por sucesso a realização de um mesmo acontecimento de interesse em ambas as populações.

- **Odds Ratio**

O *odds ratio* a favor de sucesso na população subjacente ao grupo 1 em relação à do grupo 2 é dado por

$$\theta = \text{OR} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)},$$

sendo este valor estimado por

$$\hat{\theta} = \frac{a/b}{c/d} = \frac{ad}{bc} \quad (b, c \neq 0).$$

A distribuição amostral assintótica de  $\ln \hat{\theta}$  também é normal, com variância estimada por

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \quad (a, b, c, d \neq 0).$$

- **Diferença de Proporções**

Apesar de uma diferença de proporções, ou diferença de riscos, ter uma interpretação menos intuitiva do que a de um risco relativo ou a de um *odds ratio*, pode ser igualmente usada como tamanho do efeito. Define-se por

$$\theta = \text{RD} = p_1 - p_2,$$

sendo estimada por

$$\hat{\theta} = \hat{p}_1 - \hat{p}_2 = \frac{a}{n_1} - \frac{c}{n_2}.$$

Para amostras de dimensão elevada, assume-se que a distribuição de  $\hat{\theta}$  é normal, sendo a variância estimada por

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}.$$

*Observação 2.1.* Quando numa tabela de contingência pelo menos uma das frequências  $a$ ,  $b$ ,  $c$  ou  $d$  é nula, podemos obter estimativas infinitas ou indeterminadas para um risco relativo ou para um *odds ratio*. Existem algumas estratégias para lidarmos com estas situações numa meta-análise. As mais utilizadas são as seguintes:

- i) Eliminar da análise os estudos que contêm frequências observadas nulas;
- ii) Não eliminar os estudos com frequências observadas nulas e, para todos estes, considerar um fator de correção de continuidade, ou seja, adicionar um pequeno incremento a cada frequência observada (por exemplo, 0.01, 0.1 ou 0.5).<sup>8</sup> A prática comum consiste em adicionar o valor 0.5.

Por exemplo, Sweeting *et al.* (2004) propõem diferentes fatores de correção de continuidade para os dois grupos, mais concretamente os incrementos

$$\text{incr}_{\text{Grupo 1}} = \frac{n_1}{n_1 + n_2} \quad \text{e} \quad \text{incr}_{\text{Grupo 2}} = \frac{n_2}{n_1 + n_2}.$$

Repare-se que se nas fórmulas anteriores  $n_1 = n_2$ ,  $\text{incr}_{\text{Grupo 1}} = \text{incr}_{\text{Grupo 2}} = 0.5$ .<sup>9</sup>

*Observação 2.2.* Um tamanho do efeito também baseado em proporções, embora menos usado, é a diferença de arcos senos. É definida por

$$\theta = \text{ASD} = \arcsen \sqrt{p_1} - \arcsen \sqrt{p_2},$$

sendo estimada por

$$\hat{\theta} = \arcsen \sqrt{\hat{p}_1} - \arcsen \sqrt{\hat{p}_2} = \arcsen \sqrt{\frac{a}{n_1}} - \arcsen \sqrt{\frac{c}{n_2}}.$$

---

<sup>8</sup>Também se pode aplicar o mesmo fator de correção de continuidade a todos os outros estudos para os quais não se observe frequências nulas.

<sup>9</sup>Se se desejar evitar o uso de fatores de correção de continuidade no cálculo de um *odds ratio*, pode usar-se em alternativa o *odds ratio* de Peto (Yusuf *et al.*, 1985).

A variância de  $\hat{\theta}$  é estimada por

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{4} \left( \frac{1}{n_1} + \frac{1}{n_2} \right).$$

Após a escolha do tamanho do efeito adequado para descrever os resultados dos estudos selecionados, a etapa seguinte numa meta-análise passa por combinar os tamanhos do efeito estimados nos diferentes estudos num único valor, ou seja, num *efeito síntese* ou *efeito global combinado*.<sup>10</sup> A fórmula mais usada para combinar tamanhos do efeito em Meta-Análise deve-se a Cochran (1937, 1954).

Sejam  $K$  estudos independentes e seja  $\theta_k$  o verdadeiro valor do tamanho do efeito associado ao estudo  $k$ , sendo este estimado por  $\hat{\theta}_k$ ,  $k = 1, \dots, K$ . A estimativa para o efeito síntese  $\theta$  será dada pela média ponderada

$$\hat{\theta} = \frac{\sum_{k=1}^K \alpha_k \hat{\theta}_k}{\sum_{k=1}^K \alpha_k},$$

onde  $\alpha_k$  representa o peso a atribuir ao estudo  $k$ .

O uso de uma média ponderada, em vez de uma média simples, na combinação dos tamanhos do efeito é justificado pelo facto de haver estudos que, devido às suas características, produzem estimativas mais precisas e que, por isso, devem contribuir com um maior peso para o cálculo da estimativa global. Quanto aos pesos  $\alpha_k$ , estes são normalmente definidos em função do modelo que se assumir para descrever o comportamento dos tamanhos do efeito, sendo que a maioria das meta-análises assenta num modelo de efeitos fixos ou num modelo de efeitos aleatórios, ambos casos particulares do modelo linear geral.

---

<sup>10</sup>Quando não é possível obter diretamente a partir dos estudos o mesmo tipo de índice, existem fórmulas de conversão entre alguns índices mais usados (consultar, por exemplo, o Capítulo 7 de Borenstein *et al.* (2009)).

# Capítulo 3

## Modelos em Meta-Análise

*Essentially, all models are wrong, but some are useful.*

George Box (1987)

### 3.1 Modelo de Efeitos Fixos

Um modelo de efeitos fixos (MEF) em Meta-Análise pressupõe que os  $K$  estudos partilham um tamanho do efeito comum  $\theta$ , estando este a ser estimado no estudo  $k$  por  $\hat{\theta}_k$ . Este tipo de modelo é apropriado quando os estudos selecionados para a análise são idênticos quanto aos seus métodos e características das amostras envolvidos. Por isso, considera-se que os tamanhos do efeito  $\theta_1, \dots, \theta_K$  são constantes desconhecidas, estando no entanto implícita a hipótese de homogeneidade  $\theta_1 = \dots = \theta_K = \theta$ .

Assim, um MEF em Meta-Análise é o modelo linear

$$Y_k = \theta_k + \epsilon_k = \theta + \epsilon_k, \quad k = 1, \dots, K, \quad (3.1)$$

onde  $Y_k = \hat{\theta}_k$  representa o estimador para o tamanho do efeito  $\theta_k$  e  $\epsilon_k$  o erro de amostragem dentro do estudo  $k$ . No modelo (3.1) tem-se que  $E(\epsilon_k) = 0$  e  $\text{Var}(\epsilon_k) = \sigma_k^2$ ,  $k = 1, \dots, K$ , e  $\text{Cov}(\epsilon_k, \epsilon_j) = 0$ ,  $k \neq j$ ,

$k, j = 1, \dots, K$ .<sup>1</sup> Consequentemente,  $E(\hat{\theta}_k) = \theta_k = \theta$  e  $\text{Var}(\hat{\theta}_k) = \sigma_k^2$ ,  $k = 1, \dots, K$ .

O parâmetro comum  $\theta$  é estimado pela média ponderada

$$\hat{\theta} = \frac{\sum_{k=1}^K \omega_k \hat{\theta}_k}{\sum_{k=1}^K \omega_k}, \quad (3.2)$$

onde  $\omega_k$  é o peso a atribuir à estimativa  $\hat{\theta}_k$ . Também é usual supor-se para o modelo (3.1) que  $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$ ,<sup>2</sup> pelo que

$$\hat{\theta}_k \sim \mathcal{N}(\theta, \sigma_k^2), \quad k = 1, \dots, K.$$

Embora se assumam *a priori* para um MEF que as variâncias  $\sigma_k^2$ ,  $k = 1, \dots, K$ , são conhecidas, na prática raramente o são, pelo que nestes casos há que as estimar.<sup>3</sup> Neste cenário, é usual assumir-se que a distribuição de  $\hat{\theta}_k$  é aproximadamente normal com valor médio  $\theta$  e variância  $\hat{\sigma}_k^2$ .

Por outro lado, como o estimador (3.2) é não enviesado para o parâmetro  $\theta$ , independentemente dos valores considerados para os pesos (considerados constantes), a situação ideal surge quando os pesos são escolhidos de modo que  $\hat{\theta}$  seja o estimador não enviesado de variância uniformemente mínima (UMVUE).

Para a classe dos estimadores não enviesados de  $\theta$ , a desigualdade de Cramér-Rao estabelece como limite inferior para a variância destes estimadores  $\left\{E\left[\left(\frac{d}{d\theta} \ln f(\mathbf{Y}; \theta)\right)^2\right]\right\}^{-1}$ , onde  $\mathbf{Y} = (Y_1, \dots, Y_K)$  e  $f$  é a função densidade de probabilidade conjunta do vetor aleatório  $\mathbf{Y}$ . Dado que a função log-verossimilhança para o modelo (3.1) é

$$\ell(\theta; \mathbf{y}) = \ln f(\mathbf{y}; \theta) = -\frac{K}{2} \ln(2\pi) - \frac{1}{2} \sum_{k=1}^K \ln \sigma_k^2 - \frac{1}{2} \sum_{k=1}^K \frac{(y_k - \theta)^2}{\sigma_k^2},$$

<sup>1</sup>A notação  $\text{Cov}(X, Y)$  representa a covariância entre as variáveis aleatórias  $X$  e  $Y$ .

<sup>2</sup>A notação  $X \sim \mathcal{N}(a, b)$  representa uma variável aleatória  $X$  com distribuição normal, com valor médio  $a$  e variância  $b$ .

<sup>3</sup>Aconselha-se, sempre que possível, o uso de estimadores não enviesados para as variâncias.

tem-se que

$$E \left[ \left( \frac{d}{d\theta} \ell(\theta; \mathbf{Y}) \right)^2 \right] = \text{Var} \left( \frac{d}{d\theta} \ell(\theta; \mathbf{Y}) \right) = \text{Var} \left( \sum_{k=1}^K \frac{(Y_k - \theta)}{\sigma_k^2} \right) = \sum_{k=1}^K \frac{1}{\sigma_k^2}.$$

Conseqüentemente, se num MEF considerarmos  $\omega_k = 1/\sigma_k^2$ ,  $k = 1, \dots, K$ ,  $\hat{\theta}$  definido em (3.2) tem variância dada por

$$\text{Var}(\hat{\theta}) = \sigma_{\hat{\theta}}^2 = \left( \sum_{k=1}^K \frac{1}{\sigma_k^2} \right)^{-1} = \left( \sum_{k=1}^K \omega_k \right)^{-1}, \quad (3.3)$$

ou seja,  $\hat{\theta}$  é o estimador UMVUE de  $\theta$ . Portanto, os estudos que apresentarem menor variabilidade, sendo estes geralmente os estudos maiores, contribuirão com maiores pesos para o cálculo de  $\hat{\theta}$ . Note-se que o estimador  $\hat{\theta}$  definido em (3.2) também é o estimador de máxima verosimilhança de  $\theta$  quando os pesos considerados são aqueles.

Pelo exposto acima segue-se que

$$\hat{\theta} \sim \mathcal{N}(\theta, \sigma_{\hat{\theta}}^2), \quad (3.4)$$

onde  $\sigma_{\hat{\theta}}^2$  está definido em (3.3). O resultado (3.4) permite-nos obter intervalos de confiança (IC) para  $\theta$ , assim como testar hipóteses sobre este parâmetro. A hipótese nula mais testada sobre  $\theta$  é  $H_0 : \theta = 0$ , ou seja, sobre o tamanho do efeito comum ser nulo (em cada estudo).<sup>4</sup>

Quanto à hipótese de homogeneidade dos tamanhos do efeito, implícita num MEF, esta pode ser testada usando a estatística  $Q$  de Cochran (Cochran, 1954)

$$Q = \sum_{k=1}^K \omega_k (\hat{\theta}_k - \hat{\theta})^2, \quad (3.5)$$

com  $\hat{\theta}$  definido em (3.2) e  $\omega_k = 1/\sigma_k^2$ ,  $k = 1, \dots, K$ . Sob a validade da hipótese de homogeneidade,  $Q$  tem distribuição do qui-quadrado com

---

<sup>4</sup>Note-se que num MEF as inferências realizadas sobre  $\theta$  estão condicionadas ao conjunto dos  $K$  estudos, pelo que estender conclusões para fora deste conjunto não faz sentido.

$K - 1$  graus de liberdade, ou seja,  $Q \sim \chi_{K-1}^2$ .<sup>5</sup> Logo, o valor- $p$  ( $p$ -value) associado ao teste é dado por  $P(Q > Q_{\text{obs}})$ , onde  $Q_{\text{obs}}$  denota o valor observado para a estatística (3.5).

## 3.2 Modelo de Efeitos Aleatórios

A maioria das revisões sistemáticas se baseia em estudos que podem diferir, por exemplo, quanto ao tipo de delineamento experimental, características das amostras envolvidas, tipo de intervenção ou resultados. Nestes casos não faz sentido considerarmos um MEF, mas um modelo de efeitos aleatórios (MEA).

Um MEA em Meta-Análise admite que o tamanho do efeito pode variar entre os estudos, pelo que a hipótese de homogeneidade dos tamanhos do efeito pode não ser válida. Este modelo pode ser apropriado quando, por exemplo, os tamanhos do efeito observados tendem a ser maiores (ou menores) em estudos cujos participantes são mais jovens do que em estudos cujos participantes são mais velhos. Por isso, assume-se que os tamanhos do efeito  $\theta_1, \dots, \theta_K$  formam uma amostra aleatória de uma população  $\theta$  de tamanhos do efeito, população esta com valor médio  $\mu$  e variância  $\tau^2$ . Como neste contexto o tamanho do efeito  $\theta_k$  passa a ser estocástico, é decomposto da forma  $\theta_k = \mu + \xi_k$ , onde  $\xi_k$  representa o desvio de  $\theta_k$  em relação ao valor médio populacional  $\mu$ ,  $k = 1, \dots, K$ .

Consequentemente, um MEA em Meta-Análise é o modelo linear

$$Y_k = \mu + \xi_k + \epsilon_k, \quad k = 1, \dots, K, \quad (3.6)$$

com  $Y_k = \hat{\theta}_k$ , e para o qual se tem que  $E(\xi_k) = 0$ ,  $\text{Var}(\xi_k) = \tau^2$ ,  $E(\epsilon_k) = 0$  e  $\text{Var}(\epsilon_k) = \sigma_k^2$ ,  $k = 1, \dots, K$ , e  $\text{Cov}(\epsilon_k, \epsilon_j) = 0$ ,  $\text{Cov}(\xi_k, \xi_j) = 0$ ,  $k \neq j$ ,  $k, j = 1, \dots, K$ , e  $\text{Cov}(\epsilon_k, \xi_j) = 0$ ,  $k, j = 1, \dots, K$ . Sob estas condições, verifica-se que  $E(\hat{\theta}_k) = \mu$  e  $\text{Var}(\hat{\theta}_k) = \sigma_k^2 + \tau^2$ ,  $k = 1, \dots, K$ .

---

<sup>5</sup>Se os  $\sigma_k^2$ 's forem estimados, a prática comum consiste em assumir que  $Q$  tem distribuição aproximada do qui-quadrado com  $K - 1$  graus de liberdade, ou seja,  $Q \stackrel{\approx}{\sim} \chi_{K-1}^2$ .

Tal como num MEF, assume-se para o modelo (3.6) que as variâncias  $\tau^2$  e  $\sigma_k^2$ ,  $k = 1, \dots, K$ , são conhecidas, mesmo quando na prática são estimadas. Em relação aos dois tipos de erro aleatório, admite-se que  $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$  e  $\xi_k \sim \mathcal{N}(0, \tau^2)$ , pelo que

$$\widehat{\theta}_k \sim \mathcal{N}(\mu, \sigma_k^2 + \tau^2), \quad k = 1, \dots, K.$$

Repare-se que, ao contrário de um MEF onde há apenas uma fonte de variabilidade, nomeadamente a variabilidade dentro dos estudos, num MEA há duas fontes de variabilidade independentes, a variabilidade dentro dos estudos e a variabilidade entre os estudos, estando esta última quantificada no valor  $\tau^2$ .

No que respeita ao valor médio populacional  $\mu$ , este também é estimado por uma média ponderada do tipo (3.2), mais precisamente por

$$\widehat{\mu} = \frac{\sum_{k=1}^K \omega_k^* \widehat{\theta}_k}{\sum_{k=1}^K \omega_k^*}, \quad (3.7)$$

com pesos  $\omega_k^* = 1/(\sigma_k^2 + \tau^2)$ ,  $k = 1, \dots, K$ . Neste caso, verificamos que  $\widehat{\mu}$  também é o estimador de máxima verosimilhança e o estimador UMVUE de  $\mu$ .<sup>6</sup> Quanto à variância de  $\widehat{\mu}$ , esta é dada por

$$\text{Var}(\widehat{\mu}) = \sigma_{\widehat{\mu}}^2 = \left( \sum_{k=1}^K \frac{1}{\sigma_k^2 + \tau^2} \right)^{-1} = \left( \sum_{k=1}^K \omega_k^* \right)^{-1}.$$

Assim, as inferências sobre o parâmetro  $\mu$  assentam no facto de

$$\widehat{\mu} \sim \mathcal{N}(\mu, \sigma_{\widehat{\mu}}^2), \quad (3.8)$$

sendo que a hipótese nula mais testada é  $H_0 : \mu = 0$ , ou seja, que o tamanho médio dos efeitos é nulo (na população). Porém, o resultado (3.8) será apenas uma aproximação se se tiver que estimar os pesos  $\omega_k^*$ .<sup>7</sup>

---

<sup>6</sup>Demonstração análoga à do MEF.

<sup>7</sup>Note-se que agora as inferências realizadas sob um MEA não estão condicionadas ao conjunto dos estudos selecionados, pelo que as conclusões podem ser estendidas à população de tamanhos do efeito.

## Método de Mantel-Haenszel

O método que vimos para combinar, de forma ótima, os tamanhos do efeito num MEF (e num MEA) é conhecido pelo *método da variância inversa*. No entanto, para tamanhos do efeito baseados em proporções existem outros métodos de combinação como, por exemplo, o método de Mantel-Haenszel para *odds ratios*, riscos relativos e diferenças de proporções, e o método de Peto para *odds ratios* de Peto (Yusuf *et al.*, 1985).

Considerando para o estudo  $k$  que  $n_{1k} = a_k + b_k$ ,  $n_{2k} = c_k + d_k$  e  $n_k = n_{1k} + n_{2k}$ , os pesos para um *odds ratio* combinado de Mantel-Haenszel (1959), usando a fórmula (3.2), são  $\omega_k = b_k c_k / n_k$ ,  $k = 1, \dots, K$ . Um estimador para a variância do logaritmo do estimador do *odds ratio* combinado de Mantel-Haenszel ( $\widehat{\text{OR}}_{\text{MH}}$ ) é

$$\widehat{\text{Var}}(\ln(\widehat{\text{OR}}_{\text{MH}})) = \frac{\sum_{k=1}^K P_k R_k}{2 \left( \sum_{k=1}^K R_k \right)^2} + \frac{\sum_{k=1}^K (P_k S_k + Q_k R_k)}{2 \left( \sum_{k=1}^K R_k \right) \left( \sum_{k=1}^K S_k \right)} + \frac{\sum_{k=1}^K Q_k S_k}{2 \left( \sum_{k=1}^K S_k \right)^2},$$

onde  $P_k = (a_k + d_k) / n_k$ ,  $Q_k = (b_k + c_k) / n_k$ ,  $R_k = a_k d_k / n_k$  e  $S_k = b_k c_k / n_k$ .

Greenland e Robins (1985) estenderam o método de Mantel-Haenszel a riscos relativos e diferenças de proporções. Para o cálculo de um risco relativo combinado de Mantel-Haenszel ( $\widehat{\text{RR}}_{\text{MH}}$ ), os pesos são  $\omega_k = n_{1k} c_k / n_k$ ,  $k = 1, \dots, K$ . Um estimador para a variância do logaritmo do estimador é

$$\widehat{\text{Var}}(\ln(\widehat{\text{RR}}_{\text{MH}})) = \frac{\sum_{k=1}^K \frac{n_{1k} n_{2k} (a_k + c_k) - a_k c_k n_k}{n_k^2}}{\left( \sum_{k=1}^K \frac{a_k n_{2k}}{n_k} \right) \left( \sum_{k=1}^K \omega_k \right)}.$$

Para o cálculo de uma diferença de proporções combinada de Mantel-Haenszel ( $\widehat{\text{RD}}_{\text{MH}}$ ) devem-se considerar os pesos  $\omega_k = n_{1k} n_{2k} / n_k$ ,  $k = 1, \dots, K$ . Um estimador para a variância do logaritmo do estimador é

$$\widehat{\text{Var}}(\ln(\widehat{\text{RD}}_{\text{MH}})) = \frac{\sum_{k=1}^K \frac{(a_k b_k n_{2k})^3 + (c_k d_k n_{1k})^3}{(n_{1k} n_{2k} n_k)^2}}{\left( \sum_{k=1}^K \omega_k \right)^2}.$$

### 3.2.1 Estimação do Parâmetro de Heterogeneidade

Num MEA o parâmetro  $\tau^2$ , geralmente desconhecido, expressa o grau de variabilidade, ou heterogeneidade, na população de tamanhos do efeito. Se  $\tau^2 = 0$ , a variabilidade observada entre as diferentes estimativas dos tamanhos do efeito deve-se exclusivamente ao erro de amostragem. Esta é a razão pela qual um MEF é considerado um caso particular de um MEA quando  $\tau^2 = 0$ . Quando  $\tau^2 > 0$ , isto significa que há uma fonte de variabilidade adicional à que é esperada somente com base no erro de amostragem dentro dos estudos.

As possíveis causas para um  $\tau^2 > 0$  são:

- i) Variabilidade esperável numa população aleatória e que descreve o comportamento da população;
- ii) Presença de variáveis moderadoras, ou seja, existência de covariáveis que podem ser “externas” ao estudo (por exemplo, a região geográfica onde se realizou o estudo pode ter influenciado a magnitude dos tamanhos do efeito), ou “internas” ao estudo (por exemplo, doses mais elevadas de um fármaco pode aumentar, ou diminuir, os tamanhos do efeito);
- iii) Uma combinação entre uma população aleatória de tamanhos do efeito e a presença de variáveis moderadoras.

A primeira causa não carece de explicação, mas a segunda pode levar-nos à consideração de um MEF com variáveis moderadoras, ou seja, à realização de uma meta-regressão ou uma análise de subgrupos, métodos que abordaremos no Capítulo 5. Por sua vez, a última causa pode levar-nos à consideração de um modelo com efeitos mistos. Por isso, uma prática cada vez mais frequente numa meta-análise seja a realização de um teste de homogeneidade,<sup>8</sup> para detetar a existência dessa fonte de variabilidade extra, e a indicação de uma estimativa para  $\tau^2$ .

---

<sup>8</sup>Alguns autores preferem chamar teste de heterogeneidade.

Vários estimadores têm sido propostos para a estimação de  $\tau^2$  e baseados nos mais diversos métodos. São conhecidos alguns estimadores não enviesados de  $\tau^2$ , sendo que para esta classe de estimadores a desigualdade de Cramér-Rao estabelece como limite inferior para as suas variâncias

$$\left\{ \mathbb{E} \left[ \left( \frac{\partial}{\partial \tau^2} \ell(\mu, \tau^2; \mathbf{Y}) \right)^2 \right] \right\}^{-1} = 2 \left[ \sum_{k=1}^K \frac{1}{(\sigma_k^2 + \tau^2)^2} \right]^{-1},$$

em que  $\ell$  é a função log-verossimilhança para o modelo (3.6).

Alguns estimadores possíveis são:

- **Estimador de Hedges (1983)**

$$\hat{\tau}_{\text{HE}}^2 = \frac{1}{K-1} \sum_{k=1}^K (\hat{\theta}_k - \bar{\theta})^2 - \frac{1}{K} \sum_{k=1}^K \sigma_k^2,$$

onde  $\bar{\theta} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_k$ .

O estimador de Hedges baseia-se num estimador grosseiro de  $\tau^2$ , nomeadamente em  $S_{\theta}^2 = \frac{1}{K-1} \sum_{k=1}^K (\hat{\theta}_k - \bar{\theta})^2$ . Porém, como  $S_{\theta}^2$  também engloba a variabilidade dentro dos estudos, esta é subtraída a  $S_{\theta}^2$ , sendo dada pela média  $\frac{1}{K} \sum_{k=1}^K \sigma_k^2$ .

Note-se que o estimador de Hedges é não enviesado, mesmo quando se recorre a estimadores não enviesados para a estimação de  $\sigma_k^2$ ,  $k = 1, \dots, K$ ;<sup>9</sup>

- **Estimador de DerSimonian-Laird (1986)**

$$\hat{\tau}_{\text{DL}}^2 = \frac{Q - (K - 1)}{c}, \quad (3.9)$$

em que  $Q$  é a estatística de Cochran definida em (3.5) e

$$c = \sum_{k=1}^K \omega_k - \frac{\sum_{k=1}^K \omega_k^2}{\sum_{k=1}^K \omega_k}, \quad (3.10)$$

---

<sup>9</sup>Como se podem obter estimativas negativas com  $\hat{\tau}_{\text{HE}}^2$ , a prática consiste em truncar estas em zero.

com  $\omega_k = 1/\sigma_k^2$ ,  $k = 1, \dots, K$ .

O estimador de DerSimonian-Laird é um estimador dos momentos, pois sabe-se que sob um MEA,

$$E(Q) = K - 1 + \tau^2 \left( \sum_{k=1}^K \omega_k - \frac{\sum_{k=1}^K \omega_k^2}{\sum_{k=1}^K \omega_k} \right)$$

(Biggerstaff e Tweedie, 1997). É também um estimador não enviesado de  $\tau^2$ , desde que se conheçam os valores  $\sigma_k^2$ ,  $k = 1, \dots, K$ . Se estas variâncias forem estimadas,  $\hat{\tau}_{DL}^2$  tende a sobrestimar  $\tau^2$ , especialmente quando  $K$  é pequeno;<sup>10</sup>

- **Estimador de Hunter-Schmidt (1990)**

$$\hat{\tau}_{HS}^2 = \frac{\sum_{k=1}^K \nu_k (\hat{\theta}_k - \hat{\theta})^2}{\sum_{k=1}^K \nu_k} - \frac{\sum_{k=1}^K \nu_k \sigma_k^2}{\sum_{k=1}^K \nu_k},$$

onde os  $\nu_k$ ,  $k = 1, \dots, K$ , são pesos e  $\hat{\theta}$  é o efeito síntese para um MEF.

O estimador de Hunter-Schmidt é uma espécie de generalização do tipo ponderado do estimador de Hedges. Quanto à escolha dos pesos  $\nu_k$ , o usual é considerar  $\nu_k = \omega_k = 1/\sigma_k^2$ ,  $k = 1, \dots, K$ , sendo que neste caso o estimador assume a expressão

$$\hat{\tau}_{HS}^2 = \frac{Q - K}{\sum_{k=1}^K \omega_k},$$

onde  $Q$  é a estatística de Cochran;<sup>11</sup>

- **Estimador de Sidik-Jonkman (2005)**

$$\hat{\tau}_{SJ}^2 = \frac{1}{K - 1} \sum_{k=1}^K \eta_k^{-1} (\hat{\theta}_k - \tilde{\mu})^2,$$

---

<sup>10</sup>Tal como o estimador de Hedges, o estimador de DerSimonian-Laird pode produzir estimativas negativas quando  $Q < K - 1$ , pelo que estas devem ser truncadas em zero.

<sup>11</sup>Este estimador também pode produzir estimativas negativas, de modo que estas devem ser truncadas em zero.

com

$$\tilde{\mu} = \frac{\sum_{k=1}^K \eta_k^{-1} \hat{\theta}_k}{\sum_{k=1}^K \eta_k^{-1}},$$

$\eta_k = 1 + \sigma_k^2/\hat{\tau}_0^2$ ,  $k = 1, \dots, K$ , e  $\hat{\tau}_0^2$  uma estimativa inicial para  $\tau^2$ .

Este estimador tem uma desvantagem em relação aos estimadores anteriores que é nunca produzir estimativas nulas para  $\tau^2$ ;

- **Estimador de Máxima Verosimilhança**

$$\hat{\tau}_{\text{ML}}^2 = \frac{\sum_{k=1}^K \hat{\omega}_k^{*2} [(\hat{\theta}_k - \hat{\mu}_{\text{ML}})^2 - \sigma_k^2]}{\sum_{k=1}^K \hat{\omega}_k^{*2}},$$

em que  $\hat{\mu}_{\text{ML}} = \sum_{k=1}^K \hat{\omega}_k^* \hat{\theta}_k / \sum_{k=1}^K \hat{\omega}_k^*$  e  $\hat{\omega}_k^* = 1/(\sigma_k^2 + \hat{\tau}_{\text{ML}}^2)$ ,  $k = 1, \dots, K$ .

Para se obter  $\hat{\mu}_{\text{ML}}$  é necessário recorrer-se a métodos numéricos, começando-se com uma estimativa inicial para  $\tau^2$ , fornecida por exemplo por um dos métodos de estimação não iterativos, e usada no cálculo de  $\hat{\mu}_{\text{ML}}$ . Em cada iteração há que averiguar se a estimativa de  $\tau^2$  é negativa, porque se for, deve ser truncada em zero;

- **Estimador de Máxima Verosimilhança Restrita**

$$\hat{\tau}_{\text{REML}}^2 = \frac{\sum_{k=1}^K \hat{\omega}_k^{*2} [(\hat{\theta}_k - \hat{\mu}_{\text{ML}})^2 - \sigma_k^2]}{\sum_{k=1}^K \hat{\omega}_k^{*2}} + \frac{1}{\sum_{k=1}^K \hat{\omega}_k^*},$$

em que  $\hat{\mu}_{\text{ML}} = \sum_{k=1}^K \hat{\omega}_k^* \hat{\theta}_k / \sum_{k=1}^K \hat{\omega}_k^*$  e  $\hat{\omega}_k^* = 1/(\sigma_k^2 + \hat{\tau}_{\text{REML}}^2)$ ,  $k = 1, \dots, K$ .

O estimador de máxima verosimilhança restrita procura reduzir, ou eliminar, o viés negativo normalmente associado ao estimador de máxima verosimilhança;<sup>12</sup>

---

<sup>12</sup>Estimativas negativas também devem ser truncadas em zero em cada iteração.

Dos estimadores acima indicados, os mais usados são o estimador de DerSimonian-Laird e o estimador da máxima verosimilhança restrita, sendo este último geralmente preferido pelos estatísticos que realizam meta-análises.

A popularidade do estimador de DerSimonian-Laird, principalmente em meta-análises ligadas à área da Saúde, resulta da sua interpretação simples.<sup>13</sup> Num MEA as diferenças observadas entre as diferentes estimativas  $\hat{\theta}_k$  devem-se a dois fatores, um à verdadeira heterogeneidade entre os tamanhos do efeito e outro ao erro de amostragem dentro dos estudos, ambos incorporados na medida de variabilidade total  $Q$ . Sob a hipótese de homogeneidade, tem-se que  $E(Q) = K - 1$ , pelo que a diferença  $Q - (K - 1)$  mede o excesso da variabilidade entre os estudos, ou seja, o que se espera que venha a ser excedido com base no erro de amostragem. O factor  $c$  na expressão (3.9) permite recolocar o índice na mesma escala que foi usada para reportar a variância entre os estudos.<sup>14</sup>

*Observação 3.1.* Se num MEA as variâncias  $\tau^2$  e  $\sigma_k^2$ ,  $k = 1, \dots, K$ , são estimadas, a variância de  $\hat{\mu}$  passa a ser estimada por  $\widehat{\text{Var}}(\hat{\mu}) = \left(\sum_{k=1}^K \hat{\omega}_k^*\right)^{-1}$ , em que  $\hat{\omega}_k^* = 1/(\hat{\sigma}_k^2 + \hat{\tau}^2)$ ,  $k = 1, \dots, K$ . Nestes casos, e para efeitos de inferência, assume-se que a distribuição de  $\hat{\mu}$  é aproximadamente normal. Porém, se  $\tau^2$  for estimado pelo estimador de DerSimonian-Laird, Hartung e Knapp (2001, 2003) propõem que se use em alternativa para a variância de  $\hat{\mu}$ ,

$$\widehat{\text{Var}}_{\text{HK}}(\hat{\mu}) = \frac{1}{K-1} \sum_{k=1}^K \frac{\hat{\omega}_k^*}{\hat{\omega}^*} (\hat{\theta}_k - \hat{\mu})^2,$$

onde  $\hat{\omega}^* = \sum_{k=1}^K \hat{\omega}_k^*$ .

Por outro lado, como a estatística  $T = (\hat{\mu} - \mu) / \sqrt{\widehat{\text{Var}}_{\text{HK}}(\hat{\mu})}$  tem distribuição (exata) de  $t$  de Student com  $K - 1$  graus de liberdade, esta

<sup>13</sup>O *software* estatístico RevMan5 fornecido pela *Cochrane Collaboration* considera o estimador de DerSimonian-Laird para  $\tau^2$ .

<sup>14</sup>Note-se que a estatística  $Q$  é adimensional.

pode ser usada para fazer inferências sobre o valor médio populacional  $\mu$ . Esta alteração na variância de  $\hat{\mu}$  é conhecida pelo ajustamento de Hartung-Knapp.

Vietchbauer (2005) comparou os estimadores de Hedges (HE), Hunter-Schmidt (HS), DerSimonian-Laird (DL), máxima verosimilhança (ML) e máxima verosimilhança restrita (REML) quanto ao viés e eficiência. Verificou que os resultados obtidos com os cinco estimadores podem diferir bastante, sendo no entanto todos eles consistentes para  $\tau^2$ . Concluiu também que:

- i) Os estimadores HS e ML possuem, de um modo geral, um erro quadrático médio inferior aos outros três estimadores;
- ii) Nenhum estimador consegue suplantiar totalmente os outros quanto ao erro quadrático médio, mesmo quando comparados com o estimador HE que, de todos os estimadores, é aquele que possui, geralmente, um erro quadrático médio mais elevado;
- iii) Se o critério de escolha for apenas o viés e a eficiência, deve-se evitar os estimadores HS e ML (possuem um viés negativo);
- iv) Os estimadores DL e HS apresentam alguns problemas quando as variâncias  $\sigma_k^2$ ,  $k = 1, \dots, K$ , são pequenas;
- v) O estimador REML apresenta um bom equilíbrio entre não enviesamento e eficiência, pelo que deve ser de um modo geral preferido.

Por outro lado, quando  $\sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$ , observam-se as seguintes igualdades:

i)

$$\hat{\tau}_{\text{HS}}^2 = \hat{\tau}_{\text{ML}}^2 = \frac{1}{K} \sum_{k=1}^K (\hat{\theta}_k - \bar{\theta})^2 - \sigma^2,$$

$$\text{em } \bar{\theta} = \frac{1}{K} \sum_{k=1}^K \hat{\theta}_k;$$

ii)

$$\widehat{\tau}_{\text{HE}}^2 = \widehat{\tau}_{\text{DL}}^2 = \widehat{\tau}_{\text{REML}}^2 = \frac{1}{K-1} \sum_{k=1}^K (\widehat{\theta}_k - \bar{\theta})^2 - \sigma^2.$$

Neste cenário mais restrito tem-se que

$$\text{Var}(\widehat{\tau}_{\text{HS}}^2) = \text{Var}(\widehat{\tau}_{\text{ML}}^2) = \frac{2(K-1)}{K^2} (\sigma^2 + \tau^2)^2,$$

e

$$\text{Var}(\widehat{\tau}_{\text{HE}}^2) = \text{Var}(\widehat{\tau}_{\text{DL}}^2) = \text{Var}(\widehat{\tau}_{\text{REML}}^2) = \frac{2}{K-1} (\sigma^2 + \tau^2)^2,$$

verificando-se que os estimadores HE, DL e REML são estimadores UMVUE para  $\tau^2$ .

Por outro lado, se assumirmos para a população de tamanhos do efeito que  $\theta \sim \mathcal{N}(\mu, \tau^2)$ , existem vários métodos que se podem considerar para a construção de intervalos de confiança para  $\tau^2$ .<sup>15</sup> Em alternativa, Vietchbauer (2007) propõe um intervalo de confiança baseado na estatística  $Q$  de Cochran generalizada definida por

$$Q(\tau^2) = \sum_{k=1}^K \frac{(\widehat{\theta}_k - \widehat{\mu})^2}{\sigma_k^2 + \tau^2} = \sum_{k=1}^K \omega_k^* (\widehat{\theta}_k - \widehat{\mu})^2,$$

com  $\widehat{\mu}$  definido em (3.7). Sob os pressupostos de um MEA,  $Q(\tau^2) \sim \chi_{K-1}^2$ . Assim, denotando por  $\chi_{K-1; \gamma}^2$  o quantil de probabilidade  $\gamma$  da distribuição do qui-quadrado com  $K-1$  graus de liberdade, tem-se que

$$P\left(\chi_{K-1; \alpha/2}^2 \leq Q(\tau^2) \leq \chi_{K-1; 1-\alpha/2}^2\right) = 1 - \alpha.$$

Consequentemente, um intervalo de confiança a  $(1 - \alpha)100\%$  para  $\tau^2$ ,  $[\widetilde{\tau}_L^2, \widetilde{\tau}_U^2]$ , pode ser obtido invertendo a estatística  $Q$  generalizada de forma a que  $Q(\widetilde{\tau}_L) = \chi_{K-1; 1-\alpha/2}^2$  e  $Q(\widetilde{\tau}_U) = \chi_{K-1; \alpha/2}^2$ , sendo os limites de confiança obtidos iterativamente.<sup>16</sup>

<sup>15</sup>Por exemplo, intervalos de confiança baseados no estimador  $\widehat{\tau}_{\text{DL}}^2$  não são fiáveis dado que a distribuição amostral de  $\widehat{\tau}_{\text{DL}}^2$  não é normal.

<sup>16</sup>Este é o intervalo implementado para  $\tau^2$  no pacote *metafor* do *software* R (Vietchbauer, 2010).

Simulações efetuadas por Vietchbauer (2007) revelaram que a probabilidade de cobertura nominal é atingida com este intervalo de confiança, mesmo quando as amostras têm dimensão pequena, e desde que as variâncias  $\sigma_k^2$  sejam conhecidas e os pressupostos de um MEA não sejam violados.

### 3.2.2 Impacto da Heterogeneidade numa Meta-Análise

O teste  $Q$  de Cochran tem sido o mais usado em meta-análises para testar a homogeneidade dos tamanhos do efeito, dado o seu carácter genérico de aplicação. No entanto, numa meta-análise é insuficiente constatarmos apenas que há evidência a favor da heterogeneidade, há todo o interesse em avaliarmos qual o impacto dessa heterogeneidade nas conclusões.

Para uma avaliação do impacto poderíamos ter em conta a estatística  $Q$  de Cochran, uma vez que esta pode ser encarada como uma medida de variabilidade total. Porém,  $Q$  apresenta alguns inconvenientes como, por exemplo, depender do número de estudos e ser uma medida estandardizada, tornando-a por isso uma medida pouco intuitiva.<sup>17</sup>

Higgins e Thompson (2002) propõem medidas para quantificar o impacto da heterogeneidade numa meta-análise que não dependem do número de estudos. Uma das medidas é

$$I^2 = \begin{cases} \frac{Q-(K-1)}{Q} \times 100\% & , \text{ se } Q > K - 1 \\ 0 & , \text{ se } Q \leq K - 1 \end{cases}$$

que procura estimar a quantidade  $\frac{\tau^2}{\sigma^2 + \tau^2}$ , onde  $\sigma^2$  representa a variabilidade dentro dos estudos.<sup>18</sup> A medida  $I^2$  tem uma interpretação

<sup>17</sup>Estudos de simulação têm revelado que o teste  $Q$  possui uma fraca potência estatística quando  $K$  é pequeno e uma elevada potência estatística na deteção de heterogeneidade quando esta é na realidade irrelevante.

<sup>18</sup>Os autores assumem o cenário  $\sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$ , embora esta situação possa ser generalizada para o caso de haver variâncias diferentes.

bastante simples, dá-nos a percentagem da variabilidade total que é explicada pela heterogeneidade da população de tamanhos do efeito. Por exemplo, se  $I^2 = 0$ , toda a variabilidade observada se deve ao erro de amostragem dentro dos estudos.

Outra medida proposta por Higgins e Thompson (2002) é

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

A medida  $H^2$  dá-nos o excesso relativo de  $Q$  em relação ao que é esperado sob a hipótese de homogeneidade dos tamanhos do efeito ( $\tau^2 = 0$ ). Observe-se que na ausência de heterogeneidade  $H^2 = 1$ .<sup>19</sup>

Por outro lado, um intervalo de confiança a  $(1 - \alpha)100\%$  para  $H$  baseado na estatística  $Q$  é

$$\left[ \exp(\ln H - z_{1-\alpha/2} \sqrt{\text{Var}(\ln H)}), \exp(\ln H + z_{1-\alpha/2} \sqrt{\text{Var}(\ln H)}) \right],$$

onde  $z_\gamma$  representa o quantil de probabilidade  $\gamma$  da distribuição normal padrão e

$$\sqrt{\text{Var}(\ln H)} = \begin{cases} \frac{1}{2} \left( \frac{\ln Q - \ln(K-1)}{\sqrt{2K} - \sqrt{2K-3}} \right) & , Q > K \\ \sqrt{\frac{1}{2(K-2)} \left( 1 - \frac{1}{3(K-2)^2} \right)} & , Q \leq K \end{cases}.$$

Quanto a intervalos de confiança para  $I^2$ , estes podem ser obtidos diretamente a partir dos intervalos de confiança para  $H^2$  ( $H$ ), tendo em conta a relação  $I^2 = \frac{H^2 - 1}{H^2}$ .

### Como escolher entre um MEF e um MEA?

As opiniões não são unânimes sobre este assunto. Por exemplo, quando na dúvida entre os dois modelos, há autores que defendem que

---

<sup>19</sup>Higgins e Thompson (2002) também propõem uma terceira medida,  $R^2 = \frac{\sigma^2 + \tau^2}{\sigma^2}$ . Esta medida descreve a inflação observada no intervalo de confiança para o efeito síntese sob um MEA, quando comparado com o intervalo de confiança obtido sob um MEF. Por exemplo,  $R^2 = 1$  significa que as inferências sob os dois modelos são idênticas, pelo que bastará assumir um MEF.

se deve optar primeiro por um MEF e só passar a considerar um MEA, se o resultado do teste de homogeneidade for estatisticamente significativo. Neste contexto, um efeito síntese só faz sentido ser calculado se não se rejeitar a hipótese de homogeneidade. Por outro lado, mesmo quando se julgue adequado um MEF e se observe uma variabilidade considerável entre as estimativas, poder-se-á procurar explicar essa heterogeneidade recorrendo, por exemplo, a uma meta-regressão.

No entanto, a abordagem anterior é criticada por outros autores que argumentam que a ausência de significância estatística para o teste de homogeneidade é, frequentemente, consequência da fraca potência estatística do teste por estarem envolvidos poucos estudos na análise.<sup>20</sup> Por isso, estes autores afirmam ser mais prudente assumir por defeito um MEA, sendo um MEF um caso particular para  $\tau^2 = 0$ . No que respeita a obtenção de um efeito síntese neste cenário, esta faz sentido, porque o que se estará a estimar é o valor médio da população de tamanhos do efeito.

## Exemplo 1

*A suplementação de antioxidantes traz benefícios para a saúde?*

Sabe-se que o *stress* oxidativo é a causa de várias doenças, razão pela qual muitas pessoas decidem recorrer à suplementação de antioxidantes na esperança de conseguir prevenir algumas patologias e adiar o processo de envelhecimento.<sup>21</sup> Mas, ao contrário do que é largamente publicitado pelos média, haverá mesmo evidência a favor dos benefícios da suplementação de antioxidantes?

Para avaliar o risco de mortalidade (por qualquer causa) associado à suplementação de antioxidantes, Bjelakovic *et al.* (2007) realizaram

---

<sup>20</sup>Por exemplo, segundo a *Cochrane Collaboration* o número mediano de estudos usados numa meta-análise é 6.

<sup>21</sup>O *stress* oxidativo é consequência do desequilíbrio entre a produção de radicais livres e a capacidade do corpo eliminar, ou neutralizar, a toxicidade dos efeitos nocivos destes radicais através da ação de antioxidantes.

uma meta-análise com base em 68 estudos que envolveram um grupo de tratamento, cujos elementos receberam um único suplemento ou combinações de vários suplementos de antioxidantes (betacaroteno, vitaminas A, C e E ou selênio), e um grupo de controlo, cujos elementos receberam ou não um placebo.

Para além dos autores reportarem na sua análise o número de mortes e o número total de participantes por grupo e por estudo, também identificaram várias outras variáveis que poderiam ter algum interesse, como o tipo de delineamento usado no estudo, quais os suplementos administrados e as suas dosagens, a idade média dos participantes, a duração do tratamento e do *follow-up*, e o tipo de prevenção considerado com a suplementação (primária ou secundária).<sup>22</sup>

Para exemplificarmos os procedimentos anteriores, usaremos apenas os 47 estudos que segundo os autores têm um baixo risco de viés, porque, na opinião destes, cumprem simultaneamente os seguintes requisitos: a geração adequada de uma sequência aleatória para a alocação dos participantes por grupo e a ocultação do tratamento administrado.

Na Tabela 3.1 regista-se a informação sobre a identificação do estudo (autores), o ano de publicação do estudo, o número de mortes, o total de participantes por grupo, a duração do tratamento (em anos) e o tipo de prevenção (P – primária; S – secundária).

O tamanho do efeito considerado em Bjelakovic *et al.* (2007) foi o risco relativo, definido por

$$\theta = \text{RR} = \frac{p_1}{p_2},$$

onde  $p_1$  e  $p_2$  representam, respetivamente, o risco de mortalidade nas populações subjacentes ao grupo de tratamento (antioxidantes) e grupo de controlo.

---

<sup>22</sup>Os autores excluíram estudos sobre prevenção terciária, como estudos sobre doenças agudas, infecciosas e malignas, à exceção de cancro da pele não melanoma.

Tabela 3.1: Informação sobre os 47 estudos com um baixo risco de viés em Bjelakovic *et al.* (2007)

Estudo	Ano	Tratamento Nº mortes/Total	Controlo Nº mortes/Total	Duração do tratamento	Tipo de prevenção
Greenberg et al.	1990	79/913	72/892	5	S
Murphy et al.	1992	4/53	2/56	0.003	S
Li et al.	1993	157/1657	167/1661	6	S
Greenberg et al.	1994	30/650	14/214	4	S
Pike et al.	1995	1/24	0/23	1	P
Clark et al.	1996	108/653	129/659	7.4	S
Hennekens et al.	1996	979/11036	968/11035	12.9	P
Richer	1996	2/39	2/39	1.5	S
Stephens et al.	1996	68/1035	52/967	1.4	S
Moon et al.	1997	62/1157	53/1140	3.8	S
Girodon et al.	1999	155/543	51/182	2	S
Green et al.	1999	15/801	22/820	4.5	S
Boaz et al.	2000	31/97	29/99	1.42	S
Correa et al.	2000	16/739	2/237	6	S
Jacobson et al.	2000	0/57	1/55	0.5	P
AREDS	2001	251/2370	240/2387	6.3	S
Brown et al.	2001	17/84	13/76	3	S
Desnuelle et al.	2001	34/144	35/144	1	S
Chylack et al.	2002	9/149	3/148	3	S
Graat et al.	2002	3/499	5/153	1	P
Heart Protection Study	2002	1446/10269	1389/10267	5	S
Hodis et al.	2002	2/177	1/176	3	P
Waters et al.	2002	16/212	6/211	3	S
White et al.	2002	1/50	1/50	0.23	S
Wluka et al.	2002	1/67	0/69	2	S
Collins et al.	2003	1/26	1/26	0.5	S
Prince et al.	2003	1/29	0/32	0.25	S
Salonen et al.	2003	19/390	3/130	6	P
Virtamo et al.	2003	8226/21846	2605/7287	6.1	P
Allsup et al.	2004	4/81	4/83	0.15	S
Goodman et al.	2004	1855/9420	1509/8894	4	P
Hercberg et al.	2004	76/6481	98/6536	7.54	P
Manuel-y-Keenoy et al.	2004	1/12	0/12	0.5	S
McNeil et al.	2004	20/595	11/598	4	S
Meydani et al.	2004	39/311	44/306	1	P
Mezey et al.	2004	4/25	5/26	0.25	S
Richer et al.	2004	0/30	2/31	1	S
Avenell et al.	2005	8/456	4/454	1	P
Graf et al.	2005	31/83	28/77	1.5	S
Lee et al.	2005	636/19937	615/19939	10.1	P
Limburg et al.	2005	1/180	0/180	0.83	S
Lonn et al.	2005	799/4761	801/4780	4.5	S
Marras et al.	2005	154 /399	142 /401	2.6	S

Tabela 3.1: Informação sobre os 47 estudos com um baixo risco de viés em Bjelakovic *et al.* (2007)

Estudo	Ano	Tratamento	Controlo	Duração do tratamento	Tipo de prevenção
		Nº mortes/Total	Nº mortes/Total		
Mooney et al.	2005	1/142	0/142	1.25	P
Tam et al.	2005	1/20	1/19	0.23	S
Witte et al.	2005	1/16	1/16	0.75	S
Rayman et al.	2006	1/380	0/121	0.5	P

Para obtermos o risco relativo combinado temos que, em primeiro lugar, calcular para cada estudo as estimativas do risco relativo e da variância do estimador deste dentro do estudo. Como os estudos considerados em Bjelakovic *et al.* (2007) não usaram, por exemplo, o mesmo tipo delineamento experimental, nem as mesmas combinações de antioxidantes e dosagens, faz todo o sentido assumirmos à partida um MEA para descrevermos o comportamento dos tamanhos do efeito, o que aliás foi feito pelos autores. Contudo, aplicaremos os dois modelos para podermos comparar as conclusões obtidas sob ambos.

*Observação 3.2.* A combinação de estimativas de riscos relativos (ou *odds ratios*) pelo método da variância inversa é feita na escala logarítmica, ou seja, usando os logaritmos das estimativas, dado que a distribuição amostral assintótica do logaritmo do estimador do risco relativo (ou do *odds ratio*) é normal.

Para nos auxiliar nos cálculos usaremos o *software* R ([www.r-project.org](http://www.r-project.org)) que contém alguns pacotes para a Meta-Análise, entre os quais o pacote *meta* de Guido Schwarzer, o *rmeta* de Thomas Lumley, o *metafor* de Wolfgang Viechtbauer, o *mymeta* de Antonio Gasparrini e o *mada* de Philipp Doebler. Por exemplo, o pacote *meta* é de fácil utilização, realizando somente meta-análises univariadas. Este pacote disponibiliza diferentes funções consoante o tipo de tamanho do efeito considerado, como as funções *metabin* para dados binários, *metacont* para médias, *metacor* para correlações, *metaprop* para uma proporção e *metagen* para tamanhos do efeito mais genéricos.

Para a meta-análise da suplementação de antioxidantes recorreremos ao pacote `meta`.<sup>23</sup>

```
> library(meta)
> meta.antiox=metabin(event.e=ai,n.e=nli,event.c=ci,n.c=n2i,
+                      data=bjelakovic,sm="RR",method="Inverse",
+                      method.tau="REML",
+                      studlab=paste(Estudo,Ano,sep=", "))
> meta.antiox
```

	RR	95%-CI	%W(fixed)	%W(random)
Greenberg et al.,1990	1.072	[0.790; 1.455]	0.6	1.4
Murphy et al.,1992	2.113	[0.404; 11.062]	0.0	0.1
Li et al.,1993	0.942	[0.766; 1.159]	1.3	2.9
Greenberg et al. ,1994	0.705	[0.381; 1.305]	0.1	0.4
Pike et al.,1995	2.878	[0.123; 67.163]	0.0	0.0
Clark et al.,1996	0.845	[0.670; 1.065]	1.0	2.4
Hennekens et al.,1996	1.011	[0.929; 1.101]	7.6	10.6
Richer,1996	1.000	[0.148; 6.747]	0.0	0.0
Stephens et al.,1996	1.222	[0.861; 1.734]	0.4	1.1
Moon et al.,1997	1.153	[0.806; 1.648]	0.4	1.1
Girodon et al.,1999	1.019	[0.779; 1.332]	0.8	1.8
Green et al.,1999	0.698	[0.365; 1.336]	0.1	0.3
Boaz et al.,2000	1.091	[0.716; 1.664]	0.3	0.8
Correa et al.,2000	2.566	[0.594; 11.077]	0.0	0.1
Jacobson et al.,2000	0.322	[0.013; 7.731]	0.0	0.0
AREDS,2001	1.053	[0.891; 1.245]	1.9	4.1
Brown et al.,2001	1.183	[0.616; 2.271]	0.1	0.3
Desnuelle et al.,2001	0.971	[0.644; 1.466]	0.3	0.8
Chylack et al.,2002	2.980	[0.823; 10.789]	0.0	0.1
Graat et al.,2002	0.184	[0.044; 0.761]	0.0	0.1
Heart Protection Study,2002	1.041	[0.972; 1.114]	11.7	13.0
Hodis et al.,2002	1.989	[0.182; 21.733]	0.0	0.0
Waters et al.,2002	2.654	[1.059; 6.651]	0.1	0.2
White et al.,2002	1.000	[0.064; 15.548]	0.0	0.0
Wluka et al.,2002	3.089	[0.128; 74.503]	0.0	0.0
Collins et al.,2003	1.000	[0.066; 15.150]	0.0	0.0

<sup>23</sup>O método de combinação por defeito no pacote `meta` é o método da variância inversa. No entanto, quando os dados são binários a função `metabin` aplica por defeito o método de combinação de Mantel-Haenszel. Por outro lado, o estimador por defeito de  $\tau^2$  é o de DerSimonian-Laird para um MEA.

Prince et al.,2003	3.305 [0.140; 78.025]	0.0	0.0
Salonen et al.,2003	2.111 [0.635; 7.019]	0.0	0.1
Virtamo et al.,2003	1.053 [1.017; 1.091]	44.0	19.0
Allsup et al.,2004	1.025 [0.265; 3.959]	0.0	0.1
Goodman et al.,2004	1.161 [1.091; 1.234]	14.4	14.2
Hercberg et al.,2004	0.782 [0.581; 1.053]	0.6	1.5
Manuel-y-Keenoy et al.,2004	3.000 [0.135; 66.801]	0.0	0.0
McNeil et al.,2004	1.827 [0.883; 3.780]	0.1	0.3
Meydani et al.,2004	0.872 [0.584; 1.303]	0.3	0.8
Mezey et al.,2004	0.832 [0.252; 2.748]	0.0	0.1
Richer et al.,2004	0.207 [0.010; 4.129]	0.0	0.0
Avenell et al.,2005	1.991 [0.604; 6.566]	0.0	0.1
Graf et al.,2005	1.027 [0.684; 1.542]	0.3	0.8
Lee et al.,2005	1.034 [0.927; 1.153]	4.6	7.8
Limburg et al.,2005	3.000 [0.123; 73.154]	0.0	0.0
Lonn et al.,2005	1.001 [0.916; 1.095]	6.8	10.0
Marras et al.,2005	1.090 [0.909; 1.306]	1.7	3.6
Mooney et al.,2005	3.000 [0.123; 73.024]	0.0	0.0
Tam et al.,2005	0.950 [0.064; 14.132]	0.0	0.0
Witte et al.,2005	1.000 [0.068; 14.640]	0.0	0.0
Rayman et al.,2006	0.958 [0.039; 23.362]	0.0	0.0

Number of studies combined: k = 47

	RR	95%-CI	z	p-value
Fixed effect model	1.055	[1.030; 1.080]	4.48	< 0.0001
Random effects model	1.046	[1.007; 1.086]	2.33	0.0199

Quantifying heterogeneity:

tau<sup>2</sup> = 0.0016; H = 1.04 [1.00; 1.23]; I<sup>2</sup> = 6.9% [0.0%; 34.3%];  
Rb = 9.3% [0.0%; 34.5%]

Test of heterogeneity:

Q	d.f.	p-value
49.41	46	0.3387

Details on meta-analytical method:

- Inverse variance method
- Restricted maximum-likelihood estimator for tau<sup>2</sup>
- Continuity correction of 0.5 in studies with zero cell frequencies

Considerando que é adequado um MEF, obtemos para estimativa do tamanho do efeito comum  $\hat{\theta} = 1.055$ , ou seja, estima-se que há um aumento de 5.5% no risco de mortalidade com a suplementação de antioxidantes, quando comparado com a não suplementação. Por outro lado, como o IC a 95% para  $\theta$  é  $[1.030, 1.080]$ , estima-se que o risco de mortalidade aumente entre 3.0 a 8.0%. Quanto à hipótese do tamanho do efeito comum ser nulo, isto é,  $H_0 : \ln \theta = 0$  ( $\theta = 1$ ), esta é rejeitada quando se considera *a priori* níveis de significância maiores que 0.0001 ( $p_{\text{val}} < 0.0001$ ). Observe-se também que há alguma evidência a favor da hipótese de homogeneidade dos tamanhos do efeito, ou seja, a favor de  $H_0 : \tau^2 = 0$  ( $Q_{\text{obs}} = 49.41$  e  $p_{\text{val}} = 0.339$ ).<sup>24</sup>

Considerando agora um MEA, obtemos para estimativa do valor médio da população de tamanhos do efeito  $\hat{\mu} = 1.046$ , ou seja, há um aumento médio de 4.6% no risco de mortalidade com a suplementação de antioxidantes. Como o IC a 95% para  $\mu$  é  $[1.007, 1.086]$ , estima-se que esse aumento médio varie entre 0.7 a 8.6%. Quanto à estimativa de máxima verosimilhança restrita de  $\tau^2$ , tem-se que  $\hat{\tau}_{\text{REML}}^2 = 0.0016$ .

Apesar de não haver evidência contra a hipótese de  $\tau^2 = 0$ , 6.9% da variabilidade total observada entre as estimativas pode ser atribuída à verdadeira heterogeneidade da população ( $I^2 = 6.9\%$ ), podendo no entanto essa percentagem chegar aos 34.3% (IC a 95% para  $I^2$ :  $[0.0, 34.3]$ ). Por outro lado, o valor  $H^2 = 1.04^2 = 1.082$  indica que há apenas um aumento de cerca de 8.2% na variabilidade total, quando comparada com a variabilidade que se espera observar sob a hipótese de homogeneidade, ou seja, quando apenas se espera variabilidade associada ao erro de amostragem dentro dos estudos.

Como nos resultados da função *metabin* não aparecem por defeito os erros padrão associados a  $\hat{\theta}$  e  $\hat{\mu}$ , podemos obtê-los com os comandos `meta.antiox$seTE.fixed` e `meta.antiox$seTE.random`, respectivamente. Assim,  $\hat{\sigma}_{\hat{\theta}} = 0.0119$  e  $\hat{\sigma}_{\hat{\mu}} = 0.0192$ .<sup>25</sup>

<sup>24</sup>A hipótese  $H_0 : \tau^2 = 0$  é equivalente a  $H_0 : \theta_1 = \dots = \theta_K = \theta$ .

<sup>25</sup>Os erros padrão para as estimativas dos tamanhos do efeito dos estudos podem ser obtidos com o comando `meta.antiox$seTE`.

Repare-se que neste exemplo as conclusões sob os dois modelos apontam no mesmo sentido, ou seja, que há alguma evidência a favor da suplementação de antioxidantes aumentar o risco de mortalidade, embora a variabilidade associada à estimativa do efeito síntese sob um MEA seja ligeiramente maior, como já era expetável.

*Observação 3.3.* Chama-se a atenção para o facto de no conjunto dos 47 estudos existir 43 estudos, ou seja, 91.5% dos estudos, onde não há evidência contra a hipótese do tamanho do efeito ser nulo, isto é, do risco de mortalidade ser igual nos dois grupos. Contudo, com a meta-análise conclui-se que afinal há alguma evidência a favor do risco ser maior com a suplementação de antioxidantes, quando apenas 3 estudos individuais, isto é, 6.4% dos estudos, evidenciam esse maior risco.

Em relação aos pesos atribuídos por cada modelo, observamos que os estudos Virtamo *et al.* (2003) e Goodman *et al.* (2004) são os que mais contribuem para o cálculo do efeito síntese nos dois modelos. De um modo geral, os pesos tendem a ser maiores num MEA, quando comparados com os correspondentes pesos no MEF, exceto para os estudos com maiores pesos neste último que serão menores no MEA.

*Observação 3.4.* Diferentes escolhas para o estimador de  $\tau^2$  podem levar a decisões diferentes sobre a hipótese  $H_0 : \ln \mu = 0$  ( $\mu = 1$ ), como podemos constatar no quadro abaixo para o Exemplo 1.

Estimador	$\hat{\tau}^2$	$\hat{\mu}$	$p_{\text{val}}$
HE	0	1.055	< 0.0001
DL	0.0006	1.050	0.0021
HS	0.0003	1.052	0.0004
SJ	0.1214	1.066	0.3976
ML	0.0012	1.047	0.0111
REML	0.0016	1.046	0.0199

Se para o exemplo da suplementação de antioxidantes nos tivéssemos limitado a adicionar as frequências observadas homólogas dos 47 estudos, ou seja, se tivéssemos tido em conta apenas a informação

Grupo	Mortalidade		Total
	Sim	Não	
Tratamento	15366	83729	99095
Controlo	9131	72719	81850

obteríamos para a estimativa do risco relativo o valor 1.390, com um IC a 95% igual a [1.357, 1.424]. Consequentemente, concluiríamos que a suplementação de antioxidantes aumenta o risco de mortalidade entre 35.7 a 42.4%, valores bastante acima dos obtidos com a meta-análise. Porém, esta abordagem não é adequada, porque não tem em conta a dispersão observada dentro de cada estudo para o tamanho do efeito.

Apesar de não se ter verificado para o exemplo anterior, podíamos não ter concluído o mesmo sentido para o efeito síntese ao combinarmos todas as observações dos estudos numa única tabela de contingência. Um exemplo onde as conclusões são completamente diferentes, consoante a abordagem adotada, é o da revisão realizada em Van Howe (1999), onde se procurou averiguar a existência de uma relação entre a circuncisão e a infeção pelo VIH no continente africano. Neste caso, Van Howe (1999) limitou-se a adicionar as frequências observadas homólogas dos 33 estudos selecionados, cujas informações se encontram na Tabela 3.2, tendo concluído que

*When the raw data are combined, a man with a circumcised penis is at greater risk of acquiring and transmitting HIV than a man with a non-circumcised penis.*

Tabela 3.2: Informação sobre os 33 estudos considerados em Van Howe (1999)

Estudo	Circuncidado		Não Circuncidado	
	VIH <sup>+</sup>	Total	VIH <sup>+</sup>	Total
Bwayo (a)	160	772	92	178
Bwayo (b)	37	237	22	68
Kreiss	254	422	59	77
Hira	10	20	418	590
Cameron	6	214	18	79
Pepin	13	256	5	47
Greenblatt	8	76	11	39
Diallo	212	1085	38	84
Simonsen	21	253	17	87
Tyndall	105	632	85	178
Nasio	137	717	86	164
Mehendal	38	291	837	4248
Bollinger	1	15	50	291
Chiasson	14	556	36	833
Sassan	415	636	75	93
Hunter	165	3930	43	373
Carael	34	79	90	195
Chao	75	307	442	5286
Moss	12	29	24	40
Allen	132	456	275	887
Sedlin	33	73	32	58
Konde-Luc	6	133	153	1669
Barongo	42	684	55	1411
Grosskurth	61	1087	158	4762
Van de Perre	6	32	46	270
Seed	51	243	171	593
Malamba	21	68	111	225
Quigley	48	169	101	373
Urssa 1	42	642	56	1357
Urssa 2	32	458	105	2145
Urssa 3	19	177	38	347
Urssa 4	54	746	112	828
Urssa 5	48	184	101	466

A informação combinada dos 33 estudos origina a tabela de contingência

Grupo	Infecção pelo VIH		Total
	Sim	Não	
Circuncidados	2312	13367	15679
Não circuncidados	3962	24379	28341

da qual se obtém  $\theta = \text{OR} = 1.064$ , com um IC a 95% igual a  $[1.007, 1.125]$ .<sup>26</sup>

Porém, quando se analisa o conjunto dos estudos observa-se que a conclusão de Van Howe entra em conflito com o facto de em apenas 4 estudos (12.1% dos estudos) se verificar uma maior proteção para homens não circuncidados ( $\text{OR} > 1$ ), enquanto em 15 estudos (45.5% dos estudos) uma maior proteção para homens circuncidados ( $\text{OR} < 1$ ). Nos restantes 14 estudos (42.4%) as hipóteses de infeção são iguais nos dois grupos.

A realização de uma meta-análise nos moldes aqui descritos levamos a concluir que afinal os homens circuncidados têm menos hipóteses de ser infetados pelo VIH do que os homens não circuncidados ( $\text{OR} = 0.603$  com um IC a 95% de  $[0.463, 0.784]$  sob um MEA).

```
> metabin(event.e=ai,n.e=n1i,event.c=ci,n.c=n2i,data=vanhowe,sm="OR",
+         method="Inverse",method.tau="REML",studlab=Estudo,
+         comb.fixed=F)
```

	OR	95%-CI	%W(random)
Bwayo (a)	0.244	[0.174; 0.344]	3.4
Bwayo (b)	0.387	[0.209; 0.717]	3.0
Kreiss	0.461	[0.263; 0.810]	3.1
Hira	0.411	[0.168; 1.006]	2.6
Cameron	0.098	[0.037; 0.257]	2.4
Pepin	0.449	[0.152; 1.326]	2.2
Greenblatt	0.299	[0.109; 0.823]	2.4
Diallo	0.294	[0.186; 0.463]	3.3
Simonsen	0.373	[0.186; 0.745]	2.9

<sup>26</sup>O tamanho do efeito usado em Van Howe (1999) foi o *odds ratio*.

Tyndall	0.218	[0.152; 0.313]	3.4
Nasio	0.214	[0.150; 0.307]	3.4
Mehendal	0.612	[0.432; 0.868]	3.4
Bollinger	0.344	[0.044; 2.678]	1.1
Chiasson	0.572	[0.306; 1.070]	3.0
Sassan	0.451	[0.263; 0.773]	3.1
Hunter	0.336	[0.236; 0.479]	3.4
Carael	0.881	[0.520; 1.493]	3.2
Chao	3.543	[2.683; 4.678]	3.5
Moss	0.471	[0.178; 1.245]	2.4
Allen	0.907	[0.708; 1.161]	3.5
Sedlin	0.670	[0.335; 1.340]	2.9
Konde-Luc	0.468	[0.203; 1.080]	2.7
Barongo	1.613	[1.068; 2.436]	3.3
Grosskurth	1.732	[1.279; 2.346]	3.5
Van de Perre	1.124	[0.438; 2.884]	2.5
Seed	0.656	[0.459; 0.936]	3.4
Malamba	0.459	[0.258; 0.817]	3.1
Quigley	1.068	[0.713; 1.602]	3.3
Urassa 1	1.626	[1.078; 2.454]	3.3
Urassa 2	1.459	[0.969; 2.198]	3.3
Urassa 3	0.978	[0.546; 1.752]	3.1
Urassa 4	0.499	[0.355; 0.702]	3.4
Urassa 5	1.275	[0.858; 1.895]	3.4

Number of studies combined: k = 33

	OR	95%-CI	z	p-value
Random effects model	0.603	[0.463; 0.784]	-3.77	0.0002

Quantifying heterogeneity:

$\tau^2 = 0.4974$ ;  $H = 3.62$  [3.21; 4.08];  $I^2 = 92.4\%$  [90.3%; 94.0%];  
 $R_b = 83.5\%$  [78.3%; 88.8%]

Test of heterogeneity:

Q	d.f.	p-value
419.01	32	< 0.0001

Details on meta-analytical method:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$

O exemplo de Van Howe ilustra o chamado *paradoxo de Simpson* que ocorre quando apesar de um grupo apresentar melhores resultados em cada estudo, apresenta os piores resultados quando os dados originais são simplesmente agregados.

## Exemplo 2

*Os programas de reabilitação ou reinserção social para jovens delinquentes funcionam?*

Na década de 70 do século passado a ideia que predominava na população em geral dos Estados Unidos (e Canadá) sobre a eficácia dos programas de reabilitação e reinserção social para jovens delinquentes, era que nada funcionava. O declínio do ideal da reabilitação levou a que alguns estados se procedesse à alteração de leis no sentido de desvalorizar o papel desses programas.

Por forma a comparar a taxa de reincidência de jovens infratores sujeitos e não sujeitos a programas de reabilitação ou reinserção social, Whitehead e Lab (1989) combinaram a evidência existente entre 1975 a 1984 através de uma meta-análise, baseada em 50 estudos selecionados a partir de artigos nas áreas da criminologia, psicologia e sociologia.<sup>27</sup>

Para o tamanho do efeito foi considerado o coeficiente de associação  $\phi = \mp \sqrt{X^2/n}$ , onde  $X^2$  representa a estatística do teste do qui-quadrado de Pearson, calculado para as variáveis binárias  $A$  – participação num programa de reabilitação ou reinserção (sim, não) e  $B$  – reincidência (sim, não), e  $n$  o número total de participantes no estudo. O coeficiente  $\phi$  foi usado porque possui uma interpretação semelhante ao do coeficiente de correlação de Pearson para variáveis binárias. Neste caso concreto, um sinal negativo para  $\phi$  indica uma maior reincidência para os jovens delinquentes no grupo experimental, ou seja, submetidos a um programa de reabilitação ou reinserção social.

---

<sup>27</sup>Não foram incluídos na meta-análise estudos sobre jovens com problemas de toxicod dependência.

Na Tabela 3.3 regista-se a informação sobre 45 dos 50 estudos considerados em Whitehead e Lab (1989), para os quais foi possível validar e completar a informação a partir de Andrews *et al.* (1990). Também se indica o tipo de programa usado: NSDP (*non system diversion programs*) – abrange programas que envolvem centros de correção não oficialmente ligados ao sistema judicial; SDP (*system diversion programs*) – inclui programas ligados ao sistema judicial; PPCC – liberdade condicional ou prestação de serviço comunitário; IR – institucionalização ou prisão domiciliária; SP – abrange outros programas mais específicos como, por exemplo, a visita dos jovens em risco a prisões para serem confrontados com a realidade dos reclusos (chamados os *scared straight programs*).

Para combinarmos os estudos da Tabela 3.3 usaremos a função *metagen* do pacote *meta*, função que nos permite combinar tamanhos do efeito mais genéricos como o coeficiente  $\phi$ . Neste caso, teremos que fornecer para argumentos da função as estimativas para os tamanhos do efeito dos estudos, ou seja, os valores  $\phi_k$ , e os respetivos erros padrão dados por  $\sqrt{\widehat{\text{Var}}(\phi_k)} = \sqrt{1/n_k}$ , em que  $n_k$  é o número total de participantes no estudo  $k$ ,  $k = 1, \dots, K$ .

```
> meta.programas=metagen(TE=phi, seTE=sqrt(1/n), data=whitehead, sm="phi",
+                          method.tau="REML", studlab=paste(Estudo, Ano,
+                                                              sep=","))
> summary(meta.programas)
```

Number of studies combined: k = 45

	phi	95%-CI	z	p-value
Fixed effect model	0.102	[ 0.085; 0.120]	11.55	< 0.0001
Random effects model	0.040	[-0.027; 0.108]	1.16	0.2446

Quantifying heterogeneity:

tau<sup>2</sup> = 0.0437; H = 2.99 [2.66; 3.35]; I<sup>2</sup> = 88.8% [85.9%; 91.1%];  
 Rb = 81.6% [63.9%; 99.3%]

Test of heterogeneity:

Q	d.f.	p-value
392.21	44	< 0.0001

Details on meta-analytical method:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$

Observe-se que há alguma evidência contra a hipótese de homogeneidade dos tamanhos do efeito ( $Q_{\text{obs}} = 392.21$  e  $p_{\text{val}} < 0.0001$ ). Por outro lado, a percentagem da variabilidade total que é explicada pela heterogeneidade populacional é bastante elevada, neste caso 88.8% (IC a 95% : [85.9,91.1]).

Assim, considerando adequado um MEA, estima-se que o coeficiente de associação  $\phi$  médio é 0.040, com um IC a 95% igual a  $[-0.027, 0.108]$ . Consequentemente, há alguma evidência a favor da hipótese de independência entre os programas de reabilitação e reinserção social e a reincidência. Estes resultados vêm de certa forma reforçar a ideia generalizada na população de que os programas de reabilitação e reinserção social não funcionam na prevenção da reincidência. Repare-se que sob um MEF a conclusão seria totalmente diferente, ou seja, uma menor taxa de reincidência para jovens delinquentes que participam em programas de reabilitação ou reinserção social (IC a 95% para  $\phi$ : [0.085,0.120]).

### 3.3 Métodos Baseados em Valores- $p$

Embora seja cada vez mais raro, ainda se encontram estudos que descrevem os seus resultados quase exclusivamente por um valor- $p$  ( $p$ -value), impossibilitando desta forma a combinação de informação nos moldes aqui descritos (a menos que se consiga aceder aos dados originais). No entanto, mesmo dispondo de toda a informação necessária, a violação do pressuposto de normalidade para o estimador do tamanho do efeito escolhido levanta problemas em termos de fiabilidade das inferências realizadas. Quer a primeira situação, quer a última, obriga-nos a recorrer a abordagens alternativas.

Tabela 3.3: Informação sobre os 45 estudos considerados em Whitehead e Lab (1989), validados a partir de Andrews *et al.* (1990)

Estudo	Ano	$\phi$	$n$	Programa
Hackler and Hagan	1975	-0.05	155	NSDP
Quay and Love	1977	0.07	360	NSDP
Bachra and Zaba	1978	0.38	79	NSDP
Sorenson	1978	-0.35	75	NSDP
Druckman	1979	-0.17	29	NSDP
Kelly et al.	1979	0.26	1278	NSDP
O'Donnell et al.	1979	-0.07	299	NSDP
Kraus	1981	- 0.15	156	NSDP
Farrington and Bennett	1981	0.10	26	NSDP
Shichor and Binder	1982	0.14	114	NSDP
Mott	1983	-0.08	193	NSDP
Whitaker and Severy	1984	0.10	1128	NSDP
Regoli et al.	1985	0.19	104	NSDP
Alexander et al.	1976	0.64	21	SDP
Collingwood et al.	1976	0.41	1009	SDP
Viano	1976	-0.08	56	SDP
Gilbert	1977	0.30	136	SDP
Klein et al.	1977	-0.11	86	SDP
Douds and Collingwood	1978	0,35	465	SDP
Gruher	1979	0.06	78	SDP
Byles and Maurice	1979	-0.12	208	SDP
Palmer and Lewis	1980	0.07	2537	SDP
Lipsey et al.	1981	0.18	1252	SDP
Willman and Snortum	1982	-0.11	136	SDP
McPherson et al.	1983	0.20	75	SDP
Rausch	1983	-0.01	214	SDP
Williams	1984	0.33	641	SDP
Barkwell	1976	-0.16	32	PPCC
Romig	1976	0.15	428	PPCC
Stringfield	1977	-0,31	52	PPCC
Kraus	1978	-0.28	180	PPCC
Horowitz and Wasserman	1982	- 0.22	263	PPCC
Byles	1981	-0.11	66	PPCC
Kratcoski and Kratcoski	1982	0.65	121	PPCC
Jackson	1983	0.03	296	PPCC
Mitchell	1983	0.29	92	PPCC
Schneider and Schneider	1984	0.18	251	PPCC
Jesness	1975	0.10	897	IR

Tabela 3.3: Informação sobre os 45 estudos considerados em Whitehead e Lab (1989), validados a partir de Andrews *et al.* (1990)

Estudo	Ano	$\phi$	$n$	Programa
Maskin	1976	-0.39	60	IR
Davidson and Wolfred	1977	-0.26	79	IR
Ross and McKay	1977	0.27	25	IR
Clarke and Cornish	1978	-0.01	173	IR
Adams and Vetter	1982	-0.44	27	IR
Kirigin <i>et al.</i>	1982	0.21	68	IR
Buckner and Chesney-Lind	1983	-0.04	200	SP
Lewis	1983	-0.16	108	SP

Se o que estiver em jogo numa meta-análise for simplesmente averiguar se o efeito síntese é nulo, nesta secção veremos métodos estatísticos, considerados não paramétricos, que recorrem à simples combinação dos valores- $p$  reportados.<sup>28</sup>

Consideremos então  $K$  estudos independentes que foram realizados com o objetivo de testar uma hipótese comum, como  $H_{0k} : \theta_k = 0$  contra  $H_{1k} : \theta_k > 0$ ,  $k = 1, \dots, K$ . Um teste combinado considera como hipóteses globais

$$H_0 : H_{0k} \text{ é verdadeira para qualquer } k = 1, \dots, K, \quad (3.11)$$

*vs.*

$$H_1 : \text{Existe um } k \in \{1, \dots, K\} \text{ tal que } H_{1k} \text{ é verdadeira.}$$

Atendendo a que há várias maneiras da hipótese nula global ser falsa, torna-se difícil neste contexto selecionar um teste adequado. No entanto, qualquer teste combinado que se venha definir deve ser:

- i) admissível, no sentido em que fornece o teste mais potente, não necessariamente único, contra uma alternativa para uma combinação de uma coleção de testes; e

<sup>28</sup>Como nota histórica, a combinação de testes através de valores- $p$  figura entre os primeiros resultados da Meta-Análise (Tippett, 1931; Fisher, 1932).

- ii) monótono, ou seja, se o teste combinado rejeita a hipótese  $H_0$  para um conjunto de valores- $p$ , então também a deve rejeitar para qualquer subconjunto de valores- $p$  de magnitudes inferiores.

Birnbaum (1954) mostrou que qualquer teste combinado monótono é admissível e, por conseguinte, ótimo para um problema de teste de hipóteses.

Os métodos de combinação de testes conhecidos que recorrem à combinação de valores- $p$  assentam no facto de, sob a validade da hipótese nula global, a amostra de valores- $p$  observados,  $(p_1, \dots, p_K)$ , ser uma realização da amostra aleatória  $(P_1, \dots, P_K)$  de uma população uniforme padrão. Assim, estes testes têm em conta as propriedades da distribuição uniforme, ou recorrem a métodos de transformação de probabilidade.

Os métodos de combinação de testes mais conhecidos são:

- **Método de Tippett**

O método de Tippett (Tippett, 1931) usa como estatística de teste

$$P_{1:K} = \min(P_1, \dots, P_K) \sim_{|H_0} \text{Beta}(1, K).$$

Denotando por  $B_{a,b;\gamma}$  o quantil de probabilidade  $\gamma$  da distribuição  $\text{Beta}(a, b)$ ,  $H_0$  será rejeitada ao nível de significância  $\alpha$  se  $p_{1:K} < B_{1,K;\alpha} = 1 - (1 - \alpha)^{1/K}$ ;

- **Método de Wilkinson**

O método de Wilkinson (Wilkinson, 1951) generaliza o método de Tippett no sentido em que considera a  $r$ -ésima estatística ordinal da amostra, em vez do mínimo, para estatística de teste, ou seja,

$$P_{r:K} \sim_{|H_0} \text{Beta}(r, K - r + 1), \quad r = 1, \dots, K.$$

Neste caso,  $H_0$  será rejeitada ao nível de significância  $\alpha$  se  $p_{r:K} < B_{r,K-r+1;\alpha}$ ;

- **Método de Fisher**

O método de Fisher (Fisher, 1932), ao contrário dos dois métodos anteriores, baseia-se na transformação de probabilidade  $-2 \ln P \sim \chi_{2K}^2$ , com  $P \sim \text{Uniforme}[0, 1]$ . A estatística de teste deste método é

$$T = -2 \sum_{k=1}^K \ln P_k \sim_{|H_0} \chi_{2K}^2,$$

sendo  $H_0$  rejeitada ao nível de significância  $\alpha$  se  $T_{\text{obs}} > \chi_{2K; 1-\alpha}^2$ ;

- **Método de Stouffer**

O método de Stouffer (Stouffer *et al.*, 1949) tem em conta que  $\Phi^{-1}(P) \sim \mathcal{N}(0, 1)$ , onde  $\Phi^{-1}$  representa a inversa da função de distribuição da normal padrão e  $P \sim \text{Uniforme}[0, 1]$ . Assim, a estatística de teste é

$$Z = \sum_{k=1}^K \frac{\Phi^{-1}(1 - P_k)}{\sqrt{K}} \sim_{|H_0} \mathcal{N}(0, 1).$$

A hipótese nula global será rejeitada ao nível de significância  $\alpha$  se  $Z_{\text{obs}} > z_{1-\alpha}$ ;

- **Método Logit**

O método logit (George, 1977) baseia-se na função logit  $\ln\left(\frac{P}{1-P}\right)$ , com  $P \sim \text{Uniforme}[0, 1]$ , e usa como estatística de teste

$$G = - \sum_{k=1}^K \ln\left(\frac{P_k}{1 - P_k}\right) \left[ \frac{K\pi^2(5K + 2)}{3(5K + 4)} \right]^{-1/2} \underset{\sim}{\sim}_{|H_0} t_{5K+4},$$

onde  $t_\nu$  representa a distribuição de  $t$  de Student com  $\nu$  graus de liberdade. Denotando por  $t_{\nu; \gamma}$  o quantil de probabilidade  $\gamma$  da distribuição  $t$  com  $\nu$  graus de liberdade,  $H_0$  será rejeitada ao nível de significância  $\alpha$  se  $G_{\text{obs}} > t_{5K+4; 1-\alpha}$ .<sup>29</sup>

<sup>29</sup>Existem outros métodos de combinação como, por exemplo, o método da média aritmética, cuja estatística de teste é  $\bar{P} = \frac{1}{K} \sum_{k=1}^K P_k$ , e que possui distribuição aproximadamente normal para  $K$  elevado. Por outro lado, Hartung (1999) e Makambi (2003) apresentam métodos para combinar valores- $p$  dependentes.

Em relação aos métodos anteriores não há propriamente critérios que nos possam guiar na escolha do melhor método. Contudo, Hedges e Olkin (1985) sumarizam resultados sobre o desempenho de alguns métodos, concluindo que de um modo geral o teste de Fisher é o melhor. Por outro lado, o facto dos métodos de combinação de testes independentes assentar no pressuposto de uniformidade da amostra dos valores- $p$ , a violação deste invalida a aplicação dos métodos descritos.<sup>30</sup>

Com o pacote `metap` do R podemos aplicar os métodos de combinação anteriores. Por exemplo, a aplicação do método de Tippett, caso particular do método de Wilkinson, é feita com a função `wilkinsonp`, o método de Fisher com a função `sumlog`, o método de Stouffer com a função `sumz` e o método logit com a função `logitp`.

Para aplicarmos os métodos de combinação ao exemplo da suplementação de antioxidantes, consideremos os valores- $p$  associados aos testes unilaterais à direita  $H_{0k} : \theta_k = 0$  contra  $H_{1k} : \theta_k > 0$ , onde  $\theta_k$  representa o logaritmo do risco relativo no estudo  $k$ ,  $k = 1, \dots, 47$ .<sup>31</sup> Assim, considerando as hipóteses globais

$H_0$  : Todos os  $H_{0k}$  são verdadeiros vs.  $H_1$  : Algum  $H_{1k}$  é verdadeiro ,

obtemos os seguintes resultados

Método	Valor da estatística	$p_{\text{val}}$
Tippett	$1.01 \times 10^{-6}$	$4.75 \times 10^{-5}$
Fisher	143.797	$7.28 \times 10^{-4}$
Stouffer	2.671	$3.79 \times 10^{-3}$
Logit	29.816	$1.34 \times 10^{-3}$

Consequentemente, somos levados a rejeitar a hipótese nula global ao nível de significância de 5% (até mesmo a 1%) com cada um

<sup>30</sup>Por exemplo, quando há valores- $p$  associados a estatísticas de teste com distribuição discreta, pode-se aplicar fatores de correção de continuidade.

<sup>31</sup>Os *scores* normais  $z_k = \hat{\theta}_k / \sqrt{\text{Var}(\hat{\theta}_k)}$  podem ser obtidos com o comando `meta.antiox$zval` e os correspondentes valores- $p$  associados ao teste  $z$  (bilateral) com o comando `meta.antiox$pval`.

dos métodos. Repare-se que os resultados anteriores não nos fornecem qualquer informação sobre o valor do efeito síntese, apenas nos indicam se este pode ser considerado nulo ou não.

### 3.3.1 Valores- $p$ Generalizados e Aleatórios

#### Valores- $p$ Generalizados

Os métodos de combinação que acabámos de ver assentam no pressuposto de uniformidade da amostra de valores- $p$ . Todavia, não há a garantia de que uma amostra de valores- $p$  seja constituída exclusivamente por valores genuínos, isto é, por realizações (independentes) de uma variável aleatória com distribuição uniforme padrão. Por exemplo, é sabido que alguns investigadores, quando descontentes com os resultados obtidos, repetem a experiência na esperança de obter “melhores resultados”, publicando o que consideram ser o “melhor valor- $p$ ”. Por outro lado, na impossibilidade de se poder calcular um valor- $p$  vulgar, em virtude de existir parâmetros perturbadores não elimináveis por técnicas estatísticas usuais, pode-se ter na amostra *valores- $p$  generalizados*, conceito introduzido por Tsui e Weerahandi (1989), que não são realizações de uma variável aleatória uniforme padrão.

Perante a possibilidade de haver valores- $p$  problemáticos numa amostra selecionada, torna-se crucial recorrer ou desenvolver metodologias estatísticas de combinação apropriadas que contemplem este tipo de situações. Por exemplo, o conceito de valor- $p$  generalizado pode ser útil em Meta-Análise quando não se puder usar o teste  $Q$  de Cochran para testar a homogeneidade dos tamanhos do efeito, por se estar a considerar estimadores para os tamanhos do efeito com distribuição não normal.

Brilhante (2013) aplicou o conceito para testar a hipótese de uniformidade de uma amostra de valores- $p$ , face à alternativa de existir valores- $p$  problemáticos na amostra que são realizações de uma mistura convexa entre uma variável uniforme padrão e o mínimo de duas

variáveis uniformes padrão independentes, ou seja, uma Beta(1,2).

Um exemplo clássico onde não é possível obtermos valores- $p$  vulgares é no problema de Behrens-Fisher, que visa comparar os valores médios de duas ( $k = 2$ ) populações normais com variâncias desconhecidas diferentes. Apesar de ter sido sugerido várias soluções para o problema, estas soluções não coincidem por inexistência de critérios de otimalidade para a seleção do melhor teste.<sup>32</sup> Embora haja soluções aproximadas para o caso  $k > 2$ , o conceito de valor- $p$  generalizado veio abrir portas para se encontrar uma solução exata quando parâmetros perturbadores estão envolvidos e a abordagem frequentista usual falha.

O conceito de valor- $p$  generalizado, que estende o conceito vulgar de valor- $p$ , permite-nos encontrar estatísticas de teste com uma distribuição amostral que não depende de parâmetros perturbadores e, por conseguinte, ser possível calcular um valor- $p$  exato. Para percebermos o conceito, suponhamos que  $X$  é uma variável aleatória com distribuição a depender do parâmetro de interesse  $\theta$  e do parâmetro perturbador  $\zeta$  (real ou vetorial). Suponhamos também que se pretende testar, por exemplo, as hipóteses

$$H_0 : \theta \leq \theta_0 \quad \text{vs.} \quad H_1 : \theta > \theta_0. \quad (3.12)$$

Seja  $\mathbf{X} = (X_1, \dots, X_n)$  uma amostra aleatória da população alvo. A abordagem usual neste tipo de problema é tentar encontrar uma estatística de teste  $T = T(\mathbf{X}; \theta, \zeta)$  com distribuição amostral que não dependa de  $\zeta$ . Se  $T$  for estocasticamente crescente em  $\theta$ , ou seja, se  $P(T > t; \theta, \zeta)$  for uma função não decrescente em  $\theta$  para todo o  $t$  e  $\zeta$  fixos, então o valor- $p$  associado a  $T$  para (3.12) é dado por

$$p_{\text{val}} = P(T(\mathbf{X}; \theta, \zeta) \geq T(\mathbf{x}; \theta, \zeta) | \theta = \theta_0),$$

em que  $\mathbf{x} = (x_1, \dots, x_n)$  é o valor observado de  $\mathbf{X}$ .<sup>33</sup>

---

<sup>32</sup>A primeira solução conhecida deve-se a Behrens (1929), que foi igualmente obtida por Fisher (1939) usando um argumento fiducial, mas vários outros autores se debruçaram sobre este problema recorrendo a abordagens diferentes. Para mais informações sobre o assunto, consultar, por exemplo, Brillhante *et al.* (2011).

<sup>33</sup>Se  $T$  for estocasticamente decrescente em  $\theta$ ,  $p_{\text{val}} = P(T(\mathbf{X}; \theta, \zeta) \leq T(\mathbf{x}; \theta, \zeta) | \theta = \theta_0)$ .

Para introduzirmos o conceito de valor- $p$  generalizado é necessário introduzirmos primeiro o conceito de variável de teste generalizada.

**Definição 3.1.** Uma variável aleatória do tipo  $T(\mathbf{X}; \mathbf{x}, \theta, \zeta)$  é uma variável de teste generalizada para o parâmetro  $\theta$  se verificar as seguintes propriedades:

- i) O valor observado de  $T(\mathbf{X}; \mathbf{x}, \theta, \zeta)$ , isto é,  $T(\mathbf{x}; \mathbf{x}, \theta, \zeta)$ , não depende de  $\theta$  nem de  $\zeta$ ;
- ii) Quando o valor de  $\theta$  é especificado, a distribuição de  $T(\mathbf{X}; \mathbf{x}, \theta, \zeta)$  não depende de  $\zeta$ ;
- iii) Para valores fixos de  $\mathbf{x}$  e  $\zeta$ ,  $P(T(\mathbf{X}; \mathbf{x}, \theta, \zeta) \leq t)$  é uma função monótona em  $\theta$  para qualquer valor de  $t$ .

Consequentemente, se  $T(\mathbf{X}; \mathbf{x}, \theta, \zeta)$  for estocasticamente crescente em  $\theta$ , o valor- $p$  generalizado será calculado da forma

$$p_{\text{gen}} = P(T(\mathbf{X}; \mathbf{x}, \theta, \zeta) \geq T(\mathbf{x}; \mathbf{x}, \theta, \zeta) | \theta = \theta_0) ,$$

valor este que não dependerá de  $\zeta$ .

Sobre a relevância dos valores- $p$  generalizados, Weerahandi (1993) afirma que

*Generalized p-values exist even in those situations where fixed level tests based on sufficient statistics do not exist.*

Por outro lado, sabemos que os valores- $p$  generalizados se comportam de forma semelhante à abordagem frequencista usual.<sup>34</sup> Todavia, não existe um método geral para a obtenção de variáveis fulcrais generalizadas, que possam depois ser usadas na construção de intervalos de confiança generalizados e de variáveis de teste generalizadas. Sobre este assunto Weerahandi (1993) afirma que

---

<sup>34</sup>Uma das críticas apontadas à solução de Behrens-Fisher é que os intervalos de confiança obtidos não têm uma interpretação frequencista.

*The problem of finding an appropriate pivotal quantity is a nontrivial task (...).*

Porém, Iyer e Patterson (2002) avançam com uma “receita” simples, não necessariamente “ótima”, que permite construir variáveis fulcrais e variáveis de teste generalizadas.

Suponhamos que um problema de inferência envolve  $m$  parâmetros desconhecidos  $\xi = (\xi_1, \dots, \xi_m)$  e que o parâmetro de interesse é o escalar  $\theta = f(\xi_1, \dots, \xi_m)$ . Se existir um conjunto de estatísticas  $\{U_1, \dots, U_m\}$  e um outro conjunto de estatísticas fulcrais invertíveis  $\{V_1, \dots, V_m\}$ , este último relacionando  $\{U_1, \dots, U_m\}$  e  $(\xi_1, \dots, \xi_m)$ , então ao explicitarmos o parâmetro  $\theta$  em função dos  $U_j$  e  $V_j$ 's, ou seja,  $\theta = g(U_1, \dots, U_m, V_1, \dots, V_m)$ , obteremos uma variável fulcral generalizada que pode ser usada para efeitos de inferência sobre  $\theta$ , nomeadamente  $T = \theta - g(u_1, \dots, u_m, V_1, \dots, V_m)$ , em que  $u_j$  é o valor observado de  $U_j$ ,  $j = 1, \dots, m$ . Por outro lado, a variável  $R = g(u_1, \dots, u_m, V_1, \dots, V_m)$  é uma variável fulcral generalizada que pode ser usada para construir intervalos de confiança generalizados para  $\theta$ .

O procedimento de Iyer e Patterson garante-nos que  $T$  é uma variável de teste generalizada para testar  $H_0 : \theta \leq \theta_0$  vs.  $H_1 : \theta > \theta_0$ , e como  $T$  é estocasticamente crescente em  $\theta$ , tem-se que  $p_{\text{gen}} = P(T \geq T_{\text{obs}} | \theta = \theta_0) = P(T \geq 0 | \theta = \theta_0)$ . Apesar do procedimento anterior ser muito vago sobre o vetor de estatísticas  $(U_1, \dots, U_m)$ , deve-se procurar que este se baseie em estatísticas suficientes para o parâmetro  $\xi$ .

Vejamos como o conceito de valor- $p$  generalizado fornece uma solução exata para o problema de Behrens-Fisher.

Sejam  $\mathbf{X}_1 = (X_{11}, \dots, X_{1n_1})$  e  $\mathbf{X}_2 = (X_{21}, \dots, X_{2n_2})$  duas amostras aleatórias independentes provenientes das populações  $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$  e  $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ , respetivamente. Suponhamos que se pretende testar

$$H_0 : \theta = \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \theta = \mu_1 - \mu_2 \neq 0, \quad (3.13)$$

com  $\sigma_1^2 \neq \sigma_2^2$  desconhecidos. Neste caso temos  $\xi = (\theta, \sigma_1, \sigma_2)$ , sendo

o parâmetro de interesse  $\theta = \mu_1 - \mu_2 = \xi_1$  e  $\zeta = (\xi_2, \xi_3) = (\sigma_1, \sigma_2)$  o parâmetro perturbador.

Consideremos as estatísticas  $U_1 = \hat{\theta} = \bar{X}_1 - \bar{X}_2$ ,  $U_2 = \hat{\sigma}_1^2 = S_1^2$  e  $U_3 = \hat{\sigma}_2^2 = S_2^2$ , e as variáveis fulcrais invertíveis

$$V_1 = \frac{\bar{X}_1 - \bar{X}_2 - \theta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim \mathcal{N}(0, 1), V_2 = \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi_{n_1-1}^2 \text{ e } V_3 = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi_{n_2-1}^2$$

independentes. Como

$$\begin{aligned} \theta &= g(U_1, U_2, U_3, V_1, V_2, V_3) = U_1 - V_1 \sqrt{\frac{(n_1-1)U_2}{n_1V_2} + \frac{(n_2-1)U_3}{n_2V_3}} \\ &= \bar{X}_1 - \bar{X}_2 - V_1 \sqrt{\frac{(n_1-1)S_1^2}{n_1V_2} + \frac{(n_2-1)S_2^2}{n_2V_3}}, \end{aligned}$$

a variável de teste generalizada para testarmos  $H_0^* : \theta \leq 0$  vs.  $H_1^* : \theta > 0$  é

$$T = T(\mathbf{X}; \mathbf{x}, \theta, \zeta) = \theta - \bar{x}_1 + \bar{x}_2 + V_1 \sqrt{\frac{(n_1-1)s_1^2}{n_1V_2} + \frac{(n_2-1)s_2^2}{n_2V_3}},$$

onde  $\mathbf{X} = (\bar{X}_1 - \bar{X}_2, S_1^2, S_2^2)$  e  $\mathbf{x} = (\bar{x}_1 - \bar{x}_2, s_1^2, s_2^2)$ . Logo,

$$p_{\text{gen}}^* = \mathbb{P}\left(T \geq 0 \mid \theta = 0\right) = \mathbb{P}\left(V_1 \sqrt{\frac{(n_1-1)s_1^2}{n_1V_2} + \frac{(n_2-1)s_2^2}{n_2V_3}} \geq \bar{x}_1 - \bar{x}_2\right).$$

Como as hipóteses (3.13) podem ser rescritas da forma

$$H_0 : |\theta| = 0 \quad \text{vs.} \quad H_1 : |\theta| > 0,$$

vem

$$\begin{aligned} p_{\text{gen}} &= \mathbb{P}\left(\left|V_1 \sqrt{\frac{(n_1-1)s_1^2}{n_1V_2} + \frac{(n_2-1)s_2^2}{n_2V_3}}\right| \geq |\bar{x}_1 - \bar{x}_2|\right) \\ &= \mathbb{P}\left(\left(V_1 \sqrt{\frac{(n_1-1)s_1^2}{n_1V_2} + \frac{(n_2-1)s_2^2}{n_2V_3}}\right)^2 \geq (\bar{x}_1 - \bar{x}_2)^2\right), \end{aligned}$$

probabilidade que pode ser obtida por simulação de Monte-Carlo.

Por conseguinte, um intervalo de confiança generalizado a  $(1 - \alpha)100\%$  para  $\theta$  é  $[R_{\alpha/2}, R_{1-\alpha/2}]$ , em que  $R_\gamma$  denota o quantil de probabilidade  $\gamma$  da distribuição de  $R = V_1 \sqrt{\frac{(n_1-1)s_1^2}{n_1V_2} + \frac{(n_2-1)s_2^2}{n_2V_3}}$ .

### Valores- $p$ Aleatórios

Um outro aspecto que passa frequentemente despercebido quando se lida com valores- $p$  é que estes têm um comportamento estocástico. Kulinskaya *et al.* (2008) afirmam que

*If one wants to compare  $p$ -values from different experiments, or even to combine the evidence in them as in meta analysis, one must take into account their distributional properties.*

Sob a validade de uma hipótese nula, qualquer valor- $p$  observado, calculado a partir de uma estatística de teste contínua, é a realização de uma variável aleatória  $P \sim \text{Uniforme}[0, 1]$ . Por isso, se o objetivo for combinar testes independentes através de valores- $p$  quando há suspeitas da hipótese alternativa ser verdadeira, o que faz sentido é combinar esses valores sob a hipótese alternativa.

Suponhamos que  $T$  é uma estatística de teste contínua para testar hipóteses sobre um parâmetro de interesse  $\theta$ . Sejam  $F_0$  e  $F_\theta$  as funções de distribuição de  $T$  sob a validade da hipótese nula e sob a validade da hipótese alternativa, respetivamente. Se valores elevados de  $T$  revelam evidência a favor da hipótese alternativa, o valor- $p$  observado é dado por  $p_{\text{val}} = P(T \geq t | H_0) = 1 - F_0(t)$ , onde  $t = T_{\text{obs}}$ . Por analogia definimos valor- $p$  aleatório por  $P = 1 - F_0(T)$ , verificando-se que sob a validade de  $H_0$ ,  $P \sim \text{Uniforme}[0, 1]$ . Quanto à distribuição de  $P$  sob a validade de  $H_1$ , para um  $0 \leq p \leq 1$ ,

$$F_\theta(p) = P_\theta(P \leq p) = P_\theta(1 - F_0(T) \leq p) = 1 - F_\theta(F_0^{-1}(1 - p)).$$

O conceito de valor- $p$  aleatório também pode ser útil em Meta-Análise, principalmente quando na combinação de testes independentes não se puder descartar a possibilidade de existir valores- $p$  problemáticos na amostra. Por exemplo, se houver suspeitas de existir na amostra valores- $p$  problemáticos que são mínimos ou máximos de dois valores- $p$

genuínos, isto porque alguns investigadores decidiram repetir a experiência e publicar o que consideram ser o “melhor” resultado, então  $P$  passa a ser uma mistura convexa de três variáveis aleatórias com distribuições uniforme padrão, Beta(1,2) e Beta(2,1), ou seja, com função densidade de probabilidade

$$f_P(p) = (1 - \alpha - \beta + 2\alpha p + 2\beta(1 - p))I_{[0,1]}(p),$$

com  $\alpha, \beta \in [0, 1]$  e  $0 \leq \alpha + \beta \leq 1$ . Neste cenário, para podermos usar, por exemplo, o método de combinação de Fisher, teremos que o fazer atendendo à distribuição de  $P$ , ou seja, considerando a distribuição da estatística de teste  $-2 \sum_{k=1}^K \ln P_k$ , em que  $P_k \stackrel{d}{=} P$ ,  $k = 1, \dots, K$ .<sup>35</sup> Para mais detalhes sobre valores- $p$  problemáticos consultar Semblano (2016).

---

<sup>35</sup>A notação “ $\stackrel{d}{=}$ ” significa “tem a mesma distribuição de”.

## Capítulo 4

# Representações Gráficas para a Meta-Análise

*I cannot see the forest for the trees.*

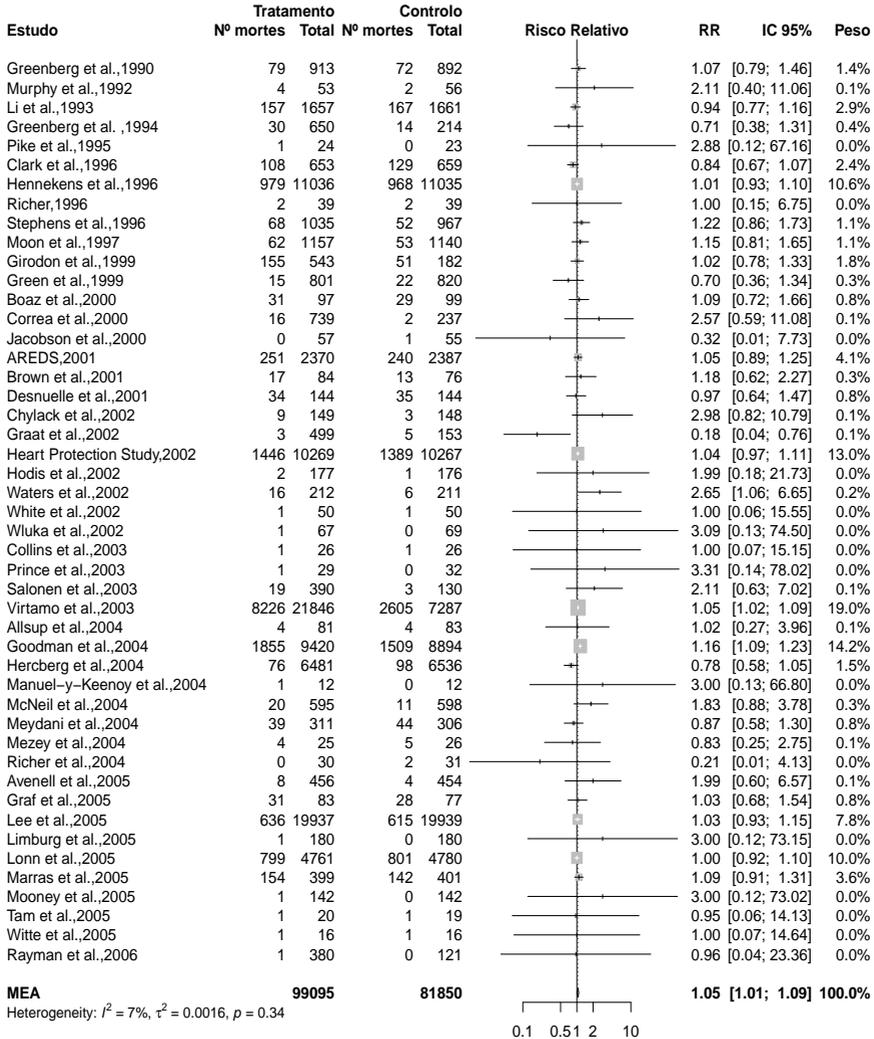
Um ditado inglês

### 4.1 Gráfico Floresta

Um gráfico floresta é basicamente um gráfico de intervalos de confiança, onde se representa as estimativas dos tamanhos do efeito para cada estudo, assim como os intervalos de confiança associados aos respectivos tamanhos do efeito. Cada estimativa é assinalada no gráfico por um quadrado com uma área proporcional ao peso atribuído ao estudo pelo modelo considerado. A estimativa do efeito síntese é indicada na parte inferior do gráfico por um losango, com a diagonal horizontal a ter um comprimento igual à amplitude do intervalo de confiança para este parâmetro.

Os gráficos floresta são obtidos no R com a função *forest* aplicada a objetos de meta-análise criados com o pacote *meta*.

Figura 4.1: Gráfico floresta para a meta-análise da suplementação de antioxidantes para um MEA



O comando usado para produzir a Figura 4.1 foi o seguinte:

```
> forest(meta.antiox,comb.fixed=F,text.random="MEA",
+       leftlabs=c("Estudo","Nº mortes","Total","Nº mortes","Total"),
+       rightlabs=c("RR","IC a 95%","Peso"),lab.e = "Tratamento",
+       lab.c = "Controlo",smlab="Risco Relativo")
```

Como podemos observar na Figura 4.1, é visível um número considerável de estudos que apresentam uma grande variabilidade associada à estimativa do tamanho do efeito, assim como um elevado número de estudos para os quais não há evidência contra a hipótese do risco de mortalidade ser igual nos dois grupos. O losango que está associado à estimativa do efeito síntese para o MEA é, neste caso, quase imperceptível face à maior precisão desta estimativa em relação à precisão de várias estimativas individuais.

Na Figura 4.2 representa-se o gráfico floresta para a meta-análise da taxa de reincidência associada a programas de reabilitação e reinserção social.

## 4.2 Gráfico Radial

Um gráfico radial, ou gráfico de Galbraith (Galbraith, 1988), é um gráfico de dispersão onde se representa a estimativa do tamanho do efeito estandardizada, isto é, o *score* normal  $z_k = \hat{\theta}_k / \sqrt{\text{Var}(\hat{\theta}_k)}$ , em função do inverso do erro padrão da estimativa, ou seja,  $1/\sqrt{\text{Var}(\hat{\theta}_k)}$ ,  $k = 1, \dots, K$ .

Num gráfico radial as estimativas menos precisas estarão mais perto da origem do referencial, enquanto as mais precisas mais afastadas da origem. Este tipo de gráfico permite-nos detetar visualmente a existência de heterogeneidade, bem como identificar possíveis *outliers*. Assim, se observarmos uma grande dispersão para os pontos, isto significa que há evidência contra a hipótese de homogeneidade dos tamanhos do efeito.

Figura 4.2: Gráfico floresta para a meta-análise da taxa de reincidência associada a programas de reabilitação e reinserção social para um MEA

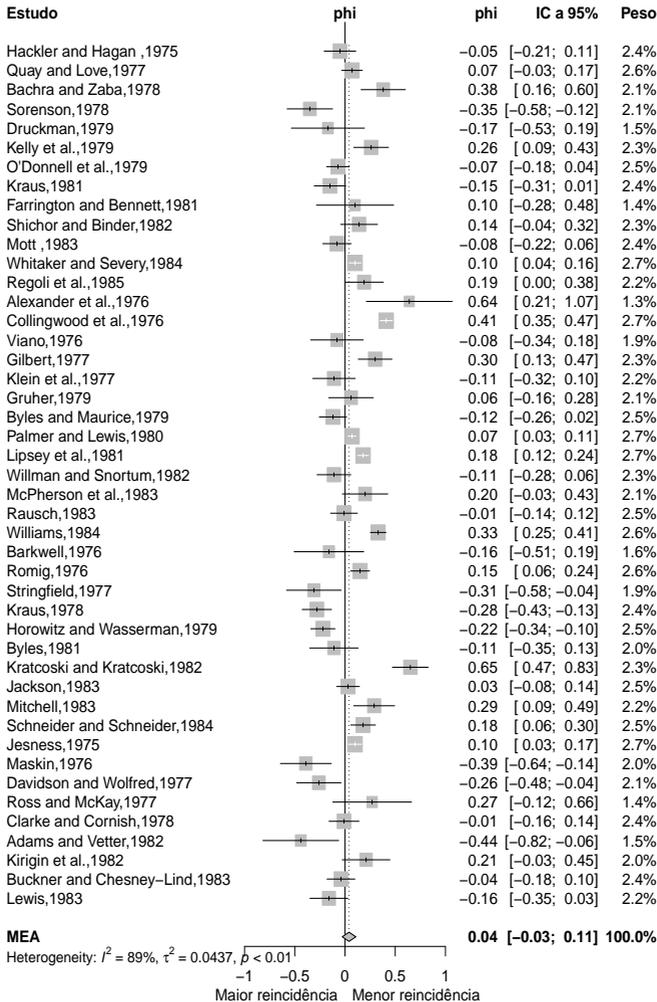
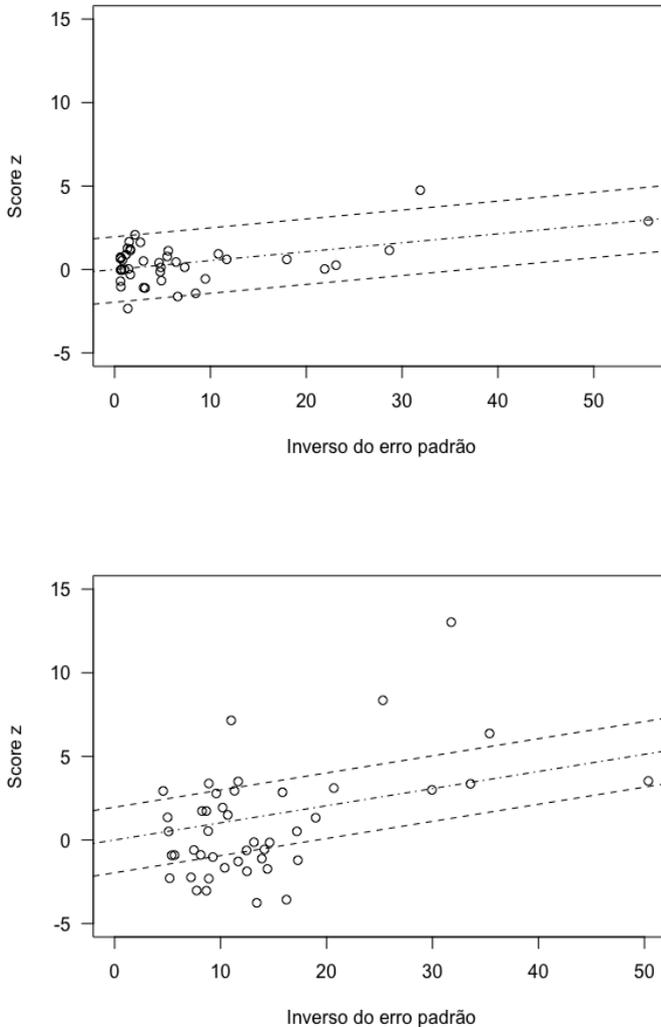


Figura 4.3: Gráficos radiais para as meta-análises da suplementação de antioxidantes (em cima) e taxa de reincidência associada a programas de reabilitação e reinserção social (em baixo)



Também se costuma indicar num gráfico radial uma reta que passa na origem e com declive igual ao  $z$ -score da estimativa do efeito síntese, obtida a partir de um MEF. Retas paralelas à reta anterior podem ser traçadas limitando uma região de confiança a 95%. Ainda poderá ser desenhado à direita do gráfico um arco, a partir do qual se poderá recuperar os valores originais para as estimativas dos tamanhos dos efeitos.

Os gráficos radiais podem ser obtidos no R com a função *radial* aplicada a objetos de meta-análise.<sup>1</sup> O gráfico radial para a meta-análise da suplementação de antioxidantes da Figura 4.3 foi obtido com o comando:

```
> radial(meta.antiox,xlab="Inverso do erro padrão",ylab="Score z",
+       level=0.95)
```

Como podemos observar no primeiro gráfico da Figura 4.3, há um maior número de estimativas que se encontram perto da origem do referencial e que, por isso, são menos precisas. Dado que não se observa muita dispersão para os pontos, não há propriamente grandes indícios a favor da existência de heterogeneidade. No gráfico radial para a meta-análise da taxa de reincidência associada a programas de reabilitação e reinserção social, já é visível uma maior dispersão para os pontos, indicando assim uma maior heterogeneidade entre as estimativas individuais.

### 4.3 Gráfico de Baujat

Um gráfico de Baujat (Baujat *et al.*, 2002) é um gráfico de dispersão que permite identificar quais os estudos que são fonte de heterogeneidade, especialmente quando a variabilidade observada entre as estimativas do tamanho do efeito não pode ser imputada à variância da

---

<sup>1</sup>O arco no gráfico radial não será desenhado quando se aplica a função *radial* a objetos criados com o pacote `meta`, mas sim para objetos criados com a função *rma* do pacote `metafor`.

população de tamanhos do efeito.

Num gráfico de Baujat representa-se no eixo horizontal a contribuição individual de cada estudo  $k$  para o valor da estatística  $Q$  de Cochran, definida em (3.5), ou seja,  $x_k = \omega_k(\hat{\theta}_k - \hat{\theta})^2$ , e no eixo vertical a influência global da estimativa no efeito síntese, isto é,  $y_k = (\hat{\theta}_{(-k)} - \hat{\theta})^2 / \sigma_{(-k)}^2$ , onde

$$\hat{\theta}_{(-k)} = \frac{\sum_{\substack{j=1 \\ j \neq k}}^K \omega_j \hat{\theta}_j}{\sum_{\substack{j=1 \\ j \neq k}}^K \omega_j}$$

é o efeito síntese sem o estudo  $k$ ,  $\omega_j = 1/\sigma_j^2$ , e

$$\sigma_{(-k)}^2 = \text{Var}(\hat{\theta}_{(-k)}) = \left( \sum_{\substack{j=1 \\ j \neq k}}^K \omega_j \right)^{-1} \quad k = 1, \dots, K.$$

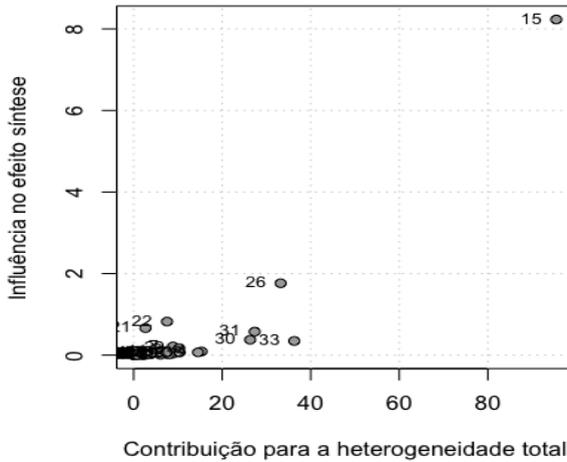
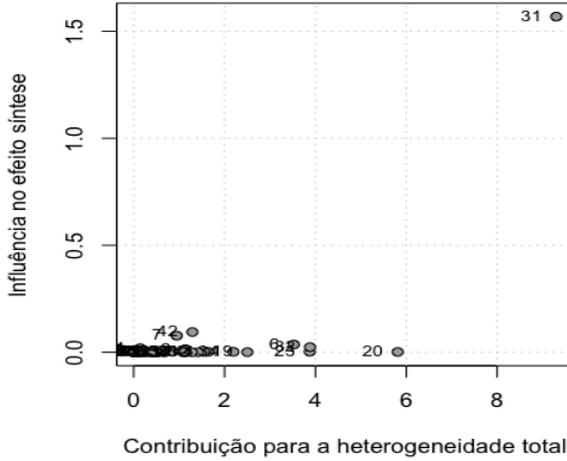
A lógica por detrás deste tipo de representação gráfica é que a exclusão de um estudo influente da meta-análise provocará um maior aumento na variância global do que a exclusão de um estudo menos influente. Assim, os pontos associados a estudos mais heterogêneos e mais influentes são aqueles que surgem no canto superior direito do gráfico. Por outro lado, como  $X_k = \omega_k(\hat{\theta}_k - \hat{\theta})^2 \stackrel{a}{\sim} \chi_1^2$ , haverá evidência do estudo  $k$  ser uma fonte de heterogeneidade ao nível de significância  $\alpha$  se  $x_k > \chi_{1;1-\alpha}^2$ .

Na Figura 4.4 representa-se os gráficos de Baujat para as meta-análises da suplementação de antioxidantes e taxa de reincidência associada a programas de reabilitação e reinserção social, tendo sido usado para produzir o primeiro gráfico o seguinte comando:

```
> baujat(meta.antiox,xlab="Contribuição para a heterogeneidade total",
+        ylab="Influência no efeito síntese")
```

Como podemos observar na Figura 4.4, os estudos para a meta-análise da suplementação de antioxidantes que mais contribuem para

Figura 4.4: Gráficos Baujat para as meta-análises da suplementação de antioxidantes (em cima) e taxa de reincidência associada a programas de reabilitação e reinserção social (em baixo)



a heterogeneidade total, ao nível de 5%, são os estudos 20 (Graat *et al.*, 2002), 23 (Waters *et al.*, 2002), 31 (Goodman *et al.*, 2004) e 32 (Hercberg *et al.*, 2004). Para a meta-análise da taxa de reincidência associada a programas de reabilitação e reinserção social existem no total 20 estudos que contribuem para a heterogeneidade total ao nível de 5%. Por outro lado, o estudo com maior influência no efeito síntese para a meta-análise da suplementação de antioxidantes é o estudo 31 (Goodman *et al.*, 2004), enquanto para a meta-análise da taxa de reincidência o estudo 15 (Collingwood *et al.*, 1976).

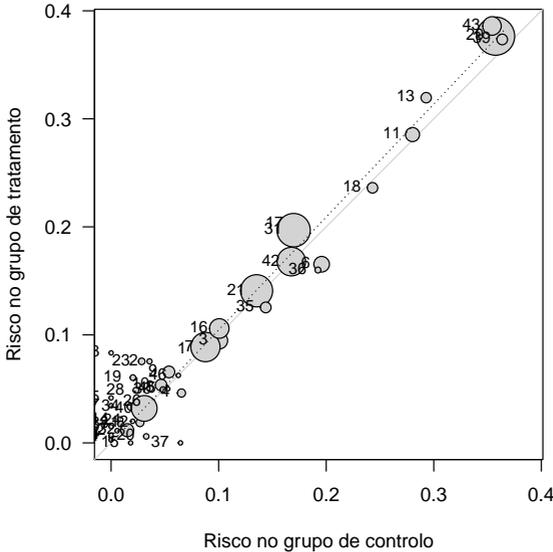
## 4.4 Gráfico de L'Abbé

Um gráfico de L'Abbé (L'Abbé *et al.*, 1987) é um gráfico de dispersão que permite mostrar os resultados de estudos que envolvem a comparação de dois grupos em que as respostas são binárias. Neste tipo de gráfico é representado no eixo horizontal a proporção  $\hat{p}_{1k}$  de sucessos observada num dos grupos (por exemplo, no grupo de controlo) e no eixo vertical a proporção  $\hat{p}_{2k}$  de sucessos observada no outro grupo (grupo de tratamento). Os pontos  $(\hat{p}_{1k}, \hat{p}_{2k})$ ,  $k = 1, \dots, K$ , são representados por círculos com diâmetros proporcionais aos pesos atribuídos aos estudos pelo modelo considerado.

São também desenhadas no gráfico duas retas, uma delas é a bissetriz dos quadrantes ímpares, ou seja, a reta de equação  $y = x$  (linha a cheio), e uma outra que tem em conta a estimativa obtida para o efeito síntese de acordo com o modelo considerado (linha a tracejado), ou seja, a reta de equação  $\theta = \hat{\theta}$  para um MEF (ou  $\mu = \hat{\mu}$  para um MEA). Por exemplo, se o tamanho do efeito for o risco relativo, então esta segunda reta terá equação  $p_2/p_1 = \hat{\theta}$ , ou seja,  $y = \hat{\theta}x$ .

Consequentemente, círculos sobre a bissetriz dizem respeito a estudos cujas proporções de sucesso observadas são iguais nos dois grupos, enquanto círculos a acima desta reta a estudos onde a proporção de sucesso observada é superior no grupo de tratamento.

Figura 4.5: Gráfico de L'Abbé para a meta-análise da suplementação de antioxidantes para um MEA



Para obtermos o gráfico de L'Abbé para a meta-análise da suplementação de antioxidantes na Figura 4.5 foi usada a função *labbe* e o comando

```
> labbe(meta.antiox,comb.fixed=F,studlab = T,xlab="Risco no grupo de
+       controle",ylab="Risco no grupo de tratamento")
```

Como podemos observar Figura 4.5, há um elevado número de estudos que estimam que o risco de mortalidade é superior no grupo de tratamento (são no total 30 estudos nestas condições). Uma desvantagem deste tipo de representação gráfica é que os erros associados às estimativas não são considerados.

## Capítulo 5

# Outros Métodos em Meta-Análise

*When reporting results as a random effect model, researchers can assume they have controlled for heterogeneity between studies but have not explained why the studies vary.*

I. Allen e J. Seaman (2017)

### 5.1 Análise de Subgrupos

As características específicas dos estudos selecionados para uma meta-análise podem propiciar a formação de grupos e, por conseguinte, o interesse em averiguar se existem diferenças entre estes quanto ao efeito síntese. Numa Meta-Análise a comparação dos tamanhos do efeito entre diferentes grupos de estudos pode ser efetuada com uma análise de subgrupos, que basicamente é uma espécie de análise de variância aplicada aos estudos.

Para realizarmos uma análise de subgrupos devemos assumir um dos seguintes modelos:

**Modelo 1:** um MEF dentro de cada subgrupo;

**Modelo 2:** um MEA dentro de cada subgrupo com  $\tau^2$  diferente entre os subgrupos;

**Modelo 3:** um MEA dentro de cada subgrupo com  $\tau^2$  comum entre os subgrupos.

Após a escolha de um dos modelos anteriores, a comparação dos efeitos síntese entre os subgrupos pode ser feita, por exemplo, recorrendo a um teste  $Q$  para a homogeneidade. Neste contexto, cada subgrupo funciona como se fosse um único estudo, pelo que primeiro é realizada uma meta-análise em separado para cada subgrupo de estudos, sendo depois combinadas as estimativas dos efeitos síntese num efeito síntese global.

Para o que se segue, suporemos que existem  $G$  subgrupos e que no subgrupo  $g$  existem  $K_g$  estudos, com  $K_1 + \dots + K_G = K$ .

### Modelo 1

Este modelo é apropriado quando se pode admitir dentro de cada subgrupo de estudos um MEF (e também entre os subgrupos). Assim, se para o subgrupo  $g$  o verdadeiro tamanho do efeito for  $\theta_g$ , assume-se o modelo

$$\hat{\theta}_{gk} = \theta_g + \epsilon_{gk},$$

$g = 1, \dots, G$  e  $k = 1, \dots, K_g$ , em que  $\epsilon_{gk} \sim \mathcal{N}(0, \sigma_{gk}^2)$  e  $\text{Cov}(\epsilon_{gk}, \epsilon_{gj}) = 0$ ,  $k \neq j$ ,  $k, j = 1, \dots, K_g$ . Consequentemente, um estimador para  $\theta_g$  é

$$\hat{\theta}_g = \frac{\sum_{k=1}^{K_g} \omega_{gk} \hat{\theta}_{gk}}{\sum_{k=1}^{K_g} \omega_{gk}}, \quad (5.1)$$

onde  $\omega_{gk} = 1/\sigma_{gk}^2$ ,  $g = 1, \dots, G$  e  $k = 1, \dots, K_g$ . Neste caso tem-se que

$$\text{Var}(\widehat{\theta}_g) = \sigma_{\widehat{\theta}_g}^2 = \left( \sum_{k=1}^{K_g} \omega_{gk} \right)^{-1}, \quad g = 1, \dots, G.$$

Se assumirmos novamente um MEF, as estimativas  $\widehat{\theta}_1, \dots, \widehat{\theta}_G$  são combinadas da forma

$$\widehat{\theta}_{\text{Global}} = \frac{\sum_{g=1}^G \omega_g \widehat{\theta}_g}{\sum_{g=1}^G \omega_g},$$

com  $\omega_g = 1/\sigma_{\widehat{\theta}_g}^2 = \sum_{k=1}^{K_g} \omega_{gk}$ ,  $g = 1, \dots, G$ .

Assim, para testarmos a hipótese de não existir diferenças entre os efeitos síntese dos  $G$  subgrupos, testamos a hipótese de homogeneidade

$$H_0 : \theta_1 = \dots = \theta_G, \quad (5.2)$$

usando para tal a estatística  $Q$  de Cochran, que neste caso define-se da forma

$$Q_{M1} = \sum_{g=1}^G \omega_g \left( \widehat{\theta}_g - \widehat{\theta}_{\text{Global}} \right)^2. \quad (5.3)$$

Sob a validade da hipótese (5.2),  $Q_{M1} \sim \chi_{G-1}^2$ .

*Observação 5.1.* Note-se que a variabilidade total dada pela estatística  $Q$  de Cochran pode ser decomposta numa soma de quadrados ponderados, nomeadamente

$$\begin{aligned} Q &= \sum_{g=1}^G \sum_{k=1}^{K_g} \omega_{gk} \left( \widehat{\theta}_{gk} - \widehat{\theta}_{\text{Global}} \right)^2 \\ &= \sum_{g=1}^G \omega_g \left( \widehat{\theta}_g - \widehat{\theta}_{\text{Global}} \right)^2 + \sum_{g=1}^G \sum_{k=1}^{K_g} \omega_{gk} \left( \widehat{\theta}_{gk} - \widehat{\theta}_g \right)^2 \\ &= \sum_{g=1}^G \omega_g \left( \widehat{\theta}_g - \widehat{\theta}_{\text{Global}} \right)^2 + \sum_{g=1}^G Q_g \\ &= Q_{\text{Entre}} + Q_{\text{Dentro}}, \end{aligned}$$

onde

$$Q_g = \sum_{k=1}^{K_g} \omega_{gk} \left( \hat{\theta}_{gk} - \hat{\theta}_g \right)^2, \quad (5.4)$$

$g = 1, \dots, G$ , e  $Q_{\text{Entre}} = Q_{M1}$ .

Quanto às somas de quadrados ponderados  $Q_{\text{Entre}}$  e  $Q_{\text{Dentro}}$ , a primeira mede a variabilidade entre as estimativas dos efeitos síntese dos subgrupos e a última a variabilidade dos tamanhos do efeito dentro dos  $G$  subgrupos. Consequentemente, sob a hipótese de homogeneidade (5.2),  $Q_{\text{Entre}} \sim \chi_{G-1}^2$ , enquanto sob as hipóteses de homogeneidade  $H_{0g} : \theta_{g1} = \dots = \theta_{gK_g} = \theta_g$ ,  $g = 1, \dots, G$ ,  $Q_{\text{Dentro}} \sim \chi_{K-G}^2$ .

## Modelo 2

Este modelo é adequado quando faz sentido assumir dentro de cada subgrupo um MEA e o parâmetro de heterogeneidade  $\tau^2$  não toma o mesmo valor em todos os subgrupos. Para o subgrupo  $g$  admite-se que  $(\theta_{g1}, \dots, \theta_{gK_g})$  é uma amostra aleatória de uma população de tamanhos de efeito com valor médio  $\mu_g$  e variância  $\tau_g^2$ ,  $g = 1, \dots, G$ . O modelo para o subgrupo  $g$  é o MEA

$$\hat{\theta}_{gk} = \mu_g + \xi_{gk} + \epsilon_{gk}, \quad (5.5)$$

$g = 1, \dots, G$  e  $k = 1, \dots, K_g$ , onde  $\epsilon_{gk} \sim \mathcal{N}(0, \sigma_{gk}^2)$ ,  $\xi_{gk} \sim \mathcal{N}(0, \tau_g^2)$ ,  $\text{Cov}(\epsilon_{gk}, \epsilon_{gj}) = 0$ ,  $\text{Cov}(\xi_{gk}, \xi_{gj}) = 0$ ,  $k \neq j$ ,  $k, j = 1, \dots, K_g$ , e  $\text{Cov}(\epsilon_{gk}, \xi_{gj}) = 0$ ,  $k, j = 1, \dots, K_g$ .

Assim, um estimador para o valor médio da população do subgrupo  $g$  é

$$\hat{\mu}_g = \frac{\sum_{k=1}^{K_g} \omega_{gk}^* \hat{\theta}_{gk}}{\sum_{k=1}^{K_g} \omega_{gk}^*}, \quad (5.6)$$

em que  $\omega_{gk}^* = 1/(\sigma_{gk}^2 + \tau_g^2)$ ,  $g = 1, \dots, G$ ,  $k = 1, \dots, K_g$ . Por conseguinte,

$$\text{Var}(\hat{\mu}_g) = \sigma_{\hat{\mu}_g}^2 = \left( \sum_{k=1}^{K_g} \omega_{gk}^* \right)^{-1}. \quad (5.7)$$

Considerando novamente um MEA, a estimativa do efeito síntese global é dada por

$$\hat{\mu}_{\text{Global}} = \frac{\sum_{g=1}^G \omega_g^* \hat{\mu}_g}{\sum_{g=1}^G \omega_g^*}, \quad (5.8)$$

onde  $\omega_g^* = 1/\sigma_{\hat{\mu}_g}^2 = \sum_{k=1}^{K_g} \omega_{gk}^*$ ,  $g = 1, \dots, G$ .

A hipótese de interesse sob o Modelo 2 é a de não haver diferenças entre os tamanhos médios do efeito, ou seja,

$$H_0 : \mu_1 = \dots = \mu_G, \quad (5.9)$$

podendo esta hipótese também ser testada com a estatística  $Q$  de Cochran, que neste caso assume a expressão

$$Q_{M2} = \sum_{g=1}^G \omega_g^* (\hat{\mu}_g - \hat{\mu}_{\text{Global}})^2. \quad (5.10)$$

Sob a validade da hipótese (5.9),  $Q_{M2} \sim \chi_{G-1}^2$ . Repare-se que como  $Q_{M2}$  está a medir a variabilidade entre as estimativas dos efeitos síntese dos subgrupos,  $Q_{M2} = Q_{\text{Entre}}^*$ .

As fórmulas anteriores pressupõem à partida que as variâncias  $\tau_g^2$  (e  $\sigma_{gk}^2$ ),  $g = 1, \dots, G$ , são conhecidas, mas como este cenário é pouco provável, estas devem ser estimadas usando a informação do respetivo subgrupo. Por exemplo, se se decidir usar o estimador de DerSimonian-Laird, então

$$\hat{\tau}_{g;\text{DL}}^2 = \frac{Q_g - (K_g - 1)}{c_g},$$

onde  $Q_g$  está definido em (5.4) e

$$c_g = \sum_{k=1}^{K_g} \omega_{gk} - \frac{\sum_{k=1}^{K_g} \omega_{gk}^2}{\sum_{k=1}^{K_g} \omega_{gk}}, \quad g = 1, \dots, G. \quad (5.11)$$

### Modelo 3

Este modelo difere do Modelo 2 apenas pelo facto de assumir que o parâmetro de heterogeneidade  $\tau^2$  tem o mesmo valor nos diferentes subgrupos de estudos. Assim, para o subgrupo  $g$  também se considera o modelo (5.5), com a diferença de  $\xi_{gk} \sim \mathcal{N}(0, \tau^2)$ ,  $g = 1, \dots, G$ ,  $k = 1, \dots, K_g$ .

Neste caso o estimador  $\hat{\mu}_g$  também é dado pela expressão (5.6), em que  $\omega_{gk}^* = 1/(\sigma_{gk}^2 + \tau^2)$ ,  $g = 1, \dots, G$  e  $k = 1, \dots, K_g$ . Quanto à variância de  $\hat{\mu}_g$ , esta é igualmente dada pela fórmula (5.7). Consequentemente, sob um MEA, o efeito síntese global pode do mesmo modo ser estimado a partir de (5.8).

Por outro lado, a hipótese de não haver diferenças entre os tamanhos médios do efeito (5.9) é testada usando a estatística  $Q$  de Cochran (5.10), com  $\omega_g^* = 1/\sigma_{\hat{\mu}}^2 = \sum_{k=1}^{K_g} \omega_{gk}^* = \sum_{k=1}^{K_g} 1/(\sigma_{gk}^2 + \tau^2)$ ,  $g = 1, \dots, G$ .

Para o caso de ser necessário estimar  $\tau^2$ , se for usado, por exemplo, o estimador de DerSimonian-Laird, então

$$\hat{\tau}_{\text{DL,comb}}^2 = \frac{\sum_{g=1}^G c_g \hat{\tau}_{g;\text{DL}}^2}{\sum_{g=1}^G c_g} = \frac{\sum_{g=1}^G (Q_g - (K_g - 1))}{\sum_{g=1}^G c_g} = \frac{\sum_{g=1}^G Q_g - (K - G)}{\sum_{g=1}^G c_g},$$

com  $Q_g$  definido em (5.4) e  $c_g$  em (5.11).

Para o exemplo da suplementação de antioxidantes, os estudos foram classificados pelos autores em função do tipo de prevenção pretendido com o tratamento, nomeadamente em prevenção primária ou prevenção secundária.<sup>1</sup> Neste caso, podemos querer averiguar se o tipo de prevenção tem um efeito sobre o risco relativo. Atendendo às características dos estudos dentro dos dois subgrupos (prevenção primária e prevenção secundária), podemos assumir *a priori* o Modelo 2 para a análise dos subgrupos.

---

<sup>1</sup>Os estudos classificados como prevenção primária foram aqueles em que os participantes eram indivíduos saudáveis, enquanto os classificados como prevenção secundária os restantes em que os participantes eram indivíduos que apresentavam uma determinada patologia.

```
> meta.antiox.M2=metabin(event.e=ai,n.e=n1i,event.c=ci,n.c=n2i,
+                         data=bjelakovic,sm="RR",method="Inverse",
+                         method.tau="REML",byvar=Prevencao,
+                         studlab=paste(Estudo,Ano,sep=","))
> summary(meta.antiox.M2)
```

Number of studies combined: k = 47

	RR	95%-CI	z	p-value
Fixed effect model	1.055	[1.030; 1.080]	4.48	< 0.0001
Random effects model	1.046	[1.007; 1.086]	2.33	0.0199

Quantifying heterogeneity:

tau<sup>2</sup> = 0.0016; H = 1.04 [1.00; 1.23]; I<sup>2</sup> = 6.9% [0.0%; 34.3%];  
Rb = 9.3% [0.0%; 34.5%]

Test of heterogeneity:

Q	d.f.	p-value
49.41	46	0.3387

Results for subgroups (fixed effect model):

	k	RR	95%-CI	Q	tau <sup>2</sup>	I <sup>2</sup>
Prevencao = S	33	1.030	[0.986; 1.076]	23.36	0	0.0%
Prevencao = P	14	1.065	[1.036; 1.095]	24.47	0.0037	46.9%
		Rb				
Prevencao = S		0.0%				
Prevencao = P		22.7%				

Test for subgroup differences (fixed effect model):

	Q	d.f.	p-value
Between groups	1.58	1	0.2094
Within groups	47.83	45	0.3585

Results for subgroups (random effects model):

	k	RR	95%-CI	Q	tau <sup>2</sup>	I <sup>2</sup>
Prevencao = S	33	1.030	[0.986; 1.076]	23.36	0	0.0%
Prevencao = P	14	1.049	[0.981; 1.122]	24.47	0.0037	46.9%
		Rb				
Prevencao = S		0.0%				
Prevencao = P		22.7%				

Test for subgroup differences (random effects model):

	Q	d.f.	p-value
Between groups	0.20	1	0.6587

Details on meta-analytical method:

- Inverse variance method
- Restricted maximum-likelihood estimator for  $\tau^2$
- Continuity correction of 0.5 in studies with zero cell frequencies

Embora as estimativas pontuais sob o Modelo 2 sugiram que em cada subgrupo o risco relativo médio seja superior a um, com uma estimativa ligeiramente maior para o subgrupo da prevenção primária, não podemos descartar para nenhum subgrupo a hipótese do risco relativo médio ser igual a um (ao nível de 5%). Por outro lado, não há evidência contra a hipótese dos riscos relativos médios ser iguais nos dois subgrupos, pois  $p_{\text{val}} = 0.659$  ( $Q_{\text{Entre}}^* = Q_{\text{M2}} = 0.20$ ).

Note-se que sob o Modelo 1 também concluiríamos que não há diferenças entre os dois subgrupos ( $Q_{\text{Entre}} = Q_{\text{M1}} = 1.58$  e  $p_{\text{val}} = 0.209$ ), embora a evidência a favor de tal seja menor que a sob o Modelo 2. No entanto, quando se considera as meta-análises dos subgrupos em separado, vemos que apenas no da prevenção primária há evidência a favor do risco relativo ser superior a 1 (IC a 95%: [1.036,1.095]).

Caso tivéssemos assumido o Modelo 3, teríamos obtido os seguintes resultados:

```
> meta.antiox.M3=metabin(event.e=ai,n.e=n1i,event.c=ci,n.c=n2i,
+                          data=bjelakovic,sm="RR",method="Inverse",
+                          method.tau="REML",byvar=Prevencao,
+                          studlab=paste(Estudo,Ano,sep=","),
+                          tau.common=T)
> summary(meta.antiox.M3)
```

Number of studies combined: k = 47

	RR	95%-CI	z	p-value
Fixed effect model	1.055	[1.030; 1.080]	4.48	< 0.0001
Random effects model	1.046	[1.007; 1.086]	2.33	0.0199

Quantifying heterogeneity:

$\tau^2 = 0.0016$ ;  $H = 1.04$  [1.00; 1.23];  $I^2 = 6.9\%$  [0.0%; 34.3%];  
 $R_b = 9.3\%$  [0.0%; 34.5%]

Test of heterogeneity:

Q	d.f.	p-value
49.41	46	0.3387

Results for subgroups (fixed effect model):

	k	RR	95%-CI	Q	$\tau^2$	$I^2$
Prevencao = S	33	1.030	[0.986; 1.076]	23.36	0.0015	0.0%
Prevencao = P	14	1.065	[1.036; 1.095]	24.47	0.0015	46.9%
		$R_b$				
Prevencao = S		5.8%				
Prevencao = P		16.6%				

Test for subgroup differences (fixed effect model):

	Q	d.f.	p-value
Between groups	1.58	1	0.2094
Within groups	47.83	45	0.3585

Results for subgroups (random effects model):

	k	RR	95%-CI	Q	$\tau^2$	$I^2$
Prevencao = S	33	1.031	[0.976; 1.089]	23.36	0.0015	0.0%
Prevencao = P	14	1.058	[1.007; 1.113]	24.47	0.0015	46.9%
		$R_b$				
Prevencao = S		5.8%				
Prevencao = P		16.6%				

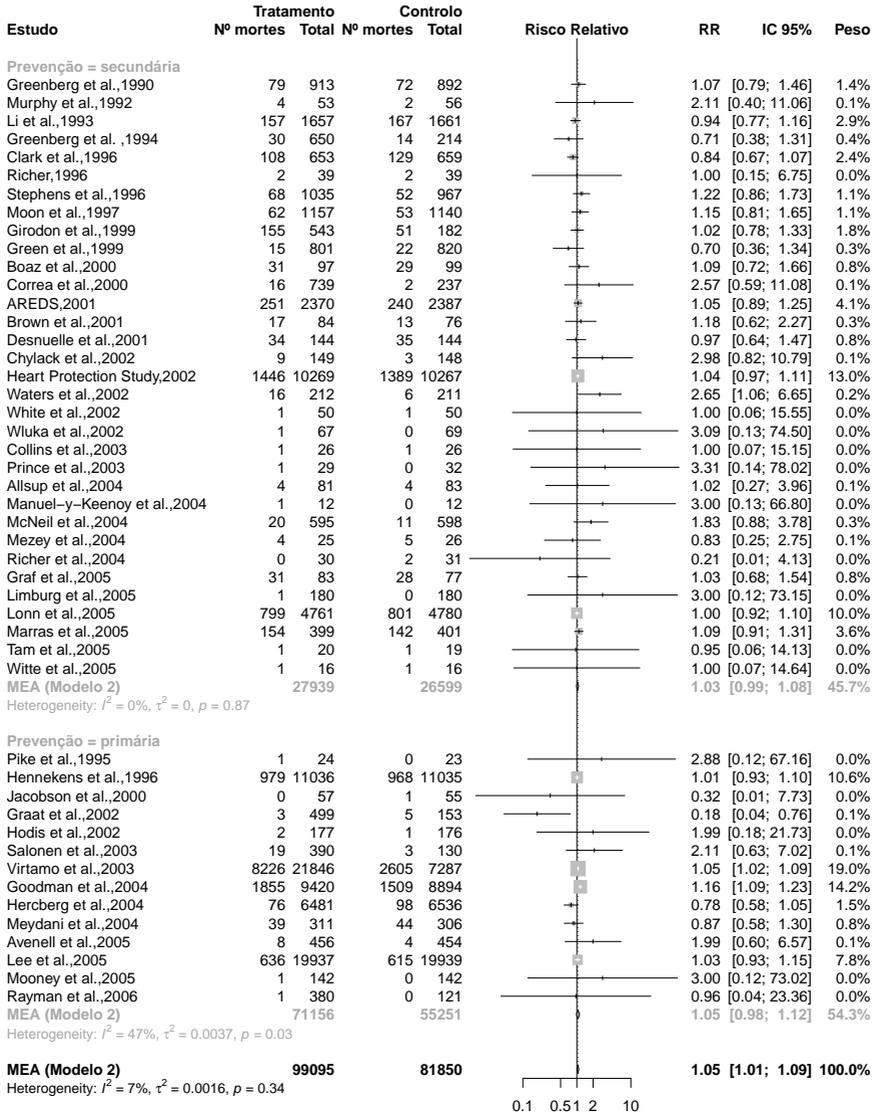
Test for subgroup differences (random effects model):

	Q	d.f.	p-value
Between groups	0.47	1	0.4912
Within groups	47.83	45	0.3585

Details on meta-analytical method:

- Inverse variance method

Figura 5.1: Gráfico floresta para a análise de subgrupos para a meta-análise da suplementação de antioxidantes sob o Modelo 2



Como podemos constatar, mesmo assumindo um  $\tau^2$  comum entre os dois subgrupos ( $\hat{\tau}^2 = 0.0015$ ), também existe alguma evidência a favor da hipótese dos riscos relativos médios ser iguais, pois  $p_{\text{val}} = 0.491$  ( $Q_{\text{Entre}}^* = Q_{\text{M3}} = 0.47$ ), embora a variabilidade associada às estimativas dos efeitos síntese seja ligeiramente superior. Por outro lado, o facto de  $Q_{\text{Dentro}} = 47.83$  e  $p_{\text{val}} = 0.359$ , revela não haver evidência contra as hipóteses de homogeneidade dos tamanhos do efeito dentro dos dois subgrupos. Por outro lado, para a meta-análise da taxa de reincidência associada a programas de reabilitação e reinserção social, se usarmos o tipo de programa para definir os subgrupos, obtemos os seguintes resultados sob o Modelo 2:

```
> meta.programas.M2=metagen(TE=phi,seTE=sqrt(1/n),data=whitehead,
+                             method.tau="REML",byvar=Programa,
+                             studlab=paste(Estudo,Ano,sep=","))
> summary(meta.programas.M2)
```

Number of studies combined: k = 45

		95%-CI	z	p-value
Fixed effect model	0.102	[ 0.085; 0.120]	11.55	< 0.0001
Random effects model	0.040	[-0.027; 0.108]	1.16	0.2446

Quantifying heterogeneity:

tau<sup>2</sup> = 0.0437; H = 2.99 [2.66; 3.35]; I<sup>2</sup> = 88.8% [85.9%; 91.1%];  
Rb = 81.6% [63.9%; 99.3%]

Test of heterogeneity:

Q	d.f.	p-value
392.21	44	< 0.0001

Results for subgroups (fixed effect model):

	k		95%-CI	Q	tau <sup>2</sup>
Programa = NSDP	13	0.050	[ 0.013; 0.086]	49.00	0.0253
Programa = SDP	13	0.162	[ 0.138; 0.187]	152.91	0.0352
Programa = PPCC	10	0.049	[ 0.002; 0.095]	106.11	0.0812
Programa = IR	7	0.040	[-0.014; 0.094]	31.37	0.0538
Programa = SP	2	-0.082	[-0.194; 0.030]	1.01	< 0.0001
		I <sup>2</sup>	Rb		
Programa = NSDP		75.5%	73.4%		

Programa = SDP 92.2% 81.6%  
 Programa = PPCC 91.5% 89.4%  
 Programa = IR 80.9% 77.2%  
 Programa = SP 1.0% 1.1%

Test for subgroup differences (fixed effect model):

	Q	d.f.	p-value
Between groups	51.81	4	< 0.0001
Within groups	340.41	40	< 0.0001

Results for subgroups (random effects model):

	k		95%-CI	Q	tau <sup>2</sup>
Programa = NSDP	13	0.032	[-0.069; 0.133]	49.00	0.0253
Programa = SDP	13	0.125	[ 0.012; 0.238]	152.91	0.0352
Programa = PPCC	10	0.029	[-0.158; 0.216]	106.11	0.0812
Programa = IR	7	-0.067	[-0.262; 0.129]	31.37	0.0538
Programa = SP	2	-0.082	[-0.195; 0.030]	1.01	< 0.0001

	I <sup>2</sup>	Rb
Programa = NSDP	75.5%	73.4%
Programa = SDP	92.2%	81.6%
Programa = PPCC	91.5%	89.4%
Programa = IR	80.9%	77.2%
Programa = SP	1.0%	1.1%

Test for subgroup differences (random effects model):

	Q	d.f.	p-value
Between groups	7.33	4	0.1195

Details on meta-analytical method:

- Inverse variance method
- Restricted maximum-likelihood estimator for tau<sup>2</sup>

Analisando separadamente a meta-análise para cada tipo de programa sob o Modelo 2, constatamos que apenas o programa SDP (programas ligados ao sistema judicial) parece ter um impacto positivo na prevenção da reincidência, pois o IC a 95% para o  $\phi$  médio é [0.012,0.238] (associação positiva fraca). Contudo, a percentagem de variabilidade total explicada pela heterogeneidade neste subgrupo é bastante elevada ( $I^2 = 92.2\%$ ). No entanto, ao nível de 5%, há alguma evidência a favor da hipótese dos  $\phi$ 's médios ser iguais nos 4 subgrupos

( $Q_{\text{Entre}}^* = Q_{M2} = 7.33$  e  $p_{\text{val}} = 0.120$ ).

Para uma análise de subgrupos sob o Modelo 3 podemos definir uma espécie de coeficiente de determinação  $R^2$  que tem em conta a variância  $\tau^2$ . Neste contexto define-se  $R^2$  da forma

$$R^2 = \frac{\text{Variância Explicada}}{\text{Variância Total}} = 1 - \frac{\text{Variância Não Explicada}}{\text{Variância Total}}, \quad (5.12)$$

onde a variância não explicada é o valor de  $\tau^2$  comum aos subgrupos e a variância total o valor de  $\tau^2$  quando não se consideram subgrupos.

Por exemplo, para a análise de subgrupos para a meta-análise da suplementação sob o Modelo 3, obtemos

$$R^2 = 1 - \frac{\hat{\tau}_{\text{comum}}^2}{\hat{\tau}^2} = 1 - \frac{0.001516}{0.001605} = 0.0555,$$

ou seja, a heterogeneidade comum aos dois subgrupos de prevenção apenas explica 5.55% da heterogeneidade total.

*Observação 5.2.* Quando numa análise de subgrupos se pretender comparar os tamanhos do efeito de dois subgrupos (1 e 2), é possível quantificar a magnitude da diferença observada, ou seja, estimar o parâmetro  $\Delta = \theta_1 - \theta_2$  sob o Modelo 1 (ou  $\Delta = \mu_1 - \mu_2$  sob os Modelos 2 e 3) através de  $\hat{\Delta} = \hat{\theta}_1 - \hat{\theta}_2$  ( $\hat{\Delta} = \hat{\mu}_1 - \hat{\mu}_2$ ). Neste caso, como

$$\text{Var}(\hat{\Delta}) = \text{Var}(\hat{\theta}_1) + \text{Var}(\hat{\theta}_2)$$

( $\text{Var}(\hat{\Delta}) = \text{Var}(\hat{\mu}_1) + \text{Var}(\hat{\mu}_2)$ ), os limites de confiança a  $(1 - \alpha)100\%$  para  $\Delta$  são  $\hat{\Delta} \mp z_{1-\alpha/2} \sqrt{\text{Var}(\hat{\Delta})}$ .

## 5.2 Meta-Regressão

Qualquer análise de subgrupos em Meta-Análise pode ser feita através de uma meta-regressão, à semelhança de uma análise de variância que pode ser resolvida como um problema de regressão. Todavia, uma

meta-regressão não é um procedimento que sirva apenas para comparar subgrupos em Meta-Análise, permite estabelecer relações entre o tamanho do efeito e outras variáveis, internas ou externas aos estudos, chamadas variáveis moderadoras.<sup>2</sup>

Vejam, em primeiro lugar, como podemos usar uma meta-regressão para efetuarmos uma análise de subgrupos.

Sob um MEF, o modelo de meta-regressão para a comparação de  $G$  subgrupos é o modelo linear

$$\hat{\theta}_k = \theta + \beta_1 x_{1k} + \cdots + \beta_{G-1} x_{G-1k} + \epsilon_k, \quad (5.13)$$

em que  $x_{gk} = 1$  se o estudo  $k$  pertencer ao subgrupo  $g$  e  $x_{gk} = 0$  caso contrário, com  $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$ ,  $k = 1, \dots, K$ , e  $\text{Cov}(\epsilon_k, \epsilon_j) = 0$ ,  $k \neq j$ ,  $k, j = 1, \dots, K$ ,  $g = 1, \dots, G$ .

Como no modelo (5.13) o subgrupo  $G$  está a ser considerado o subgrupo de referência, o parâmetro  $\theta$  representa o tamanho do efeito no subgrupo  $G$ , enquanto  $\theta + \beta_g = \theta_g$  o tamanho do efeito no subgrupo  $g$ ,  $g = 1, \dots, G - 1$ . Logo, todo o conjunto de técnicas conhecidas para uma análise de regressão pode ser aplicado ao modelo (5.13). Neste contexto, a hipótese nula de não haver diferença entre os efeitos síntese dos subgrupos equivale a testar a hipótese

$$H_0 : \beta_1 = \dots = \beta_{G-1} = 0. \quad (5.14)$$

Tal como acontece numa análise de regressão vulgar, a variabilidade total ( $Q_T$ ) também pode ser decomposta numa soma de duas parcelas independentes, ou seja,

$$Q_T = Q_M + Q_E, \quad (5.15)$$

onde  $Q_M = Q_{M1}$  mede a variabilidade que é explicada pelo modelo (5.13) e  $Q_E = Q_{\text{Dentro}}$  mede a variabilidade que não é explicada

---

<sup>2</sup>Uma meta-regressão não deve ser realizada quando o número de estudos é pequeno.

pelo modelo, sendo aqui  $Q_T$  a estatística  $Q$  de Cochran definida em (3.2).

Sob a validade da hipótese (5.14),  $Q_M \sim \chi_{G-1}^2$ , enquanto sob a validade da hipótese de homogeneidade dentro de cada subgrupo,  $Q_E \sim \chi_{K-G}^2$ . Por sua vez, testes  $z$  podem ser realizados para testar a significância de cada parâmetro  $\beta_j$  do modelo (5.13).<sup>3</sup>

Por outro lado, quando se assume uma população de tamanhos de efeito, uma meta-regressão para uma análise de subgrupos assenta neste caso no modelo linear

$$\hat{\theta}_k = \mu + \beta_1 x_{1k} + \cdots + \beta_{G-1} x_{G-1k} + \xi_k + \epsilon_k, \quad (5.16)$$

onde  $x_{gk}$  está definido tal como no caso anterior,  $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$ ,  $\xi_k \sim \mathcal{N}(0, \tau^2)$ ,  $k = 1, \dots, K$ , e  $\text{Cov}(\epsilon_k, \epsilon_j) = 0$ ,  $\text{Cov}(\xi_k, \xi_j) = 0$ ,  $k \neq j$ ,  $\text{Cov}(\epsilon_k, \xi_j) = 0$ ,  $k, j = 1, \dots, K$ ,  $g = 1, \dots, G$ .<sup>4</sup>

Para o modelo (5.16)  $\mu$  representa o efeito síntese para o subgrupo de referência  $G$ , enquanto  $\mu + \beta_g = \mu_g$  o efeito síntese para o subgrupo  $g$ ,  $g = 1, \dots, G - 1$ . Como o modelo possui termos associados a efeitos fixos (representados pelos  $\beta_j$ 's) e um termo associado a efeitos aleatórios (representado pelo  $\xi_k$ ) é considerado um modelo de efeitos mistos.

Da mesma forma, a hipótese (5.14) de não haver diferença entre os efeitos síntese dos subgrupos também será testada tendo em conta uma partição do tipo (5.15). Neste caso,  $Q_M = Q_{M3}$ , sendo que sob a validade da hipótese nula,  $Q_M \sim \chi_{G-1}^2$ . Por outro lado, a percentagem da variabilidade total que é explicada por um modelo de meta-regressão é dada por  $R^2$ , valor que se definiu em (5.12).

Voltando ao exemplo da suplementação de antioxidantes, comparemos os riscos relativos médios dos dois subgrupos de prevenção através de uma meta-regressão, assumindo para o efeito o modelo de efeitos

<sup>3</sup>Não são realizados testes  $t$  como acontece numa regressão linear vulgar porque se está a aplicar testes a estudos e não a observações.

<sup>4</sup>Note-se que no modelo (5.16) se está a exigir que o valor de  $\tau^2$  seja igual nos diferentes subgrupos, à semelhança do pressuposto de homocedasticidade numa análise de variância.

mistos

$$\widehat{\theta} = \mu + \beta x + \xi + \epsilon, \quad (5.17)$$

onde  $\widehat{\theta} = \ln \widehat{RR}$  e  $x$  é a variável indicatriz do subgrupo, tomando o valor 0 se o estudo pertencer ao subgrupo da prevenção primária e o valor 1 caso contrário. Para realizarmos a meta-regressão usaremos a função *metareg* do pacote *meta*.<sup>5</sup>

```
> library(metafor)
> metareg(meta.antiox, ~Prevencao)
```

Mixed-Effects Model (k = 47; tau<sup>2</sup> estimator: REML)

```
tau^2 (estimated amount of residual heterogeneity):      0.002 (SE = 0.002)
tau (square root of estimated tau^2 value):             0.039
I^2 (residual heterogeneity / unaccounted variability): 12.86%
H^2 (unaccounted variability / sampling variability):    1.15
R^2 (amount of heterogeneity accounted for):            5.55%
```

Test for Residual Heterogeneity:

QE(df = 45) = 47.831, p-val = 0.359

Test of Moderators (coefficient(s) 2):

QM(df = 1) = 0.474, p-val = 0.491

Model Results:

	estimate	se	zval	pval	ci.lb	ci.ub
intrcpt	0.057	0.026	2.219	0.026	0.007	0.107
PrevencaoS	-0.026	0.038	-0.688	0.491	-0.101	0.048

O modelo estimado é

$$\ln \widehat{RR} = 0.057 - 0.026 x.$$

Assim, o logaritmo da estimativa do risco relativo médio no subgrupo da prevenção primária é 0.057 ( $\widehat{RR}_P = 1.059$ ), observando-se uma diminuição média de 0.026 unidades no logaritmo do risco relativo médio no subgrupo da prevenção secundária. Porém, esta diminuição não é

<sup>5</sup>O uso da função *metareg* requer que o pacote *metafor* esteja instalado.

estatisticamente significativa dado não existir evidência contra a hipótese de  $\beta = 0$  ( $p_{\text{val}} = 0.491$ ). Repare-se que o valor- $p$  aqui associado é, como já estávamos à espera, igual ao valor- $p$  para a hipótese do modelo não explicar as diferenças observadas entre os subgrupos ( $Q_M = 0.474$  e  $p_{\text{val}} = 0.491$ ). Consequentemente, não há evidência a favor do logaritmo do risco relativo médio estar relacionado com o tipo de prevenção.

Também não existe evidência contra a hipótese de haver homogeneidade dentro de cada subgrupo dado que  $Q_E = 47.831$  e  $p_{\text{val}} = 0.359$ . Observe-se que estes valores são iguais aos valores obtidos com a análise de subgrupos sob o Modelo 3 na Secção 5.1, e para a qual se obteve o valor  $R^2 = 5.55\%$ . Por outro lado, a hipótese do efeito síntese no subgrupo da prevenção primária ser nulo ( $H_0 : \mu = 0$ ), isto é, do risco relativo médio ser 1, é rejeitada ao nível de significância de 5% ( $p_{\text{val}} = 0.026$ ).

Os modelos (5.13) e (5.16) também podem ser usados quando, em vez de termos  $G - 1$  variáveis binárias (mudas) para representar os  $G$  níveis de uma variável moderadora categórica, temos  $G - 1$  variáveis moderadoras contínuas, ou mistas.

Para exemplificarmos uma meta-regressão com uma variável moderadora contínua, consideremos novamente o exemplo da suplementação de antioxidantes. Averiguemos se existe uma relação entre o risco relativo e a duração do tratamento. O modelo a considerar é um modelo do tipo (5.17), em que  $x$  representa agora a duração do tratamento em anos ( $\mu$  não representa agora um valor médio).

```
> metareg.antiox.duracao=metareg(meta.antiox, ~Duracao)
> metareg.antiox.duracao
```

```
Mixed-Effects Model (k = 47; tau^2 estimator: REML)
```

```
tau^2 (estimated amount of residual heterogeneity):      0.001 (SE = 0.002)
tau (square root of estimated tau^2 value):              0.037
I^2 (residual heterogeneity / unaccounted variability): 12.65%
H^2 (unaccounted variability / sampling variability):    1.14
R^2 (amount of heterogeneity accounted for):              16.52%
```

Test for Residual Heterogeneity:

QE(df = 45) = 46.574, p-val = 0.407

Test of Moderators (coefficient(s) 2):

QM(df = 1) = 1.198, p-val = 0.274

Model Results:

	estimate	se	zval	pval	ci.lb	ci.ub
intrcpt	0.088	0.043	2.054	0.040	0.004	0.172
Duracao	-0.007	0.007	-1.095	0.274	-0.020	0.006

O modelo estimado é

$$\ln \widehat{RR} = 0.088 - 0.007 x .$$

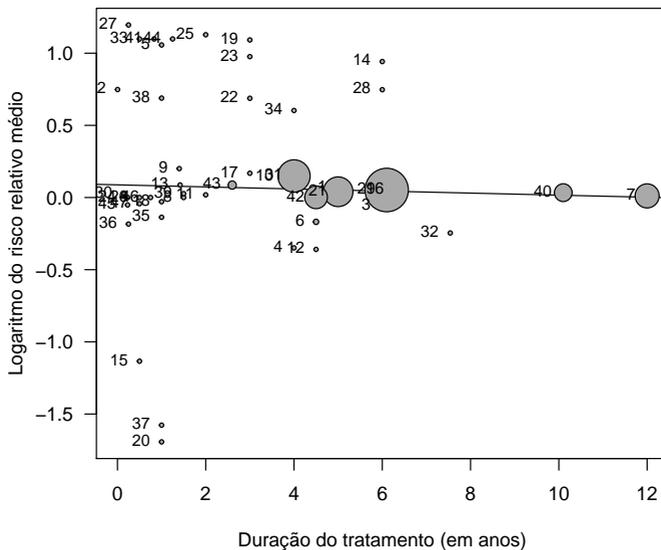
Consequentemente, estima-se que com o prolongamento do tratamento por mais um ano haja uma diminuição média de 0.007 unidades no logaritmo do risco relativo. No entanto, há alguma evidência a favor da hipótese de  $\beta = 0$  ( $p_{\text{val}} = 0.274$ ), ou seja, do logaritmo do risco relativo não estar relacionado com a duração do tratamento.

Os resultados de uma meta-regressão com uma variável moderadora contínua podem ser representados num gráfico de bolhas. As bolhas neste tipo de gráfico são na realidade círculos que representam as estimativas dos tamanhos do efeito observados, com diâmetros proporcionais aos pesos atribuídos aos estudos.

Na Figura 5.2 representamos o gráfico de bolhas para a meta-regressão da meta-análise da suplementação de antioxidantes, quando se considera para variável moderadora a duração do tratamento. Para o efeito foi usado a função *bubble* e o comando

```
> bubble(metareg.antiox.duracao,studlab=T,xlab="Duração do tratamento
+       (em anos)",ylab="Logaritmo do risco relativo médio")
```

Figura 5.2: Gráfico de bolhas para a meta-regressão da suplementação de antioxidantes



### 5.3 Meta-Análise Cumulativa

Uma meta-análise cumulativa não é um procedimento novo, trata-se de uma meta-análise que é realizada primeiro com um estudo, depois com dois estudos, de seguida com três, e assim sucessivamente, até que todos os estudos selecionados sejam incluídos na meta-análise. Geralmente, a ordem de entrada dos estudos na meta-análise é determinada pelas datas de publicação dos estudos, ou seja, por ordem cronológica de publicação, mas poderá ser utilizado qualquer outro critério. Por exemplo, quando os estudos são incluídos por ordem cronológica de publicação, uma meta-análise cumulativa mostra como a evidência se foi acumulada ao longo do tempo e se as conclusões se alteraram nesse período, e em que sentido.

As conclusões de uma meta-análise cumulativa podem ser bastante úteis, pois podem: i) desaconselhar a realização de mais estudos similares, uma vez que a questão que se procurava responder já foi adequadamente respondida com uma meta-análise; ii) direcionar recursos financeiros para a realização de outros estudos, de melhor qualidade, quando a questão de interesse ainda não foi convenientemente respondida.

Vejamos um exemplo de como as conclusões de uma meta-análise cumulativa poderiam ter desaconselhado a realização de mais experiências semelhantes para responder a um problema de interesse clínico.

O uso de sangue e seus derivados no perioperatório é cada vez mais um assunto menos pacífico por causa do aumento da morbidade e mortalidade, dos custos associados às transfusões, da recusa de alguns doentes receber sangue por questões religiosas, etc. Por isso, ao longo dos anos têm sido realizados estudos com o objetivo de avaliar a eficácia de alguns fármacos no sentido de reduzir a terapêutica de sangue. Um desses fármacos é a aprotinina, um inibidor de serina proteases, que tem sido usada há muito tempo em cirurgias cardíacas onde o risco hemorrágico é bastante elevado.

Em Fergusson *et al.* (2005) consideraram-se 64 estudos controlados e completamente aleatórios, com datas de publicação entre 1987 e 2002, que procuraram comparar a eficácia da aprotinina *versus* placebo em diminuir a necessidade de transfusões sanguíneas em cirurgias cardíacas. Para cada estudo foi registado por grupo (tratamento e placebo) o número de transfusões efetuadas e o número total de cirurgias realizadas.<sup>6</sup>

Usando, tal como os autores, o *odds ratio* para tamanho do efeito, o estimador de DerSimonian-Laird e o método de combinação de Mantel-Haenszel, obtemos os seguintes resultados sob um MEA:

```
> meta.aprot=metabin(event.e=ai,n.e=n1i,event.c=ci,n.c=n2i,
+ data=aprotinina,sm="OR",comb.fixed=F)
```

---

<sup>6</sup>Os dados estão disponíveis em Fergusson *et al.* (2005).

```
> summary(meta.aprot)
```

```
Number of studies combined: k = 63
```

```

                OR          95%-CI      z  p-value
Random effects model 0.342 [0.286; 0.410] -11.67 < 0.0001
```

```
Quantifying heterogeneity:
```

```
tau^2 = 0.1771; H = 1.37 [1.18; 1.58]; I^2 = 46.4% [27.8%; 60.1%];
Rb = 33.3% [14.8%; 51.9%]
```

```
Test of heterogeneity:
```

```

      Q d.f.  p-value
115.58  62 < 0.0001
```

```
Details on meta-analytical method:
```

- Mantel-Haenszel method
- DerSimonian-Laird estimator for tau^2
- Continuity correction of 0.5 in studies with zero cell frequencies
- Studies with double zeros included in meta-analysis

Os resultados da meta-análise revelam que a estimativa do *odds ratio* combinado médio é 0.342, com um IC a 95% igual a [0.286, 0.410]. Assim, estima-se que o uso da aprotinina reduza as hipóteses a favor da necessidade de uma transfusão de sangue em 65.8%, em média, quando comparado com o placebo. A redução verificada poderá na realidade variar entre 59 a 71.4% em média, ou seja, há claramente evidência a favor da eficácia da aprotinina.

Na Figura 5.3 representamos o gráfico floresta para a meta-análise cumulativa para o exemplo da aprotinina, em que os estudos entraram por ordem cronológica. Para realizarmos esta meta-análise cumulativa foi usada a função *metacum* e o comando

```
> metacum(meta.aprot, sortvar = Ano, pooled="random")
```

Como podemos observar, desde o primeiro estudo publicado há evidência a favor da eficácia da aprotinina. Também se observa que a estimativa do *odds ratio* médio estabiliza bastante cedo em torno do

valor 0.3, sensivelmente a partir da entrada do décimo segundo estudo (Dietrich, 1992).

A respeito da eficácia da aprotinina, Fergusson *et al.* (2005) afirmam nas conclusões que:

*This study demonstrates that investigators evaluating aprotinin were not adequately citing previous research, resulting in a large number of RCTs being conducted to address efficacy questions that prior trials had already definitively answered. Institutional review boards and journals could reduce the number of redundant trials by requiring investigators to conduct adequate searches for prior evidence and conducting systematic reviews.*

Na Figura 5.4, onde temos o gráfico floresta para a meta-análise cumulativa para o exemplo da suplementação de antioxidantes, constatamos que para a entrada dos primeiros vinte e oito estudos, publicados entre 1990 e 2003, a evidência acumulada aponta no sentido do risco de mortalidade ser igual nos dois grupos. Mas, a partir do estudo Virtamo *et al.* (2003) dá-se uma mudança, passa a haver uma evidência acumulada no sentido de um maior risco de mortalidade com a suplementação de antioxidantes, com o valor do risco relativo médio a estabilizar em torno de 1.05.

Figura 5.3: Gráfico floresta para a meta-análise cumulativa para o exemplo da aptrotinina

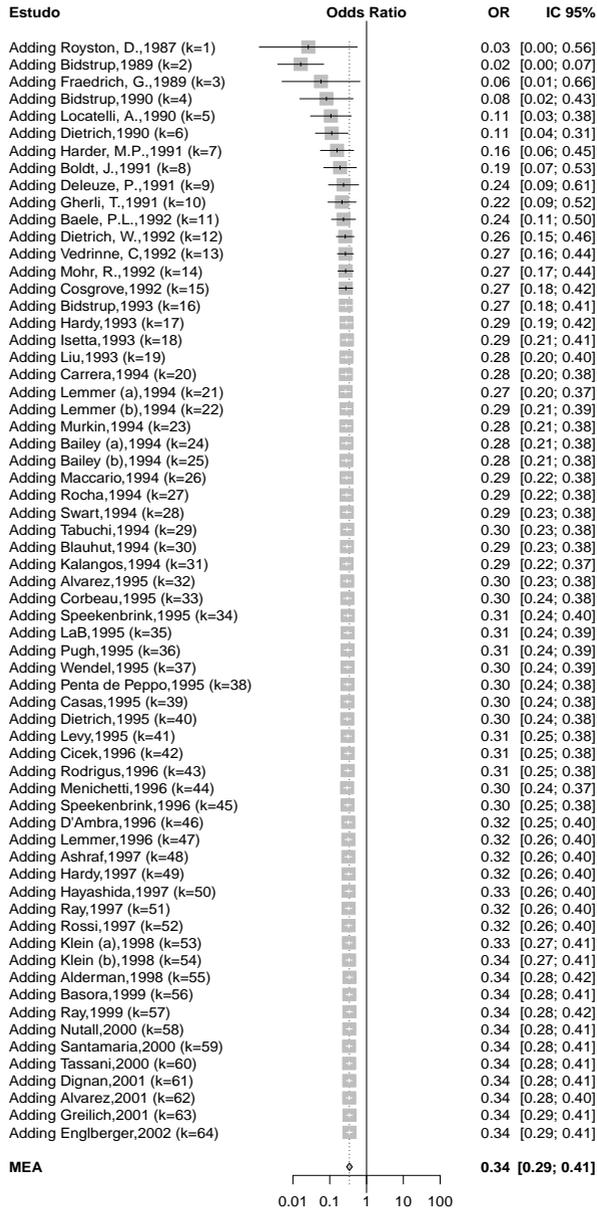
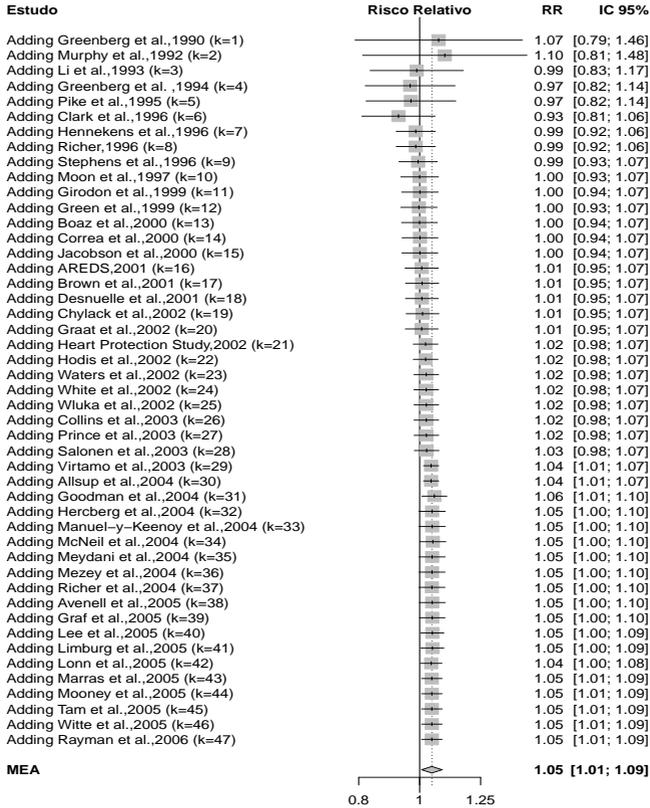


Figura 5.4: Gráfico floresta para a meta-análise cumulativa para o exemplo da suplementação de antioxidantes



## Capítulo 6

# Viés de Publicação

*The combined effects found by a meta analysis often appear to overstate the evidence and to be biased in favor of the alternative.*

Kulinskaya *et al.* (2008)

A convicção generalizada de que “rejeitar uma hipótese nula é uma decisão forte, enquanto mantê-la uma decisão fraca”, tem de certa forma contribuído para que alguns estudos relevantes não sejam publicados por “falharem a desejada” propriedade de significância estatística, isto é,  $p_{\text{val}} < 0.05$ . Sobre este assunto, vários autores têm alertado para o facto de estudos que reportam tamanhos do efeito elevados ter uma maior probabilidade de ser publicados, do que estudos que reportam tamanhos do efeito pequenos, porque são considerados “mais interessantes” pelos editores ou revisores de algumas revistas.

No entanto, as conclusões de uma meta-análise só poderão ser consideradas fiáveis se esta se basear no conjunto de estudos relevantes sobre um mesmo problema, alguns dos quais com resultados estatisticamente significativos e outros não. Mas, na maioria das vezes, a seleção dos estudos é feita exclusivamente a partir do conjunto de estudos publicados, pelo que esta situação pode enviesar as conclusões de

uma meta-análise. De facto, não se pode descartar a possibilidade de haver no meio do conjunto de estudos não publicados alguns que sejam tão ou mais informativos do que aqueles que foram publicados. Este tipo de viés de seleção é vulgarmente conhecido pelo *viés de publicação*, embora ultimamente esteja a ser mais conhecido pelo problema dos efeitos dos estudos pequenos, uma vez que são os estudos pequenos que reportam resultados “positivos” que normalmente causam problemas numa meta-análise quando selecionados.

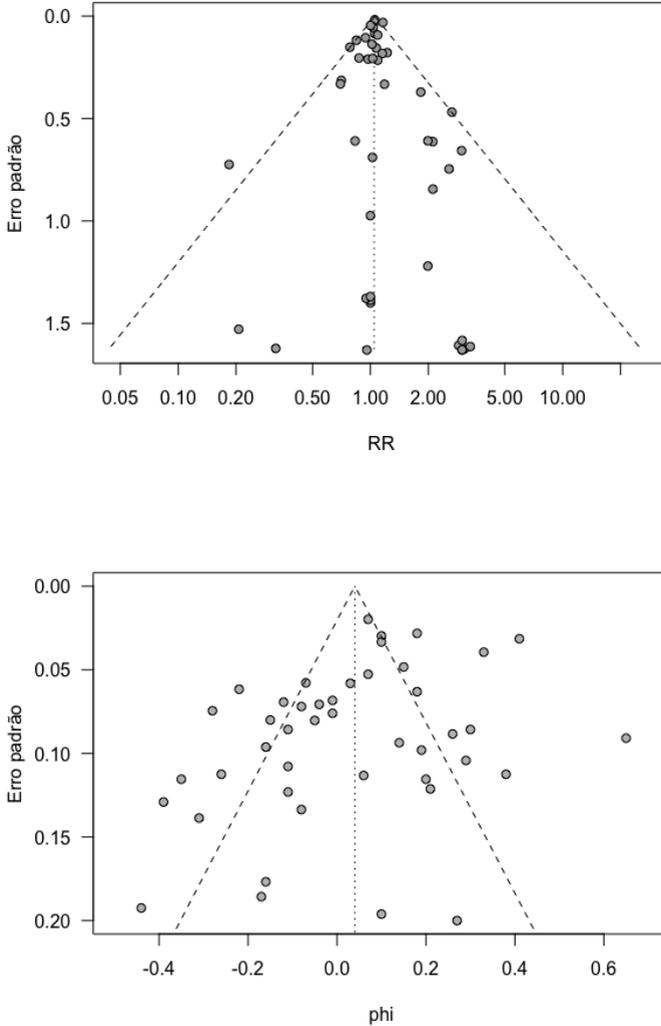
A situação ideal é que todos os estudos relevantes sobre o mesmo problema, publicados ou não, sejam incluídos na meta-análise, mas em muitos casos este objetivo é inatingível. Contudo, existem alguns métodos que nos podem ajudar a identificar a presença de viés de publicação e a procurar corrigi-lo.

## 6.1 Gráfico Funil

Uma ferramenta gráfica bastante popular para detetar o viés de publicação, e que não assenta em nenhum modelo em particular, é um tipo de gráfico de dispersão conhecido por gráfico funil, proposto por Light e Pillemer (1984), onde se representa a precisão da estimativa do tamanho do efeito *versus* o tamanho do efeito observado no estudo. Embora Light e Pillemer (1984) tenham sugerido inicialmente como medida de precisão a dimensão da amostra do estudo, chegou-se depois à conclusão que a medida mais eficaz a considerar seria o erro padrão da estimativa do tamanho do efeito.

Representando então no eixo horizontal o tamanho do efeito observado e no eixo vertical o erro padrão da estimativa, não haverá indícios de viés de publicação se a nuvem de pontos for simétrica e estiver contida num funil, com o bico deste a apontar para cima (o funil está invertido). Na ausência de viés de publicação os estudos que produzem estimativas menos precisas estarão espalhados na base do funil, enquanto os estudos que originam estimativas mais precisas concentrados no topo, em torno da estimativa do efeito síntese.

Figura 6.1: Gráficos funil para as meta-análises da suplementação de antioxidantes (em cima) e taxa de reincidência associada a programas de reabilitação e reinserção social (em baixo)



A falta de simetria na nuvem de pontos sugere a falta de estudos e, por conseguinte, a possibilidade de haver viés de publicação.

Num gráfico funil também se pode representar um segmento vertical correspondendo ao valor do efeito síntese estimado pelo modelo, podendo igualmente se considerar uma pseudo-região de confiança limitada pelos valores “efeito síntese  $\mp 1.96 \times$  erro padrão”, onde os valores para o erro padrão são obtidos a partir da amplitude dos valores registados para o eixo vertical com base nos estudos selecionados.

Na Figura 6.1 representamos os gráficos funil para as meta-análises da suplementação de antioxidantes e da taxa de reincidência associada a programas de reabilitação e reinserção social. Para obtermos gráficos funil no R usamos a função *funnel*. Por exemplo, o primeiro gráfico da Figura 6.1 foi obtido usando o comando

```
> funnel(meta.antiox,xlab="RR",ylab="Erro padrão",comb.fixed=F)
```

Observe-se que no gráfico funil para a meta-análise da suplementação existe uma ligeira assimetria na nuvem de pontos, parecendo faltar estudos que indiquem um menor risco de mortalidade com a suplementação (lado esquerdo do gráfico). Por outro lado, no gráfico funil para a meta-análise da taxa de reincidência, também parece existir uma assimetria na nuvem de pontos, aparentemente faltando estudos com conclusões a favor da eficácia dos programas (lado direito do gráfico). Repare-se que, ao contrário do que acontece com o primeiro exemplo, não existem estimativas concentradas no topo do gráfico em torno do efeito síntese. É também visível uma maior heterogeneidade entre as estimativas dos tamanhos do efeito neste segundo exemplo. No entanto, é difícil afirmarmos com total convicção que existe viés de publicação para cada caso simplesmente a partir do respetivo gráfico funil.

Aliás, o uso de um gráfico funil como ferramenta para detetar o viés de publicação deve ser feito com alguma cautela, uma vez que vários autores têm alertado para a inadequação de um gráfico funil dado o carácter subjetivo da sua interpretação. Por outro lado, alguns autores também têm chamado a atenção para o facto de uma aparente assime-

tria no gráfico funil poder-se dever a outros fatores que não apenas a um viés de seleção, como por exemplo: i) o idioma de publicação do estudo, pois de um modo geral a escolha limita-se a estudos publicados na língua inglesa; ii) a relutância de alguns investigadores publicar os seus resultados, sobretudo quando não é verificada a significância estatística, ou quando o efeito observado é no sentido contrário ao esperado/desejado; iii) o atraso na publicação dos estudos pelas revistas, principalmente quando os artigos aceites para publicação descrevem estudos onde se conclui que o tamanho do efeito é nulo; iv) a intervenção do acaso.

## 6.2 Testes para o Viés de Publicação

A dificuldade de detetar a existência de viés de publicação a partir de um gráfico funil, tem levado alguns autores a proporem testes estatísticos que procuram quantificar a assimetria observada no gráfico. Por exemplo, Begg e Mazumdar (1994) propõem um teste de correlação ordinal ajustado baseado no coeficiente *tau* de Kendall. Este teste tem em conta que se existir assimetria no gráfico funil, em virtude de estudos não estatisticamente significativos ter menor probabilidade de ser publicados, haverá uma correlação entre as estimativas dos tamanhos do efeito ( $\hat{\theta}_k$ ) e as suas variâncias ( $\sigma_k^2 = \text{Var}(\hat{\theta}_k)$ ).

O coeficiente de correlação ordinal é calculado com base nas estimativas dos tamanhos do efeito standardizadas, ou seja,

$$\hat{\theta}_k^* = \frac{\hat{\theta}_k - \hat{\theta}}{\sigma_k^*},$$

com  $\hat{\theta}$  o efeito síntese definido em (3.2), e nas variâncias

$$\sigma_k^{*2} = \text{Var}(\hat{\theta}_k - \hat{\theta}) = \sigma_k^2 - \left( \sum_{j=1}^K \omega_j \right)^{-1} = \sigma_k^2 - \left( \sum_{j=1}^K \sigma_j^{-2} \right)^{-1},$$

$k = 1, \dots, K$ .

Denotando por  $X$  o número de pares de estudos  $(k, j)$  concordantes, ou seja, que verificam

$$\widehat{\theta}_k^* < \widehat{\theta}_j^* \quad \text{e} \quad \sigma_k^{*2} < \sigma_j^{*2},$$

ou

$$\widehat{\theta}_k^* > \widehat{\theta}_j^* \quad \text{e} \quad \sigma_k^{*2} > \sigma_j^{*2},$$

e  $Y$  o número de pares de estudos  $(k, j)$  discordantes, isto é, que satisfazem

$$\widehat{\theta}_k^* < \widehat{\theta}_j^* \quad \text{e} \quad \sigma_k^{*2} > \sigma_j^{*2},$$

ou

$$\widehat{\theta}_k^* > \widehat{\theta}_j^* \quad \text{e} \quad \sigma_k^{*2} < \sigma_j^{*2},$$

a estatística de teste a usar é

$$Z = \frac{X - Y}{\sqrt{K(K-1)(2K+5)/18}}.$$

Sob a hipótese nula de não haver associação, ou seja, do gráfico funil ser simétrico,  $Z \stackrel{a}{\sim} \mathcal{N}(0, 1)$ , pelo que a hipótese nula será rejeitada ao nível de significância  $\alpha$  se  $|Z_{\text{obs}}| > z_{1-\alpha/2}$ .<sup>1</sup> No entanto, o teste de Begg não deve ser aplicado quando o número de estudos numa meta-análise é pequeno, porque este teste revelou possuir uma fraca potência estatística nestas situações.

Egger *et al.* (1997) propõem em alternativa um teste de regressão linear simples para identificar o viés de publicação, através da assimetria observada num gráfico funil. Este teste procura descrever o *score* normal  $z_k = \widehat{\theta}_k / \sqrt{\text{Var}(\widehat{\theta}_k)}$  em função da precisão da estimativa (preditor), geralmente o inverso do erro padrão da estimativa, ou seja, considera o modelo linear

$$z_k = \alpha + \beta / \sqrt{\text{Var}(\widehat{\theta}_k)}, \quad k = 1, \dots, K.$$

---

<sup>1</sup>No caso de haver empates nas observações  $\widehat{\theta}_k^*$  ou  $\sigma_k^{*2}$  há que proceder à correção da variância de  $X - Y$ .

A lógica subjacente ao teste de Egger é a seguinte: As estimativas dos estudos mais pequenos têm erros padrão maiores, originando por isso  $1/\sqrt{\text{Var}(\hat{\theta}_k)}$  e *scores*  $z_k$  próximos de zero, e mais perto da origem. Por outro lado, estudos maiores produzem estimativas mais precisas, de modo que se existir um tamanho do efeito não nulo, originarão *scores* maiores. Assim, se não houver viés de publicação, os pontos estarão dispostos em torno de uma reta que passará na origem ( $\alpha = 0$ ), cujo declive  $\beta$  indicará o valor do efeito síntese. Caso haja assimetria no gráfico, indicador de viés de publicação, a reta de regressão deixa de passar na origem e o valor de  $\alpha$  quantifica a assimetria observada no gráfico funil. Basicamente, o teste de Egger testa a significância do parâmetro de regressão  $\alpha$ . Se a hipótese nula  $H_0 : \alpha = 0$  não for rejeitada, assume-se que há simetria no gráfico funil, ou seja, que não há viés de publicação.

*Observação 6.1.* A aplicação do teste de Egger é desaconselhada se no conjunto dos estudos meta-analisados não existir pelo menos um estudo grande que possa servir de termo de comparação.

Várias generalizações do teste de Egger têm sido propostas como, por exemplo, o teste de Thompson e Sharp (1999) que se baseia numa regressão linear ponderada. Por outro lado, como geralmente os testes concebidos para dados contínuos tendem a ser liberais quando aplicados a dados binários, alguns autores propuseram modificações do teste de Egger para respostas binárias como, por exemplo, o teste *score* de Harbord (Harbord *et al.*, 2006), o teste de Peters (Peters *et al.*, 2006) e o teste de Rücker para arcosenos (Rücker *et al.*, 2008). Também existe o teste de Schwarzer (Schwarzer *et al.*, 2007) que é um teste de correlação ordinal alternativo ao teste de Begg. Todos estes testes estão disponíveis na função *metabias* do pacote *meta*.

Aplicando os testes anteriores à meta-análise da suplementação de antioxidantes, obtemos os seguintes resultados:

```
> # teste de Begg
> metabias(meta.antiox,method="rank")
```

Rank correlation test of funnel plot asymmetry

```
data: meta.antiox
z = 1.8066, p-value = 0.07083
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      ks      se.ks
197.0000 109.0459
```

```
> # teste de Egger
> metabias(meta.antiox,method="linreg")
```

Linear regression test of funnel plot asymmetry

```
data: meta.antiox
t = 0.32314, df = 45, p-value = 0.7481
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias      slope
0.05827323 0.18033443 0.05081768
```

```
> # teste de Thompson
> metabias(meta.antiox,method="mm")
```

Linear regression test of funnel plot asymmetry  
(methods of moment)

```
data: meta.antiox
t = 0.52151, df = 45, p-value = 0.6046
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias      slope
0.09397330 0.18019351 0.04212344
```

```
> # teste de Harbord
> metabias(meta.antiox,method="score")
```

Linear regression test of funnel plot asymmetry  
(efficient score)

```

data: meta.antiox
t = 0.4576, df = 45, p-value = 0.6494
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias      slope
0.09177639 0.20056195 0.06270306

> # teste de Schwarzer
> metabias(meta.antiox,method="count")

      Rank correlation test of funnel plot asymmetry
      (based on counts)

data: meta.antiox
z = 1.798, p-value = 0.07218
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      ks      se.ks
196.0000 109.0123

> # teste de Peters
> metabias(meta.antiox,method="peters")

      Linear regression test of funnel plot asymmetry
      (based on sample size)

data: meta.antiox
t = 0.30468, df = 45, p-value = 0.762
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias      slope
4.91691055 16.13770408 0.05159794

```

Todos os testes aplicados sugerem não haver evidência contra a hipótese do gráfico funil ser simétrico e, por conseguinte, não haver evidência a favor da existência de viés de publicação. Observe-se que os testes que apresentam menores valores-*p* são os testes de correlação ordinal de Begg e de Schwarzer.

Por outro lado, se aplicarmos os testes de Begg, Egger e Thompson à meta-análise da taxa de reincidência associada a programas de

reabilitação e reinserção social, obtemos os seguintes resultados:

```
> # teste de Begg
> metabias(meta.programas,method="rank")

Rank correlation test of funnel plot asymmetry

data: meta.programas
z = -0.5381, p-value = 0.5905
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      ks      se.ks
-55.0000 102.2106

> # teste de Egger
> metabias(meta.programas,method="linreg")

Linear regression test of funnel plot asymmetry

data: meta.programas
t = -2.0674, df = 43, p-value = 0.04475
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias      slope
-1.6077209  0.7776530  0.1821435

> # teste de Thompson
> metabias(meta.programas,method="mm")

Linear regression test of funnel plot asymmetry
(methods of moment)

data: meta.programas
t = -1.3968, df = 43, p-value = 0.1696
alternative hypothesis: asymmetry in funnel plot
sample estimates:
      bias      se.bias      slope
-0.9293275  0.6653212  0.1223615
```

Logo, ao nível de 5%, apenas há evidência a favor de viés de publicação com o teste de Egger. Segundo este teste existe uma assimetria negativa no gráfico funil ( $\hat{\alpha} = -1.608$ ), ou seja, faltam estimativas no lado direito do gráfico.

## 6.3 Número de Estudos na Gaveta

Os testes anteriores procuram detetar o viés de publicação através da assimetria observada no gráfico funil, mas não estimam quantos estudos ficaram de fora da análise e que poderão eventualmente afetar as conclusões globais, caso sejam incluídos na meta-análise. Este problema foi designado por Rosenthal (1979) pelo problema dos estudos na gaveta (*file-drawer problem*).

Desconhecendo-se a realidade sobre o número de estudos na gaveta, vários autores têm procurado estimar esse número com o intuito de avaliar qual será o impacto que os estudos em falta poderão ter na meta-análise.

Rosenthal (1979) propôs a seguinte abordagem ao problema:

*The fundamental idea in coping with the file drawer problem is simply to calculate the number of studies averaging null results that must be in the file drawers before the overall probability of Type I error is brought to any desired level of significance, say  $p = 0.05$ .*

Suponhamos então que existem  $K + K_0$  estudos relevantes sobre um determinado problema, dos quais  $K$  são conhecidos (publicados) e  $K_0$  desconhecidos (não publicados, em que muitos destes estudos não apresentam resultados estatisticamente significativos). Se considerarmos o método de combinação de testes de Stouffer, método usado em Rosenthal (1979),<sup>2</sup> que recordemos tem estatística de teste

$$\frac{1}{\sqrt{K}} \sum_{k=1}^K \Phi^{-1}(1 - P_k) = \frac{1}{\sqrt{K}} \sum_{k=1}^K Z_k,$$

onde  $Z_k = \Phi^{-1}(1 - P_k) \sim_{|H_{0,k}} \mathcal{N}(0, 1)$ ,  $k = 1, \dots, K$ , então os  $K$  valores- $p$  reportados levarão à rejeição da hipótese nula global (3.11),

---

<sup>2</sup>Qualquer outro método de combinação pode ser usado, com as devidas adaptações.

ao nível de significância  $\alpha$ , se

$$\frac{1}{\sqrt{K}} \sum_{k=1}^K z_k > z_{1-\alpha},$$

com  $z_k = \Phi^{-1}(1 - p_k)$ ,  $k = 1, \dots, K$ . Para estimarmos o valor  $K_0$ , número de estudos na gaveta necessários para alterarmos a decisão para aceitação da hipótese nula global, admitamos que o efeito global combinado desses  $K_0$  valores- $p$  é nulo, ou seja,  $\sum_{k=K+1}^{K+K_0} z_k = 0$ . Assim, a decisão final de rejeitar a hipótese nula global será alterada se para os  $K + K_0$  estudos verificarmos

$$\frac{1}{\sqrt{K + K_0}} \sum_{k=1}^{K+K_0} z_k = \frac{1}{\sqrt{K + K_0}} \sum_{k=1}^K z_k \leq z_{1-\alpha}.$$

Consequentemente,  $K_0$  será o menor número inteiro que satisfaz a condição

$$K_0 \geq \frac{\left(\sum_{k=1}^K z_k\right)^2}{z_{1-\alpha}^2} - K.$$

Neste contexto, a existência de viés de publicação dependerá da plausibilidade do valor  $K_0$  em função do problema sob análise. Se este valor for considerado pequeno, é provável que existam estudos na gaveta que podem afetar as conclusões de uma meta-análise. Neste caso deve-se envidar todos os esforços no sentido de localizar mais estudos relevantes.

Note-se que a determinação do valor  $K_0$  pelo método proposto por Rosenthal suscita algumas questões, nomeadamente o facto de assumir que o efeito global médio dos  $K_0$  estudos em falta é nulo (esta condição pode no entanto ser relaxada) e assentar na combinação de estudos através de valores- $p$ , método pouco usado em Meta-Análise.

Orwin (1983) introduz uma estatística semelhante à do método de Rosenthal baseada na medida  $d$  de Cohen. Apesar da medida de Cohen ser uma diferença de médias standardizada, a ideia subjacente ao método de Orwin pode ser aplicada a qualquer outro tamanho do efeito.

Seja  $\hat{\theta}_k$  a estimativa do tamanho do efeito no estudo  $k$ ,  $k = 1, \dots, K + K_0$ , em que os valores  $\hat{\theta}_{K+1}, \dots, \hat{\theta}_{K+K_0}$  estão associados aos  $K_0$  estudos em falta. A estatística para o método de Orwin é

$$\bar{\theta} = \frac{1}{K + K_0} \sum_{k=1}^{K+K_0} \hat{\theta}_k = \frac{K \bar{\theta}_K + K_0 \bar{\theta}_{K_0}}{K + K_0},$$

onde  $\bar{\theta}_K$  e  $\bar{\theta}_{K_0}$  representam as médias das estimativas dos  $K$  estudos conhecidos e dos  $K_0$  estudos em falta, respetivamente. Consideremos

$$\frac{K \bar{\theta}_K + K_0 \bar{\theta}_{K_0}}{K + K_0} = \theta_m^*, \quad (6.1)$$

onde  $\theta_m^*$  é um valor fixado *a priori* e que representa o valor desejado para o mínimo do tamanho médio do efeito. Resolvendo (6.1) em ordem a  $K_0$ , obtém-se

$$K_0 = \frac{K (\bar{\theta}_K - \theta_m^*)}{\theta_m^* - \bar{\theta}_{K_0}}.$$

Se também assumirmos um efeito global médio nulo para os  $K_0$  estudos em falta, ou seja,  $\bar{\theta}_{K_0} = 0$ , o número de estudos na gaveta é dado por

$$K_0 = \frac{K (\bar{\theta}_K - \theta_m^*)}{\theta_m^*}.$$

Rosenberg (2005) chama a atenção para o facto dos métodos de Rosenthal e Orwin não ponderar os estudos, quando a prática habitual numa meta-análise é atribuir pesos a estes. Por isso, propõe que se utilize em alternativa a estatística

$$T = \frac{\hat{\theta}}{\sqrt{\text{Var}(\hat{\theta})}} = \frac{\sum_{k=1}^{K+K_0} \omega_k \hat{\theta}_k / \sum_{k=1}^{K+K_0} \omega_k}{\sqrt{\left(\sum_{k=1}^{K+K_0} \omega_k\right)^{-1}}},$$

com distribuição  $t$  de Student, ou equivalentemente,

$$T^2 = \frac{\left(\sum_{k=1}^{K+K_0} \omega_k \hat{\theta}_k\right)^2}{\sum_{k=1}^{K+K_0} \omega_k},$$

em que  $\omega_k = 1/\sigma_k^2$ ,  $k = 1, \dots, K + K_0$ . Assumindo novamente um efeito nulo para os estudos desconhecidos, devemos considerar

$$t_{\nu;1-\alpha}^2 = \frac{\left(\sum_{k=1}^K \omega_k \hat{\theta}_k\right)^2}{\sum_{k=1}^K \omega_k + \omega'}, \quad (6.2)$$

onde  $\omega' = \sum_{k=K+1}^{K+K_0} \omega_k$  representa o peso total a atribuir aos  $K_0$  estudos na gaveta para se atingir o nível de significância desejado e  $t_{\nu;\gamma}$  o quantil de probabilidade  $\gamma$  da distribuição  $t$  de Student com  $\nu$  graus de liberdade.

Resolvendo (6.2) em ordem a  $\omega'$  vem

$$\omega' = \frac{\left(\sum_{k=1}^K \omega_k \hat{\theta}_k\right)^2}{t_{\nu;1-\alpha}^2} - \sum_{k=1}^K \omega_k, \quad (6.3)$$

pelo que o número de estudos na gaveta é dado por

$$K_0 = \frac{\omega'}{\sum_{k=1}^K \omega_k / K}. \quad (6.4)$$

Note-se que na fórmula anterior  $\sum_{k=1}^K \omega_k / K$  representa o peso médio de um estudo conhecido, enquanto  $\omega'$  pode ser interpretado como sendo a dimensão relativa de um único estudo de efeito nulo necessário para reduzir a significância estatística para o nível  $\alpha$ .

No que respeita o número de graus de liberdade  $\nu = K + K_0 - 1$ , este pode ser obtido iterativamente até que o valor de  $K_0$  estabilize em (6.4). No entanto, como aproximação, podemos substituir em (6.3) o quantil da  $t$  de Student pelo quantil da normal  $z_{1-\alpha}$  para obtermos o valor de  $\omega'$ .

A função *fsn* (de *fail and safe number*) do pacote `metafor` permite-nos aplicar os três métodos anteriores. Por exemplo, para a meta-análise da taxa de reincidência associada a programas de reabilitação e reinserção social obtemos os seguintes resultados:

```
> fsn(yi=phi,vi=1/n,data=whitehead,type="Rosenthal")
```

Fail-safe N Calculation Using the Rosenthal Approach

Observed Significance Level: <.0001

Target Significance Level: 0.05

Fail-safe N: 624

```
> fsn(yi=phi,vi=1/n,data=whitehead,type="Orwin")
```

Fail-safe N Calculation Using the Orwin Approach

Average Effect Size: 0.0362

Target Effect Size: 0.0181

Fail-safe N: 45

```
> fsn(yi=phi,vi=1/n,data=whitehead,type="Rosenberg")
```

Fail-safe N Calculation Using the Rosenberg Approach

Average Effect Size: 0.1024

Observed Significance Level: <.0001

Target Significance Level: 0.05

Fail-safe N: 1517

Observamos uma grande discrepância entre as estimativas de  $K_0$  fornecidas pelos três métodos. O método de Orwin estima que há 45 estudos na gaveta e o método de Rosenberg 1517, enquanto pelo meio o método de Rosenthal indica 624 estudos em falta. Situações deste tipo podem provocar alguma indecisão quanto à existência ou não de viés de publicação. De facto, as estimativas indicadas pelos métodos descritos devem ser interpretadas com cautela. A propósito Rosenberg (2005) afirma que:

*Fail-safe numbers are not necessarily the best way to approach publication bias, but they are a simple first step that can help identify whether more complex approaches are necessary.*

## 6.4 Métodos para a Correção do Viés de Publicação

Métodos mais sofisticados foram propostos para a estimação do número de estudos na gaveta e também para a estimação dos tamanhos do efeito dos estudos em falta, com o objetivo de procurar corrigir o viés de publicação numa meta-análise.

Duval e Tweedie (2000a, 2000b) propõem um método simples e não paramétrico de aumento computacional, conhecido pelo método apara e preenche (*trim and fill method*), que assenta exclusivamente em argumentos de simetria do gráfico funil, argumentos válidos quer para um MEF, quer para um MEA. O argumento de simetria usado na justificação do método é que, na ausência de viés de publicação, uma estimativa do tamanho do efeito tanto pode estar à direita, como à esquerda do efeito síntese. Por isso, podemos considerar as ordens dos valores absolutos das estimativas, afetando-as de um sinal que indicará a sua posição em relação ao efeito síntese, ou seja, um sinal “+” se a estimativa estiver à direita do efeito síntese, ou um sinal “-” se estiver à esquerda, muito à semelhança do que se faz no teste de Wilcoxon. No caso de haver viés de publicação serão as estimativas mais extremas (ordens mais extremas) que estarão ausentes no gráfico funil.

Em termos gerais, o método apara e preenche tem duas fases distintas: uma primeira fase de aparagem de estudos (estimativas) do gráfico funil e uma segunda fase de preenchimento do gráfico. Para explicarmos o funcionamento do método, admitamos que faltam estudos no lado esquerdo do gráfico funil.

Na primeira fase são suprimidos estudos no lado do gráfico funil onde aparentemente não faltam estudos (lado direito), de acordo com a estimativa  $\widehat{K}_0^{(j)}$  obtida na  $j$ -ésima iteração. Por exemplo, o primeiro passo do algoritmo nesta fase consiste em calcular os valores  $\widehat{\theta}_k^{(1)} = \widehat{\theta}_k - \widehat{\theta}^{(1)}$ ,  $k = 1, \dots, K$ , onde  $\widehat{\theta}^{(1)}$  é o efeito síntese obtido a partir das estimativas  $\widehat{\theta}_k$ ,  $k = 1, \dots, K$ . Obter depois  $\widehat{K}_0^{(1)}$  a partir dos valores  $\widehat{\theta}_k^{(1)}$ ,  $k = 1, \dots, K$ . Usando o argumento de simetria, remover  $\widehat{K}_0^{(1)}$  estimati-

vas  $\hat{\theta}_k$  mais extremas do lado direito do gráfico funil. Com base agora nos valores  $\hat{\theta}_1, \dots, \hat{\theta}_{K-\hat{K}_0^{(1)}}$ , calcular o efeito síntese  $\hat{\theta}^{(2)}$ . De seguida, obter os valores  $\hat{\theta}_k^{(2)} = \hat{\theta}_k - \hat{\theta}^{(2)}$ ,  $k = 1, \dots, K$ . Calcular  $\hat{K}_0^{(2)}$  a partir dos valores  $\hat{\theta}_k^{(2)}$ ,  $k = 1, \dots, K$ , procedendo-se de seguida à remoção das  $\hat{K}_0^{(2)}$  estimativas  $\hat{\theta}_k$ ,  $k = 1, \dots, K - \hat{K}_0^{(1)}$  mais extremas no lado direito. Esta fase de aparagem termina quando se verificar  $\hat{K}_0^{(j)} = \hat{K}_0^{(j+1)} = \hat{K}_0$ .

Na segunda fase do método preenche-se o gráfico funil no lado onde faltam estudos (lado esquerdo) com os  $\hat{K}_0$  valores estimados  $\hat{\theta}_i^* = 2\hat{\theta}^{(j)} - \hat{\theta}_{K-i+1}$ , com erro padrão  $\sigma_i^* = \sigma_{K-i+1}$ ,  $i = 1, \dots, \hat{K}_0$ . A partir dos valores  $\hat{\theta}_1, \dots, \hat{\theta}_K, \hat{\theta}_1^*, \dots, \hat{\theta}_{\hat{K}_0}^*$  e  $\sigma_1, \dots, \sigma_K, \sigma_1^*, \dots, \sigma_{\hat{K}_0}^*$  pode-se obter estimativas para o efeito síntese ajustado e erro padrão usando os métodos usuais.

Para a estimação de  $K_0$ , Duval e Tweedie (2000a) propõem três estimadores. Um dos estimadores é  $R_0$ , que nos dá o comprimento do maior *run* de ordens positivas subtraído de uma unidade (se os estudos em falta estiverem do lado esquerdo do gráfico funil). Denotando por  $T_K^+$  a soma das ordens das estimativas dos tamanhos do efeito afetadas de sinal positivo, sob o que os autores chamam um modelo de Bernoulli suprimido,<sup>3</sup> são definidos os outros dois estimadores

$$L_0 = \frac{4T_K^+ - K(K+1)}{2K-1} \quad \text{e} \quad Q_0 = K - \frac{1}{2} - \sqrt{2K^2 - 4T_K^+ + \frac{1}{4}}.$$

Duval e Tweedie (2000b) compararam o desempenho dos três estimadores propostos, tendo concluído que apesar do estimador  $R_0$  ser o único não enviesado para  $K_0$ , não é robusto, ao passo que  $L_0$  e  $Q_0$  o são. Por outro lado, repararam que os estimadores  $L_0$  e  $Q_0$  têm comportamentos semelhantes, mas dado que  $Q_0$  tende a apresentar ocasionalmente valores mais elevados que  $L_0$ , este último deve ser geralmente preferido. Por exemplo, em relação ao estimador  $L_0$ , tem-se que

$$E(L_0) = K_0 - \frac{K_0^2}{2K-1} \quad \text{e} \quad \text{Var}(L_0) = \frac{16 \text{Var}(T_K^+)}{(2K-1)^2},$$

<sup>3</sup>Na prática  $T_K^+$  é o valor da estatística de Wilcoxon  $W_K^+$  para a amostra dos tamanhos do efeito estimados.

em que

$$\text{Var}(T_k^+) = 24^{-1} [K(K+1)(2K+1) + 10K_0^3 + 27K_0^2 + 17K_0 - 18K K_0(K_0+1) + 6K^2K_0] .$$

A função *trimfill* permite-nos aplicar o método apara e preenche a um objeto de meta-análise. Se aplicarmos o método à meta-análise da suplementação de antioxidantes, obtemos os seguintes resultados:

```
> trimfill(meta.antiox,comb.fixed=F)
```

	RR	95%-CI	%W(random)
Greenberg et al.,1990	1.072	[0.790; 1.455]	1.4
Murphy et al.,1992	2.113	[0.404; 11.062]	0.1
Li et al.,1993	0.942	[0.766; 1.159]	2.9
Greenberg et al.,1994	0.705	[0.381; 1.305]	0.4
Pike et al.,1995	2.878	[0.123; 67.163]	0.0
Clark et al.,1996	0.845	[0.670; 1.065]	2.4
Hennekens et al.,1996	1.011	[0.929; 1.101]	10.6
Richer,1996	1.000	[0.148; 6.747]	0.0
Stephens et al.,1996	1.222	[0.861; 1.734]	1.1
Moon et al.,1997	1.153	[0.806; 1.648]	1.1
Girodon et al.,1999	1.019	[0.779; 1.332]	1.8
Green et al.,1999	0.698	[0.365; 1.336]	0.3
Boaz et al.,2000	1.091	[0.716; 1.664]	0.8
Correa et al.,2000	2.566	[0.594; 11.077]	0.1
Jacobson et al.,2000	0.322	[0.013; 7.731]	0.0
AREDS,2001	1.053	[0.891; 1.245]	4.1
Brown et al.,2001	1.183	[0.616; 2.271]	0.3
Desnuelle et al.,2001	0.971	[0.644; 1.466]	0.8
Chylack et al.,2002	2.980	[0.823; 10.789]	0.1
Graat et al.,2002	0.184	[0.044; 0.761]	0.1
Heart Protection Study,2002	1.041	[0.972; 1.114]	13.0
Hodis et al.,2002	1.989	[0.182; 21.733]	0.0
Waters et al.,2002	2.654	[1.059; 6.651]	0.2
White et al.,2002	1.000	[0.064; 15.548]	0.0
Wluka et al.,2002	3.089	[0.128; 74.503]	0.0
Collins et al.,2003	1.000	[0.066; 15.150]	0.0
Prince et al.,2003	3.305	[0.140; 78.025]	0.0
Salonen et al.,2003	2.111	[0.635; 7.019]	0.1
Virtamo et al.,2003	1.053	[1.017; 1.091]	19.0

Allsup et al.,2004	1.025	[0.265; 3.959]	0.1
Goodman et al.,2004	1.161	[1.091; 1.234]	14.2
Herberg et al.,2004	0.782	[0.581; 1.053]	1.5
Manuel-y-Keenoy et al.,2004	3.000	[0.135; 66.801]	0.0
McNeil et al.,2004	1.827	[0.883; 3.780]	0.3
Meydani et al.,2004	0.872	[0.584; 1.303]	0.8
Mezey et al.,2004	0.832	[0.252; 2.748]	0.1
Richer et al.,2004	0.207	[0.010; 4.129]	0.0
Avenell et al.,2005	1.991	[0.604; 6.566]	0.1
Graf et al.,2005	1.027	[0.684; 1.542]	0.8
Lee et al.,2005	1.034	[0.927; 1.153]	7.8
Limburg et al.,2005	3.000	[0.123; 73.154]	0.0
Lonn et al.,2005	1.001	[0.916; 1.095]	10.0
Marras et al.,2005	1.090	[0.909; 1.306]	3.6
Mooney et al.,2005	3.000	[0.123; 73.024]	0.0
Tam et al.,2005	0.950	[0.064; 14.132]	0.0
Witte et al.,2005	1.000	[0.068; 14.640]	0.0
Rayman et al.,2006	0.958	[0.039; 23.362]	0.0
Filled: Mooney et al.,2005	0.371	[0.015; 9.024]	0.0
Filled: Wluka et al.,2002	0.360	[0.015; 8.685]	0.0
Filled: Prince et al.,2003	0.337	[0.014; 7.944]	0.0

Number of studies combined: k = 50 (with 3 added studies)

	RR	95%-CI	z	p-value
Random effects model	1.045	[1.007; 1.085]	2.30	0.0215

Quantifying heterogeneity:

tau<sup>2</sup> = 0.0016; H = 1.02 [1.00; 1.19]; I<sup>2</sup> = 3.5% [0.0%; 28.8%];  
Rb = 8.8% [0.0%; 42.4%]

Test of heterogeneity:

Q	d.f.	p-value
50.76	49	0.4041

Details on meta-analytical method:

- Inverse variance method
- Restricted maximum-likelihood estimator for tau<sup>2</sup>
- Trim-and-fill method to adjust for funnel plot asymmetry

O método apara e preenche estima que estão em falta 3 estudos (estimativas) no lado esquerdo do gráfico funil, tendo  $K_0$  sido estimado a partir de  $L_0$  (estimador por defeito na função *trimfill*). Repare-se que as estimativas em falta foram obtidas a partir dos 3 estudos conhecidos com as estimativas mais extremas no sentido “positivo”: Prince *et al.* (2003) com  $\widehat{RR} = 3.31$ , Wluka *et al.* (2002) com  $\widehat{RR} = 3.09$  e Mooney *et al.* (2005) com  $\widehat{RR} = 3.00$ . Se incluirmos na meta-análise as 3 estimativas em falta, e que espelham as três estimativas mais extremas conhecidas, obtemos para a estimativa do efeito síntese corrigido  $\widehat{\mu}_{cor} = 1.045$ , com um IC a 95% igual a [1.007, 1.085]. Estes resultados são bastante semelhantes aos resultados da meta-análise com apenas os 47 estudos ( $\widehat{\mu} = 1.046$ ), pelo que não parece haver evidência a favor de viés de publicação, ou melhor, o impacto deste, caso exista, será provavelmente insignificante na meta-análise realizada.

Se, por outro lado, aplicarmos o método apara e preenche à meta-análise da taxa de reincidência associada a programas de reabilitação e reinserção social, obtemos os seguintes resultados:

```
> summary(trimfill(meta.programas,comb.fixed=F))
```

```
Number of studies combined: k = 55 (with 10 added studies)
```

	phi	95%-CI	z	p-value
Random effects model	0.123	[0.048; 0.198]	3.20	0.0014

```
Quantifying heterogeneity:
```

```
tau^2 = 0.0692; H = 3.17 [2.87; 3.51]; I^2 = 90.1% [87.9%; 91.9%];  
Rb = 85.8% [66.6%; 100.0%]
```

```
Test of heterogeneity:
```

Q	d.f.	p-value
543.18	54	< 0.0001

```
Details on meta-analytical method:
```

- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Trim-and-fill method to adjust for funnel plot asymmetry

Para esta meta-análise estima-se que estejam em falta 10 estudos. Repare-se agora que quando se tem em conta o efeito dos estudos ausentes no efeito síntese, a conclusão é que  $\widehat{\phi}_{\text{cor}} = 0.123$ , com um IC a 95% igual a  $[0.048, 0.197]$ . Consequentemente, com a correção do viés passamos a ter alguma evidência a favor da eficácia dos programas na prevenção da reincidência, embora a associação estimada seja fraca. Neste caso, há todo o interesse em localizar mais estudos relevantes para esta meta-análise.

De entre possíveis métodos para modelar o viés de publicação temos, por exemplo, o modelo de seleção de Copas (Copas, 1999). Este modelo engloba dois submodelos distintos:

- i) um MEA para descrever o comportamento dos tamanhos do efeito,

$$\widehat{\theta}_k = \theta_k + \sigma_k \epsilon_k, \quad (6.5)$$

em que  $\theta_k \sim \mathcal{N}(\mu, \tau^2)$ ,  $\epsilon_k \sim \mathcal{N}(0, 1)$  e  $\text{Cov}(\epsilon_k, \epsilon_j) = 0$ ,  $k \neq j$ ,  $k, j = 1, \dots, K$ ;

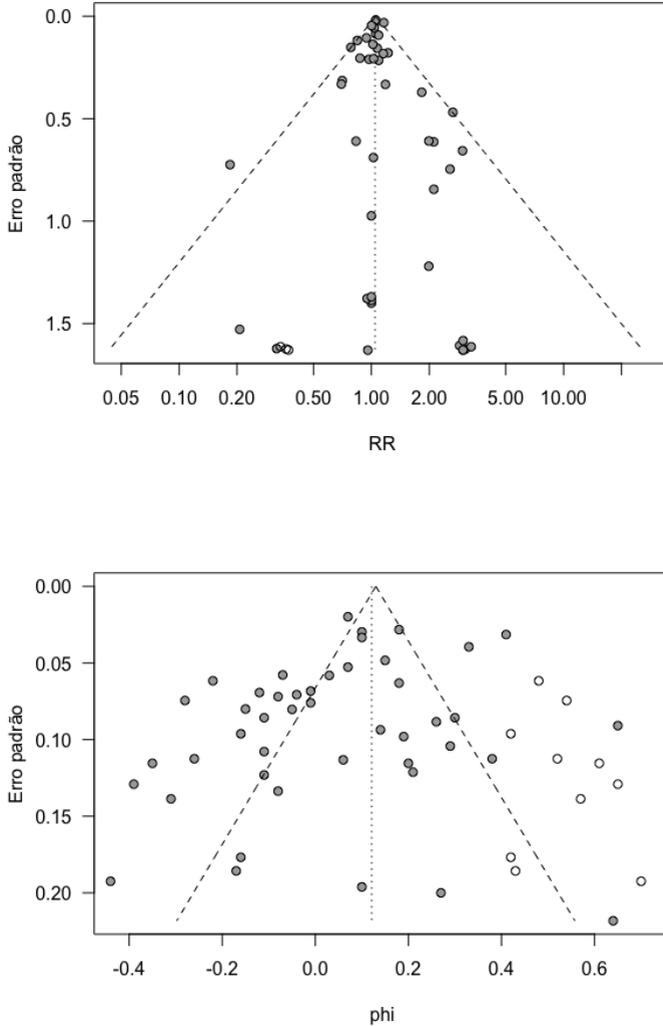
- ii) e um modelo de regressão linear para o processo de publicação (seleção) dos estudos, ou seja,

$$Z_k = \gamma_0 + \frac{\gamma_1}{s_k} + \delta_k, \quad (6.6)$$

onde  $s_k^2 = \widehat{\sigma}_k^2 = \widehat{\text{Var}}(\widehat{\theta}_k)$ ,  $\delta_k \sim \mathcal{N}(0, 1)$  e  $\text{Cor}(\epsilon_k, \delta_k) = \rho$ ,  $k = 1, \dots, K$ .

No modelo de seleção de Copas o coeficiente de correlação  $\rho$  desempenha um papel importante, pois estabelece uma ligação entre os submodelos (6.5) e (6.6) através dos resíduos destes. Note-se que, de acordo com o submodelo (6.6), a seleção de um estudo para publicação depende do seu tamanho, sendo expetável que quanto maior é a precisão de  $\widehat{\theta}_k$ , maior são as hipóteses de seleção do estudo.

Figura 6.2: Gráficos funil para as meta-análises da suplementação de antioxidantes (em cima) e taxa de reincidência associada a programas de reabilitação e reinserção social (em baixo) com a adição dos estudos em falta pelo método apara e preenche



Suponhamos então que existe uma população com  $K$  estudos relevantes para um determinado problema, não estando alguns estudos dessa população publicados (ou selecionados para uma meta-análise). No modelo de seleção Copas, a ausência de viés de publicação corresponde ao caso  $\rho = 0$ , ou seja, quando  $\hat{\theta}_k$  e  $Z_k$  são independentes, pelo que neste cenário a magnitude de  $\hat{\theta}_k$  não terá qualquer influência no facto do estudo vir a ser publicado ou não. Se  $\rho \neq 0$ , existe viés de publicação, de modo que os estudos pequenos que são aceites para publicação, tendem a ter valores elevados para a estimativa do tamanho do efeito, mais concretamente, produzem resultados estatisticamente significativos.

Suponhamos também que os estudos que reportam valores positivos mais elevados para o tamanho do efeito são considerados os “mais interessantes” ( $\rho > 0$ ). Segundo o modelo de seleção de Copas,  $\hat{\theta}_k$  será observado se a variável latente  $Z_k$  assumir um valor positivo. Logo, a probabilidade marginal do estudo  $k$ , com erro padrão  $s_k$ , ser publicado é dada por

$$P_{s_k} = P(Z_k > 0 | s_k, \gamma_0, \gamma_1) = P\left(\delta_k > -\gamma_0 - \frac{\gamma_1}{s_k}\right) = \Phi\left(\gamma_0 + \frac{\gamma_1}{s_k}\right).$$

Consequentemente,  $\Phi(\gamma_0)$  pode ser interpretado como sendo a probabilidade marginal de um estudo ser publicado com um erro padrão infinito, estando o parâmetro  $\gamma_1$  associado à mudança na probabilidade de seleção com o aumento da precisão da estimativa. Espera-se que  $\gamma_1$  seja positivo, pois estudos maiores, isto é, com menores erros padrão associados às estimativas, têm maiores probabilidades de ser publicados.

A distribuição da variável condicional  $\hat{\theta}_k | Z_k > 0$  pode ser usada como modelo para um estudo  $k$  ser incluído numa meta-análise. Mostra-se que

$$E(\hat{\theta}_k | Z_k > 0, s_k, \gamma_0, \gamma_1) = \mu + \rho \sigma_k \lambda \left( \gamma_0 + \frac{\gamma_1}{s_k} \right),$$

onde  $\lambda = \frac{\phi}{\Phi}$  é o rácio de Mills, com  $\phi$  e  $\Phi$  a denotarem a função den-

sidade de probabilidade e a função de distribuição da normal padrão, respetivamente. Assim, se  $\rho > 0$  ( $\rho < 0$ ), a meta-análise efetuada com os estudos selecionados tende a sobrestimar (subestimar) o valor de  $\mu$ . Também se mostra que

$$\text{Var}(\widehat{\theta}_k | Z_k > 0, s_k, \gamma_0, \gamma_1) = \sigma_k^2 [1 - \lambda(u)(u + \lambda(u))] ,$$

onde  $u = \gamma_0 + \frac{\gamma_1}{s_k}$ .

Como os métodos que procuram modelar o viés de publicação assumem à partida um conjunto de pressupostos que não podem ser verificados, é sugerido em Copas (1999) e Copas e Shi (2000, 2001) a realização de uma análise de sensibilidade para o tamanho do efeito, em função da variação dos valores  $\gamma_0$  e  $\gamma_1$ . Na opinião destes autores trata-se de uma abordagem mais sensata, pois permite comparar as inferências de diferentes graus de severidade para o viés de publicação.

Por outro lado, se representarmos os erros padrão dos estudos considerados para uma meta-análise com menor e maior valores por  $s_{\min}$  e  $s_{\max}$ , respetivamente, é possível determinarmos, para um dado par  $(\gamma_0, \gamma_1)$ , as probabilidades marginais de publicação com valor máximo dado por  $P_{s_{\min}} = \Phi\left(\gamma_0 + \frac{\gamma_1}{s_{\min}}\right)$  e valor mínimo  $P_{s_{\max}} = \Phi\left(\gamma_0 + \frac{\gamma_1}{s_{\max}}\right)$ .

No modelo de seleção de Copas os parâmetros  $\gamma_0$  e  $\gamma_1$  são considerados fixos, pelo que não são estimados. No entanto, para um dado par  $(\gamma_0, \gamma_1)$ , os parâmetros  $\mu$ ,  $\rho$  e  $\tau^2$  podem ser estimados pelo método da máxima verosimilhança. Quanto a inferências sobre o principal parâmetro de interesse  $\mu$ , fixado um par  $(\gamma_0, \gamma_1)$ , estas são feitas recorrendo ao teste da razão de verosimilhanças.

Porém, para cada par  $(\gamma_0, \gamma_1)$  plausível, há que verificar se o modelo resultante se ajusta satisfatoriamente ao gráfico funil. Para se efetuar o teste de ajustamento, é adicionado ao modelo (6.5) um termo linear em  $s_k$ , ou seja, considera-se o modelo

$$\widehat{\theta}_k = \theta_k + \beta s_k + \sigma_k \epsilon_k . \quad (6.7)$$

Se para o modelo anterior a hipótese nula  $H_0 : \beta = 0$  não é rejeitada, isto quer dizer que o modelo explica satisfatoriamente a relação linear

entre  $\hat{\theta}_k$  e  $s_k$ . O teste usado neste contexto é também o teste da razão de verossimilhanças, onde se compara a função de verossimilhança de (6.7) com a função de verossimilhança de (6.5) ( $\beta = 0$ ). Para mais detalhes, consultar Copas e Shi (2000). Para obtermos com o modelo de seleção de Copas uma estimativa grosseira para o número total de estudos na população (publicados ou não), com base em  $m$  estudos observados, podemos considerar  $\sum_{k=1}^m 1/P_{s_k}$ .

Com a função *copas* do pacote *metasens* do R podemos aplicar o modelo de seleção de Copas. Por exemplo, aplicando o modelo à meta-análise da suplementação de antioxidantes, obtemos os seguintes resultados:

```
> library(metasens)
> copas(meta.antiox)
```

```
Copas selection model analysis
```

```

                min  max
range of gamma0: -0.26 2.00
range of gamma1:  0.00 0.02
```

```
Largest standard error (SE): 1.6297
```

```
Range of probability publishing trial with largest SE:
```

```

    min    max
0.396 0.978
```

```
Calculation of orthogonal line:
```

level	nobs	adj.r.square	slope	se.slope
0.038	10	0.9690318	-0.5773676	0.03434398
0.039	27	0.9622914	-1.6626928	0.06450081
0.040	24	0.9994024	-4.5886803	0.02339675
0.041	23	0.9972767	-6.9746330	0.07770019
0.042	23	0.9949085	-8.4097958	0.12824959
0.043	22	0.9933760	-9.4171403	0.16778154
0.044	22	0.9919455	-10.1071402	0.19870614
0.045	12	0.9995533	-13.0155658	0.08295433

Summary of Copas selection model analysis:

	publprob	RR	95%-CI	pval.treat	pval.rsb	N.unpubl
	1.00	1.05	[1.01; 1.09]	0.0154	0.5477	0
	0.97	1.05	[1.01; 1.09]	0.0194	0.6096	1
	0.94	1.04	[1.01; 1.09]	0.0251	0.6744	2
	0.91	1.04	[1.00; 1.09]	0.0324	0.7379	4
	0.88	1.04	[1.00; 1.09]	0.0420	0.8030	5
	0.83	1.04	[1.00; 1.09]	0.0549	0.8724	8
	0.74	1.04	[1.00; 1.09]	0.0719	0.9528	14
	0.44	1.04	[0.99; 1.09]	0.0930	0.9614	51
Copas model (adj)	1.05	1.05	[1.01; 1.09]	0.0154	0.5477	0
Random effects model	1.05	1.05	[1.01; 1.09]	0.0199		

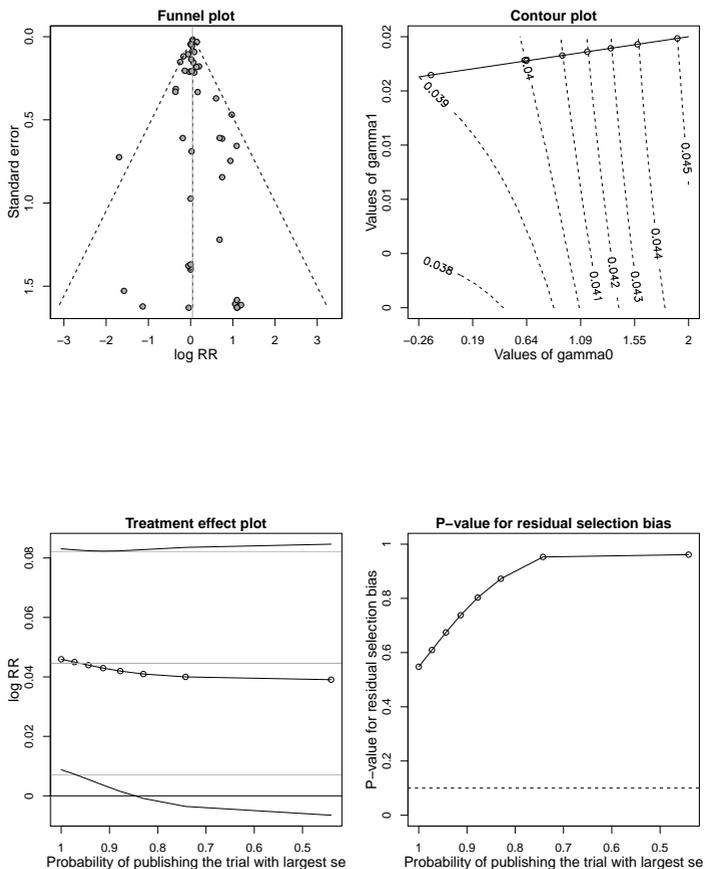
Significance level for test of residual selection bias: 0.1

Legend:

- publprob - Probability of publishing study with largest standard error
- pval.treat - P-value for hypothesis of overall treatment effect
- pval.rsb - P-value for hypothesis that no selection remains unexplained
- N.unpubl - Approximate number of unpublished studies suggested by model

Nos resultados são dadas as amplitudes para os valores  $\gamma_0$  e  $\gamma_1$ , assim como a amplitude para a probabilidade  $P_{s_{\max}}$  ( $s_{\max} = 1.6297$ ). Os valores para as linhas ortogonais são, neste caso, as estimativas do logaritmo do risco relativo para diferentes modelos (6.7) ajustados. Também se indica para cada modelo a estimativa do parâmetro  $\beta$ , o erro padrão associado a esta estimativa e o valor do  $R^2$  ajustado. Abaixo dos valores anteriores são indicados a probabilidade de publicação  $P_{s_{\max}}$ , a estimativa de  $\mu = RR$ , o IC a 95% para  $\mu$  e o valor- $p$  associado ao teste de hipóteses  $H_0 : \ln \mu = 0$  contra  $H_1 : \ln \mu \neq 0$  para alguns modelos. É igualmente indicado o valor- $p$  (*pval.rsb*) associado ao teste com hipótese nula “O modelo explica o processo de seleção dos estudos.”. É ainda indicada a estimativa do número de estudos não publicados para cada cenário.

Figura 6.3: Gráficos obtidos para o modelo de seleção de Copas para a meta-análise da suplementação de antioxidantes



Fixado *a priori* um nível de significância  $\alpha$ , o modelo proposto no fim é aquele que apresenta menor evidência a favor da hipótese do modelo explicar o processo de seleção dos estudos, ou seja, o modelo com menor valor  $pval.rsb \geq \alpha$  (o valor de  $\alpha$  considerado por defeito na função *copas* é 0.1). Assim, vemos que o modelo de seleção de Copas propõe para

a meta-análise da suplementação de antioxidantes o modelo em que  $\hat{\mu}_{aj} = 1.047$ , associado a um  $P_{s_{\max}} = 1$ , o que corresponde à situação em que nenhum estudo da população está por publicar.

Toda a informação relevante sobre a aplicação do modelo de seleção de Copas pode ser representada graficamente, usando-se para o efeito a função *plot* do R aplicada a um objeto *copas*. Na Figura 6.3 apresentamos os gráficos obtidos para a meta-análise da suplementação de antioxidantes. O gráfico do canto superior esquerdo é o gráfico funil para a meta-análise realizada com os estudos selecionados. O gráfico do canto superior direito é um gráfico de contorno cujas linhas a tracejado dizem respeito a alguns valores selecionados para  $\hat{\mu}$  ( $\ln \widehat{RR}$ ), obtidos em função dos valores  $(\gamma_0, \gamma_1)$ . Como as linhas de contorno são aproximadamente paralelas, é traçada uma linha ortogonal (a cheio) às linhas de contorno. Repare-se que à maneira que nos aproximamos do canto superior direito deste gráfico, vamos obtendo modelos cada vez menos seletivos, ou seja, modelos para os quais  $P_{s_{\max}}$  vai aumentando. No gráfico do canto inferior esquerdo, indica-se o valor de  $\hat{\mu}$  em função de  $P_{s_{\max}}$ , assim como umas bandas que delimitam uma região de confiança a 95% para  $\mu$ . Este gráfico mostra-nos qual é o efeito da variação de  $P_{s_{\max}}$  no valor de  $\hat{\mu}$ . No gráfico do canto inferior direito, representa-se o valor de *pval.rsb* em função de  $P_{s_{\max}}$ .

Se, por outro lado, aplicarmos o modelo de seleção de Copas à meta-análise da taxa de reincidência associada a programas de reabilitação e reinserção social, este também indica ausência de viés de publicação, como podemos verificar a partir dos seguintes resultados:

```
> summary(copas(meta.programas))
```

Summary of Copas selection model analysis:

publprob	phi	95%-CI	pval.treat	pval.rsb	N.unpubl
1.00	0.04	[-0.03; 0.11]	0.2374	0.2625	0
0.96	0.04	[-0.03; 0.12]	0.2183	0.2874	1
0.91	0.05	[-0.02; 0.12]	0.1860	0.3195	3
0.86	0.05	[-0.02; 0.13]	0.1611	0.3537	5
0.81	0.06	[-0.02; 0.14]	0.1414	0.3905	7

	0.76	0.06	[-0.02; 0.15]	0.1263	0.4294	10
	0.71	0.07	[-0.02; 0.16]	0.1150	0.4711	13
	0.65	0.07	[-0.02; 0.17]	0.1070	0.5155	17
	0.59	0.08	[-0.02; 0.18]	0.1018	0.5626	22
	0.53	0.08	[-0.02; 0.19]	0.0991	0.6130	30
	0.45	0.09	[-0.02; 0.20]	0.0987	0.6677	40
Copas model (adj)	0.04		[-0.03; 0.11]	0.2374	0.2625	0
Random effects model	0.04		[-0.03; 0.11]	0.2446		

Significance level for test of residual selection bias: 0.1

Legend:

- publprob - Probability of publishing study with largest standard error
- pval.treat - P-value for hypothesis of overall treatment effect
- pval.rsb - P-value for hypothesis that no selection remains unexplained
- N.unpubl - Approximate number of unpublished studies suggested by model

Este último exemplo mostra que as conclusões finais podem ser bastante diferentes, consoante se opte pelo método *apara* e preenche ou pelo modelo de seleção de Copas para corrigir o viés de publicação.



## Capítulo 7

# Calibrar e Combinar Evidência

*Traditional “significance” is only weak evidence for the alternative.*

Kulinskaya *et al.* (2008)

O uso de valores- $p$  como medida de evidência tem suscitado ao longo dos anos várias críticas. Por exemplo, Kulinskaya *et al.* (2008) afirmam que a evidência estatística fica totalmente obscurecida pela interpretação vulgar que se dá a um valor- $p$ . Um valor- $p$  deve ser encarado como uma medida de surpresa, pois sendo calculado sob a validade de uma hipótese nula, torna esta menos verosímil quanto menor for o seu valor. Estes autores também alertam para o facto de que se há entre mãos vários valores- $p$  significativos ( $p_{\text{val}} < 0.05$ ), cresce naturalmente a convicção da hipótese alternativa ser verdadeira. Consequentemente, interpretar e combinar evidência de diferentes estudos torna-se nessas situações uma tarefa difícil, mesmo que se tenha em conta a distribuição para o valor- $p$  sob a hipótese alternativa. Todavia, os valores- $p$  não deixam de ser úteis, porque, tal como defendeu Fisher quando propôs os testes de significância, permitem discriminar resultados desinteres-

santes daqueles que importam investigar mais a fundo.

Para nos mostrarem a dificuldade que existe na interpretação de um valor- $p$  quando há suspeitas da hipótese alternativa ser verdadeira, Kulinskaya *et al.* (2008) comparam os valores- $p$  0.01 e 0.05. Na prática, um valor- $p$  igual a 0.01 é considerado altamente significativo, enquanto o valor 0.05 apenas de significativo, passando a ideia de que a evidência a favor da hipótese alternativa é substancialmente maior para um  $p_{\text{val}} = 0.01$  do que para um  $p_{\text{val}} = 0.05$ . No entanto, se se aplicar a transformação *probit* aos dois valores, por forma a poder-se compará-los na mesma escala, verifica-se que  $\Phi^{-1}(1 - 0.01) = 2.326$  e  $\Phi^{-1}(1 - 0.05) = 1.645$ , pelo que há apenas 41% de evidência a mais, em média, a favor da hipótese alternativa quando  $p_{\text{val}} = 0.01$ . Por isso, Kulinskaya *et al.* (2008) defendem a necessidade de definir corretamente evidência estatística e de como medir essa evidência, sendo que na opinião deles o mais importante é a definição de evidência contra a hipótese nula e a favor da hipótese alternativa. Para o efeito, aconselham o uso de uma escala calibrada, mais concretamente a escala normal unitária, onde a comparação e combinação de evidência faz sempre sentido, independentemente da validade da hipótese nula.

A evidência é então definida como sendo uma estatística  $T$ , que resulta de uma transformação de uma estatística de teste  $S$ , com distribuição (aproximadamente) normal com variância um e valor médio a crescer a partir de zero à medida que se afasta da hipótese nula. Nesta escala calibrada as comparações e combinações de evidência fazem sentido, porque o erro padrão associado à estimativa da evidência média é sempre igual a um. Assim, uma evidência observada  $T = 1.645$ , que corresponde a um  $p_{\text{val}} = 0.05$ , indicará uma evidência fraca a favor da alternativa, o dobro desse valor ( $T = 3.3$ ) uma evidência moderada, enquanto o triplo do valor ( $T = 5$ ) uma evidência forte a favor da hipótese alternativa.

Por outro lado, a transformação de  $S$  na evidência  $T$  é conseguida à custa de uma função monótona crescente  $\mathcal{K}$  chamada *função inferência chave*, que dependerá do modelo subjacente, e para a qual se verificará

$E(T) \approx \sqrt{n}\mathcal{K}(\delta)$ , onde  $n$  representa a dimensão da amostra e  $\delta$  o tamanho do efeito estandardizado (ou não). Para um tamanho do efeito  $\delta$  em particular, o valor  $\kappa = \mathcal{K}(\delta)$  representa o tamanho do efeito (estandardizado) transformado.

Basicamente, o que está em jogo nesta abordagem é tentar encontrar uma transformação estabilizadora da variância  $h$ , de modo que a estatística  $T = h(S)$  tenha distribuição (aproximadamente) normal com variância um (ou próxima de um), ou seja,  $\text{Var}(T) = \text{Var}[h(S)] \approx 1$ . Pretende-se com esta transformação que a variância estabilize no valor um com amostras de dimensão mais pequena, sem que haja necessidade de se recorrer a teoremas limite.

Nesta escala calibrada facilmente se obtêm intervalos de confiança a  $(1 - \alpha)100\%$  para a evidência esperada,  $[L, U] = [T - z_{1-\alpha/2}, T + z_{1-\alpha/2}]$ , e a partir destes intervalos de confiança para o tamanho do efeito  $\delta$  através da transformação  $[\mathcal{K}^{-1}(L/\sqrt{n}), \mathcal{K}^{-1}(U/\sqrt{n})]$ .

No que respeita a escolha da função  $h$ , esta surge do desenvolvimento em série de Taylor da forma como se indica a seguir.

Seja  $S$  uma estatística com valor médio  $\mu_S = E(S)$  e variância  $\sigma_S^2 = \text{Var}(S)$ , e seja  $T = h(S)$ , onde  $h$  é uma função diferenciável pelo menos duas vezes. Então

$$E(T) = E[h(S)] \approx h(\mu_S) \quad \text{e} \quad \text{Var}(T) = \text{Var}[h(S)] \approx \sigma_S^2 [h'(\mu_S)]^2.$$

Para  $\text{Var}(T) = 1$ , temos  $[h'(\mu_S)]^2 = \sigma_S^{-2}$ , pelo que se  $\sigma_S^2 = g(\mu_S)$ , vem

$$h(x) = \int [g(x)]^{-1/2} dx.$$

Para exemplificarmos a abordagem acima, suponhamos que  $(X_1, \dots, X_n)$  é uma amostra aleatória de uma população normal com valor médio  $\mu$  desconhecido e variância  $\sigma^2$  conhecida, e que se pretende testar as hipóteses  $H_0 : \mu = \mu_0$  ( $\mu - \mu_0 = 0$ ) *vs.*  $H_1 : \mu > \mu_0$  ( $\mu - \mu_0 > 0$ ).

O tamanho do efeito  $\theta = \mu - \mu_0$  pode ser estimado por  $S = \hat{\theta} = \bar{X} - \mu_0$ , com  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$ , sendo que valores elevados da estatística  $S$  revelam evidência contra a hipótese de  $\theta = 0$  e a favor de  $\theta > 0$ . A calibração de  $S$  na escala normal unitária irá simplesmente permitir discriminar entre valores elevados e não elevados de  $S$ .

Neste caso, o tamanho do efeito estandardizado é  $\delta = \theta/\sigma$ , sendo este valor estimado por  $\hat{\delta} = S/\sigma$ . Consequentemente, as hipóteses anteriores são equivalentes às hipóteses  $H_0 : \delta = 0$  vs.  $H_1 : \delta > 0$ . Como  $S \sim \mathcal{N}(\theta, \sigma^2/n)$ , a transformação para evidência contra a hipótese nula e a favor da alternativa é naturalmente

$$T = \frac{\sqrt{n}S}{\sigma} = \sqrt{n}\hat{\delta} \sim \mathcal{N}(\sqrt{n}\delta, 1),$$

onde  $E(T) = \sqrt{n}\delta$  assume o valor 0 sob a validade de  $H_0$  e que vai aumentando à medida que se afasta de  $H_0$ . Quanto à função inferência chave, esta é a função identidade, ou seja,  $\mathcal{K}(\delta) = \delta$ .

Para o caso da variância  $\sigma^2$  ser desconhecida, o tamanho do efeito estandardizado  $\delta$  será estimado por  $\hat{\delta} = (\bar{X} - \mu_0)/S_{n-1}$ , onde  $S_{n-1}^2 = \hat{\sigma}^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2$ . Neste caso, valores elevados da estatística  $\hat{\delta}$  revelam evidência a favor da alternativa  $\delta > 0$ . Sob a validade de  $H_0$ ,

$$S = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S_{n-1}} = \sqrt{n}\hat{\delta} \sim t_{n-1},$$

enquanto sob a validade de  $H_1$ ,  $S$  tem distribuição  $t$  de Student não central com  $n - 1$  graus de liberdade e parâmetro de não centralidade  $\lambda = \sqrt{n}\delta$ , ou seja,  $S \sim t_{n-1}(\lambda)$ . A transformação de  $S$  em evidência a favor da alternativa  $\delta > 0$ , ou equivalentemente a favor de  $\lambda > 0$ ,<sup>1</sup> tem função inferência chave

$$\mathcal{K}(\delta) = \sqrt{2} \sinh^{-1} \left( \frac{\delta}{\sqrt{2}} \right) = \sqrt{2} \ln \left( \frac{\delta}{\sqrt{2}} + \sqrt{1 + \frac{\delta^2}{2}} \right).$$

<sup>1</sup>Testar  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu > \mu_0$  equivale a testar  $H_0 : \lambda = 0$  vs.  $H_1 : \lambda > 0$ .

**Proposição 7.1.** *A evidência numa estatística  $S \sim t_\nu$  para testar  $H_0 : \mu = \mu_0$  vs.  $H_1 : \mu > \mu_0$  é*

$$T = \sqrt{2\nu} \operatorname{senh}^{-1} \left( \frac{S}{\sqrt{2\nu}} \right). \quad (7.1)$$

De facto, se  $X \sim t_\nu(\lambda)$ ,

$$E(X) = \lambda \sqrt{\frac{\nu}{2}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} = \lambda \left( 1 + \frac{3}{\nu} + O\left(\frac{1}{\nu^2}\right) \right), \quad \nu > 1,$$

e

$$\operatorname{Var}(X) = \frac{\nu}{\nu-2} (1 + \lambda)^2 - \lambda^2 \frac{\nu}{2} \left[ \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right]^2, \quad \nu > 2,$$

em que  $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ ,  $\alpha > 0$ , representa a função gama de Euler. Dado que

$$\operatorname{Var}(X) = 1 + \frac{\lambda^2}{2\nu} + O\left(\frac{1}{\nu^2}\right) = 1 + \frac{[E(X)]^2}{2\nu} + O\left(\frac{1}{\nu^2}\right),$$

segue-se que  $g(x) = 1 + \frac{x^2}{2\nu}$ . Logo, a transformação estabilizadora para a variância de  $X$  é

$$h(x) = \int \frac{1}{\sqrt{1 + x^2/(2\nu)}} dx = \sqrt{2\nu} \operatorname{senh}^{-1} \left( \frac{x}{\sqrt{2\nu}} \right) + C,$$

sendo usual assumir-se  $C = 0$ .<sup>2</sup>

## Propriedades Desejáveis para uma Evidência Unilateral

Sejam  $\theta$  um tamanho do efeito e  $S$  uma estatística para testar  $H_0 : \theta = 0$  vs.  $H_1 : \theta > 0$ , sendo  $H_0$  rejeitada para valores elevados de  $S$ . Uma medida de evidência unilateral (à direita)  $T = h(S)$  deve satisfazer as seguintes propriedades:

- i)  $T$  é uma função monótona crescente de  $S$ ;

---

<sup>2</sup>Quando em (7.1) se tem  $\nu = n - 1$ , pode-se considerar como aproximação  $\nu = n$ .

- ii)  $T$  tem distribuição normal, independentemente dos valores dos parâmetros desconhecidos;
- iii)  $\text{Var}(T) = 1$  para todos os valores dos parâmetros desconhecidos;
- iv)  $E(T) = \eta(\theta)$  é uma função monótona crescente em  $\theta$ , com  $\eta(0) = 0$ .

## 7.1 Combinar Evidência

A questão que naturalmente se coloca neste momento é como combinar a evidência segundo esta abordagem?

Suponhamos que são considerados  $K$  estudos independentes e que se dispõe da evidência  $T_k$  contra  $H_{0k} : \delta_k = 0$  e a favor de  $H_{1k} : \delta_k > 0$ , com  $T_k \sim \mathcal{N}(\sqrt{n_k} \kappa_k, 1) \equiv \mathcal{N}(\sqrt{n_k} \mathcal{K}(\delta_k), 1)$ ,  $k = 1, \dots, K$ .<sup>3</sup> A forma de combinar a evidência fornecida pelos diferentes estudos dependerá de qual dos seguintes cenários se observar:

**Cenário 1:** Um modelo de efeitos fixos iguais, ou seja,  
 $\delta_1 = \dots = \delta_K = \delta$ ;

**Cenário 2:** Um modelo de efeitos fixos diferentes, ou seja,  $\delta_k \neq \delta_j$  para um par  $k \neq j$ ;

**Cenário 3:** Um modelo de efeitos aleatórios.

---

<sup>3</sup>Evidência a favor da alternativa unilateral à esquerda  $\delta < 0$ , pode ser tratada por simetria, ou seja, substituindo  $T$  por  $-T$ . Consequentemente, evidência negativa para a alternativa  $\delta > 0$  é evidência positiva para alternativa  $\delta < 0$ . Por outro lado, evidência para uma alternativa bilateral pode ser definida à custa da evidência unilateral  $T$  através de  $T^2 = S$ . Assim, se  $F$  denotar a função de distribuição do qui-quadrado com 1 grau de liberdade e  $M = \chi_{1;0.5}^2$  o valor da mediana, a evidência para uma alternativa bilateral é dada por

$$T^\pm = \begin{cases} \sqrt{S - M/2} - \sqrt{M/2} & , \text{ se } S \geq M \\ -\sqrt{S^* - M/2} + \sqrt{M/2} & , \text{ se } S < M \end{cases}$$

em que  $S^* = F^{-1}(1 - F(S))$ .

Uma abordagem usual em Meta-Análise consiste em recorrer ao teste  $Q$  de Cochran para testar a homogeneidade dos tamanhos do efeito e em função dos resultados deste teste, decidir entre um MEF ou um MEA. No entanto, Kulinskaya *et al.* (2008) chamam a atenção para o facto da evidência para a heterogeneidade ser diferente nos dois modelos.

- **Combinar Evidência num MEF**

Valores elevados da estatística  $Q$  de Cochran definida da forma

$$Q = \sum_{k=1}^K \omega_k (\hat{\delta}_k - \hat{\delta})^2, \quad (7.2)$$

onde  $\omega_k = 1/\sigma_k^2$  e  $\hat{\delta} = \sum_{k=1}^K \omega_k \hat{\delta}_k / \sum_{k=1}^K \omega_k$ , revelam evidência contra a hipótese de efeitos fixos iguais e a favor da alternativa de efeitos fixos diferentes.

Sob a validade da hipótese alternativa de efeitos fixos diferentes,  $Q$  tem distribuição do qui-quadrado não central com  $K - 1$  graus de liberdade e parâmetro de não centralidade  $\lambda = \sum_{k=1}^K \omega_k (\delta_k - \delta)^2$ , em que  $\delta = \sum_{k=1}^K \omega_k \delta_k / \sum_{k=1}^K \omega_k$ , ou seja,  $Q \sim \chi_{K-1}^2(\lambda)$ . Consequentemente, testar as hipóteses anteriores equivale a testar as hipóteses  $H_0 : \lambda = 0$  vs.  $H_1 : \lambda > 0$ .<sup>4</sup>

**Proposição 7.2.** *Sejam  $F_\nu$  a função de distribuição de uma variável aleatória com distribuição do qui-quadrado central com  $\nu$  graus de liberdade e  $M_\nu = \chi_{\nu,0.5}^2$  o valor da mediana de  $F_\nu$ . A transformação de uma estatística  $S_\nu \sim \chi_\nu^2(\lambda)$  em evidência  $T$  contra  $H_0 : \lambda = 0$  e a favor de  $H_1 : \lambda > 0$  é*

$$T = \begin{cases} \sqrt{S_\nu - M_\nu/2} - \sqrt{M_\nu/2} & , \text{ se } S_\nu \geq M_\nu \\ -\sqrt{S_\nu^* - M_\nu/2} + \sqrt{M_\nu/2} & , \text{ se } S_\nu < M_\nu \end{cases}, \quad (7.3)$$

onde  $S_\nu^* = F_\nu^{-1}(1 - F_\nu(S_\nu))$ .

---

<sup>4</sup>Caso se tenha que estimar os  $\omega_k$ 's, sabe-se que se  $K$  é fixo e  $\hat{\sigma}_k^2 = \widehat{\text{Var}}(\hat{\delta}_k) \rightarrow 0$ ,  $k = 1, \dots, K$ , então  $Q \stackrel{\mathcal{L}}{\sim} \chi_{K-1}^2(\lambda)$ .

A função inferência chave para o modelo do qui-quadrado não central, com  $\nu$  graus de liberdade e parâmetro de não centralidade  $\lambda$ , para testar  $H_0 : \theta = 0$  vs.  $H_1 : \theta > 0$ , onde  $\theta = \lambda/n$ , é

$$\mathcal{K}_{n,\nu}(\theta) = \mathcal{L}_{n,\nu}(\theta) - \frac{\theta + \frac{\nu}{2n}}{2n\mathcal{L}_{n,\nu}^3(\theta)} - \sqrt{\frac{M_\nu}{2n}},$$

em que

$$\mathcal{L}_{n,\nu}(\theta) = \sqrt{\theta + \frac{2\nu - M_\nu}{2n}}.$$

Logo, a evidência contra a hipótese de homogeneidade dos tamanhos do efeito e a favor da alternativa de tamanhos do efeito diferentes é obtida a partir de (7.3), substituindo  $S_\nu$  por  $Q$ .

Quanto aos tamanhos do efeito transformados  $\kappa_k = \mathcal{K}(\delta_k)$ , como  $T_k \sim \mathcal{N}(\sqrt{n_k} \kappa_k, 1)$ , estes podem ser estimados por  $\hat{\kappa}_k = T_k / \sqrt{n_k}$ , pelo que  $\hat{\kappa}_k \sim \mathcal{N}(\kappa_k, n_k^{-1})$ ,  $k = 1, \dots, K$ . Assim, para testarmos a hipótese de homogeneidade dos tamanhos do efeito transformados, ou seja,  $H_0^* : \kappa_1 = \dots = \kappa_K = \kappa$ , contra a hipótese dos tamanhos do efeito transformados ser diferentes,  $H_1^* : \kappa_k \neq \kappa_j$ , para um par  $k \neq j$ , devemos considerar para estatística de teste

$$Q^* = \sum_{k=1}^K n_k (\hat{\kappa}_k - \hat{\kappa})^2, \quad (7.4)$$

onde  $\hat{\kappa} = \frac{1}{N} \sum_{k=1}^K n_k \hat{\kappa}_k$  e  $N = \sum_{k=1}^K n_k$ .

Para  $k$  fixo e  $n_k \rightarrow \infty$ ,  $k = 1, \dots, K$ , tem-se que sob a validade de  $H_0^*$ ,  $Q^* \stackrel{a}{\sim} \chi_{K-1}^2$ , enquanto sob a validade de  $H_1^*$ ,  $Q^* \stackrel{a}{\sim} \chi_{K-1}^2(\lambda^*)$ , em que  $\lambda^* = \sum_{k=1}^K n_k (\kappa_k - \kappa)^2$  e  $\kappa = \frac{1}{N} \sum_{k=1}^K n_k \kappa_k$ . Consequentemente, a evidência contra  $H_0^* : \lambda^* = 0$  e a favor de  $H_1^* : \lambda^* > 0$  é também dada por (7.3), onde  $S_\nu$  é substituído agora por  $Q^*$ .

*Observação 7.1.* Uma vantagem de  $Q^*$  em relação a  $Q$  é que a primeira estatística não necessita dos pesos  $\omega_k = 1/\sigma_k^2$ , geralmente estimados. Todavia, pode haver homogeneidade dos tamanhos do efeito sem que haja homogeneidade dos tamanhos do efeito transformados, e vice-versa.

Sob o Cenário 1 a evidência combinada é

$$T_{\text{comb}} = \frac{\sum_{k=1}^K \sqrt{n_k} T_k}{\sqrt{N}} \sim \mathcal{N}(\sqrt{N}\kappa, 1), \quad (7.5)$$

onde  $N = \sum_{k=1}^K n_k$ . Na eventualidade de  $T_{\text{comb}}$  ser negativo, então  $|T_{\text{comb}}|$  revela evidência positiva para  $\delta < 0$ . Por outro lado, sob a validade da hipótese de homogeneidade dos tamanhos do efeito transformados, tem-se que o estimador do tamanho do efeito transformado  $\kappa$  combinado verifica

$$\hat{\kappa}_{\text{comb}} = \frac{T_{\text{comb}}}{\sqrt{N}} \sim \mathcal{N}(\kappa, N^{-1}).$$

Por conseguinte, o efeito  $\delta$  combinado obtém-se a partir da relação  $\delta = \mathcal{K}^{-1}(\kappa)$ , ou seja,  $\hat{\delta}_{\text{comb}} = \mathcal{K}^{-1}(\hat{\kappa}_{\text{comb}})$ .

Sob o Cenário 2, ou seja, para tamanhos do efeito fixos diferentes, a evidência combinada define-se da mesma forma que em (7.5). No entanto, o tamanho do efeito transformado combinado é estimado por

$$\hat{\kappa}_{\text{comb}} = \frac{\sum_{k=1}^K n_k \hat{\kappa}_k}{N}, \quad (7.6)$$

com  $\hat{\kappa}_k = \mathcal{K}(\hat{\delta}_k)$ ,  $k = 1, \dots, K$ . Logo, o  $\delta$  combinado é também obtido através da relação  $\hat{\delta}_{\text{comb}} = \mathcal{K}^{-1}(\hat{\kappa}_{\text{comb}})$ , onde  $\hat{\kappa}_{\text{comb}}$  está definido em (7.6).<sup>5</sup>

Repare-se que as únicas diferenças entre os Cenários 1 e 2 dizem respeito à forma como se obtém o tamanho do efeito  $\delta$  combinado (ou  $\kappa$  combinado) e à sua interpretação.

### • Combinar Evidência num MEA

Sob o Cenário 3, os tamanhos do efeito  $\delta_k$  passam a ser estocásticos, com  $\delta_k \sim \mathcal{N}(\mu, \tau^2)$ ,  $k = 1, \dots, K$ . Consequentemente,

---

<sup>5</sup>No caso da função inferência chave depender de um parâmetro perturbador  $\zeta$ , e por conseguinte termos  $\kappa_k = \mathcal{K}(\delta_k, \zeta_k)$ , é usado um valor para representar  $\zeta$ , nomeadamente  $\zeta = \frac{1}{N} \sum_{k=1}^K n_k \zeta_k$ . Se  $\mathcal{K}$  for uma função estritamente monótona em ambos os argumentos, a estimativa  $\hat{\zeta} = \frac{1}{N} \sum_{k=1}^K n_k \hat{\zeta}_k$  pode ser usada para resolver a equação  $\hat{\kappa} = \mathcal{K}(\hat{\delta}, \hat{\zeta})$  em ordem a  $\hat{\delta}$ .

os tamanhos do efeito transformados  $\kappa_k = \mathcal{K}(\delta_k)$  também passam a ser estocásticos, com  $\kappa_k \sim \mathcal{N}(\kappa, \tau^2)$ , seguindo-se que  $\hat{\kappa}_k \sim \mathcal{N}(\kappa, n_k^{-1} + \tau^2)$ ,  $k = 1, \dots, K$ .

Por exemplo, para testarmos  $H_0 : \tau^2 = 0$  vs.  $H_1 : \tau^2 > 0$ , podemos usar as estatísticas  $Q$  e  $Q^*$  definidas em (7.2) e (7.4), respetivamente, que sob a validade de  $H_0$  têm ambas distribuição do qui-quadrado com  $K - 1$  graus de liberdade. No entanto, sob a validade de  $H_1$ ,  $Q$  e  $Q^*$  não têm distribuição do qui-quadrado não central.

Considerando  $M_r = \sum_{k=1}^K \omega_k^r$ ,  $a = M_1 - M_2/M_1$ ,  $b = M_2 - 2M_3/M_1 + (M_2/M_1)^2$ ,  $c = b/a^2$  e  $d = c(K - 1) - 1$ , Biggerstaff e Tweedie (1997) mostram que

$$E(Q) = K - 1 + a\tau^2 \quad \text{e} \quad \text{Var}(Q) = 2(K - 1) + 4a\tau^2 + 2b\tau^4.$$

Estes autores propõem como aproximação para a distribuição de  $Q$  a distribuição gama com os dois momentos anteriores. Se assumirmos então um MEA, a transformação de  $Q$  em evidência contra  $\tau^2 = 0$  e a favor de  $\tau^2 > 0$  é

$$T'_Q = \frac{1}{\sqrt{2c}} \ln \left( \frac{cQ - d + \sqrt{(cQ - d)^2 + d}}{1 + \sqrt{1 + d}} \right). \quad (7.7)$$

Por outro lado, a transformação de  $Q^*$  na evidência  $T'_{Q^*}$  a favor da alternativa  $\tau^2 > 0$  também é dada pela fórmula (7.7), em que se substitui  $Q$  por  $Q^*$  e se considera  $\omega_k = n_k$ ,  $k = 1, \dots, K$ .

Quanto à combinação de evidência neste cenário, consideremos  $\bar{\kappa} = \frac{1}{K} \sum_{k=1}^K \hat{\kappa}_k$  e  $S_\kappa^2 = \frac{1}{K-1} \sum_{k=1}^K (\hat{\kappa}_k - \bar{\kappa})^2$ . Como  $E(\bar{\kappa}) = \kappa$ ,  $\text{Var}(\bar{\kappa}) = \sigma_\kappa^2 = \tau^2 + \frac{1}{K} \sum_{k=1}^K n_k^{-1}$  e  $\bar{\kappa} \sim \mathcal{N}(\kappa, \sigma_\kappa^2)$ , em que  $\hat{\sigma}_\kappa^2 = S_\kappa^2$ , tem-se que a estatística de teste para as hipóteses  $H_0 : \kappa = 0$  vs.  $H_1 : \kappa > 0$  é

$$S_{K-1}^* = \frac{\sqrt{K}(\bar{\kappa} - 0)}{S_\kappa} = \frac{\sqrt{K} \bar{\kappa}}{S_\kappa}.$$

Não se conhece a distribuição exata da estatística anterior, mas quando  $n_k = n$ ,  $k = 1, \dots, K$ , sabe-se que  $S_{K-1}^* \sim t_{K-1}$  sob  $H_0$

e  $S_{K-1}^* \sim t_{K-1}(\lambda)$  sob  $H_1$ , com parâmetro de não centralidade  $\lambda = \sqrt{K} \kappa / \sqrt{n^{-1} + \tau^2}$ . O resultado anterior pode ser usado como aproximação se cada  $1/n_k$  for pequeno quando comparado com o valor de  $\tau^2$ .

Se, por outro lado, atendermos à Proposição 7.1, segue-se que a combinação de evidência para  $\kappa = \mathcal{K}(\delta) > 0$ , e para  $\delta > 0$ , num MEA é dada por

$$T_{\text{comb}}^* = \sqrt{2K} \sinh^{-1} \left( \frac{S_{K-1}^*}{\sqrt{2K}} \right) = \sqrt{2K} \sinh^{-1} \left( \frac{\bar{\kappa}}{\sqrt{2} S_\kappa} \right),$$

desde que  $n_k \geq 10$ ,  $k = 1, \dots, K$ , e  $\tau > 2s_{1/n_k}$ , com  $s_{1/n_k}$  a denotar o desvio padrão amostral corrigido da amostra dos valores  $1/n_k$ .

A abordagem apresentada neste capítulo para uma meta-análise não está isenta de críticas, sendo as principais as seguintes:

- i) Dificuldade em encontrar uma função inferência chave para alguns modelos — são no entanto conhecidas transformações estabilizadoras da variância para os modelos mais comuns como o binomial, Poisson,  $t$  de Student, qui-quadrado e  $F$  de Snedecor;
- ii) Não estabilização da variância, ou aproximação inválida pela distribuição normal à distribuição da evidência, especialmente quando amostras de dimensão pequena estão envolvidas.

Para aplicarmos a abordagem anterior, consideremos novamente o exemplo da suplementação de antioxidantes. Vejamos, em primeiro lugar, qual é a função inferência chave para a distribuição do estimador do risco relativo. Neste caso, como a evidência para o risco relativo e para o *odds ratio* é a mesma que para a diferença entre duas proporções, abordaremos esta última situação.

Sejam  $X_1 \sim \text{Binomial}(n_1, p_1)$  e  $X_2 \sim \text{Binomial}(n_2, p_2)$  variáveis aleatórias independentes. Suponhamos que se pretende definir a evidência contra  $H_0 : \theta = p_1 - p_2 = 0$  e a favor de  $H_1 : \theta = p_1 - p_2 > 0$ . Denotando por  $N = n_1 + n_2$  e  $q = n_2/N$ , onde  $q$  representa a proporção da amostra combinada atribuída à segunda amostra, podemos considerar o parâmetro definido por Brown e Li (2005)

$$p = qp_1 + (1 - q)p_2,$$

e a partir do qual se obtêm  $p_1 = p + (1 - q)\theta$  e  $p_2 = p - q\theta$ .

Neste caso,  $\hat{\theta} = \hat{p}_1 - \hat{p}_2 = X_1/n_1 - X_2/n_2$ , com

$$\text{Var}(\hat{\theta}) = \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2} = \frac{\zeta - \theta^2}{N},$$

em que  $\zeta = p(1 - p)/[q(1 - q)]$ .<sup>6</sup> Dado que  $\text{Var}(\hat{\theta}) = g(\text{E}(\hat{\theta})) = g(\theta)$ , tem-se que  $g(x) = a - bx^2$ , com  $a = \zeta/N$  e  $b = 1/N$ . Logo, a função estabilizadora da variância para o estimador  $\hat{\theta}$  é

$$h(x) = \int (a - bx^2)^{-1/2} dx = \sqrt{N} \arcsen(\zeta^{-1/2}x) + C,$$

onde também se considera para esta família de transformações  $C = 0$ . Consequentemente, a transformação de  $\hat{\theta}$  em evidência contra  $\theta = 0$  e a favor de  $\theta > 0$  é

$$T = \sqrt{N} \arcsen(\hat{\zeta}^{-1/2}\hat{\theta}),$$

em que

$$T \stackrel{a}{\sim} \mathcal{N}(\sqrt{N} \arcsen(\zeta^{-1/2}\theta), 1). \quad (7.8)$$

Para o tamanho do efeito estandardizado  $\delta = \theta/\sqrt{N \text{Var}(\hat{\theta})} = \theta/\sqrt{\zeta - \theta^2}$ , a função inferência chave é

$$\mathcal{K}(\delta) = \arcsen\left(\frac{\delta}{\sqrt{1 + \delta^2}}\right) = \arcsen\left(\frac{\theta}{\sqrt{\zeta}}\right).$$

<sup>6</sup>Kulinskaya *et al.* (2008) sugerem usar em alternativa  $\tilde{p}_k = (X_k + 0.5)/(n_k + 1)$ , em vez de  $\hat{p}_k = X_k/n_k$ ,  $k = 1, 2$ .

Por outro lado, como

$$\text{RR} = \frac{p_1}{p_2} = \frac{p + (1 - q)\theta}{p - q\theta},$$

tem-se para  $p$  fixo que o RR é uma função crescente em  $\theta$ . Assim, testar as hipóteses  $H_0 : \text{RR} = 1$  vs.  $H_1 : \text{RR} > 1$  equivale a testar as hipóteses  $H_0 : \theta = 0$  vs.  $H_1 : \theta > 0$ . Consequentemente, a evidência para um tamanho do efeito positivo, quer se use o parâmetro  $\theta$ , quer se use o parâmetro RR, é precisamente a mesma.<sup>7</sup> Na prática a evidência observada para um estudo  $k$  será calculada a partir da fórmula  $T_k = \sqrt{N_k} \arcsen(\widehat{\zeta}_k^{-1/2} \widehat{\theta}_k)$ .

A partir do resultado (7.8) podemos obter um intervalo de confiança aproximado a  $(1 - \alpha)100\%$  para  $\theta$ , nomeadamente

$$\left[ \sqrt{\widehat{\zeta}} \operatorname{sen}\left(\frac{T - z_{1-\alpha/2}}{\sqrt{N}}\right), \sqrt{\widehat{\zeta}} \operatorname{sen}\left(\frac{T + z_{1-\alpha/2}}{\sqrt{N}}\right) \right].$$

Para obtermos um intervalo de confiança aproximado a  $(1 - \alpha)100\%$  para o risco relativo a partir do intervalo de confiança de  $\theta$ , basta atendermos a que

$$\theta = p_1 - p_2 = p_2 \left( \frac{p_1}{p_2} - 1 \right) = p_2(\text{RR} - 1),$$

ou seja, considerarmos o intervalo de confiança

$$\left[ \frac{\sqrt{\widehat{\zeta}}}{\widehat{p}_2} \operatorname{sen}\left(\frac{T - z_{1-\alpha/2}}{\sqrt{N}}\right) + 1, \frac{\sqrt{\widehat{\zeta}}}{\widehat{p}_2} \operatorname{sen}\left(\frac{T + z_{1-\alpha/2}}{\sqrt{N}}\right) + 1 \right].$$

Quanto ao risco relativo combinado, este é dado por

$$\widetilde{\text{RR}}_{\text{comb}} = \frac{\sqrt{\widehat{\zeta}_{\text{comb}}}}{\widehat{p}_{2,\text{comb}}} \operatorname{sen}\left(\frac{T_{\text{comb}}}{\sqrt{N^*}}\right) + 1, \quad (7.9)$$

---

<sup>7</sup>Da mesma forma, testar hipóteses sobre o *odds ratio* do tipo  $H_0 : \text{OR} = 1$  vs.  $H_1 : \text{OR} > 1$ , equivale a testar as hipóteses  $H_0 : \theta = 0$  vs.  $H_1 : \theta > 0$ , porque para  $p$  fixo, OR também é uma função crescente em  $\theta$ .

onde  $\hat{p}_{i,\text{comb}} = \sum_{k=1}^K X_{ki} / \sum_{k=1}^K n_{ki}$ , com  $X_{ki}$  a representar o número de sucessos na amostra  $i$  no estudo  $k$  e  $n_{ki}$  a dimensão da amostra  $i$  no estudo  $k$ ,  $i = 1, 2$ ,  $k = 1, \dots, K$ , e  $\hat{\zeta}_{\text{comb}} = \hat{p}_{\text{comb}}(1 - \hat{p}_{\text{comb}}) / [q_{\text{comb}}(1 - q_{\text{comb}})]$ , com  $q_{\text{comb}} = \frac{1}{N^*} \sum_{k=1}^K n_{k2}$ ,  $N^* = \sum_{k=1}^K N_k = \sum_{k=1}^K (n_{k1} + n_{k2})$  e  $\hat{p}_{\text{comb}} = q_{\text{comb}} \hat{p}_{1,\text{comb}} + (1 - q_{\text{comb}}) \hat{p}_{2,\text{comb}}$ .

*Observação 7.2.* Sobre o tamanho do efeito  $\theta = \ln \text{RR}$ , Kulinskaya *et al.* (2008) afirmam que apesar do estimador estandardizado  $S = \hat{\theta} / \sqrt{\text{Var}(\hat{\theta})}$  convergir para uma normal com variância um, não é claro que  $S$  satisfaça as propriedades de uma medida de evidência. O mesmo tipo de comentário se aplica ao tamanho do efeito  $\theta = \ln \text{OR}$ .

Para o exemplo da suplementação de antioxidantes, das 47 evidências observadas, 12 são evidências negativas, ou seja, evidências positivas para  $\text{RR} < 1$ . Por outro lado, 5 estudos mostram evidência fraca a favor da alternativa, enquanto apenas 1 estudo revela evidência moderada a favor da alternativa (Goodman *et al.*, 2004).

Para definirmos a evidência contra a hipótese dos tamanhos do efeito transformados fixos ser iguais, isto é,  $H_0 : \kappa_1 = \dots = \kappa_{47} = \kappa$ , e a favor da hipótese dos tamanhos do efeito transformados fixos ser diferentes, ou seja,  $H_1 : \kappa_k \neq \kappa_j$ , para um par  $k \neq j$ , temos que calcular em primeiro lugar o valor da estatística  $Q^*$  de Cochran definida por

$$Q^* = \sum_{k=1}^K N_k (\hat{\kappa}_k - \hat{\kappa})^2,$$

onde  $\hat{\kappa} = \frac{1}{N^*} \sum_{k=1}^K N_k \hat{\kappa}_k$  e  $N^* = \sum_{k=1}^K N_k$ .

Como  $Q_{\text{obs}}^* = 62.13076$  ( $N^* = 180945$  e  $\hat{\kappa} = 8.82 \times 10^{-3}$ ), transformando  $Q^*$  em evidência de acordo com a fórmula (7.3) ( $Q_{\text{obs}}^* > \chi_{46;0.5}^2 = 45.335$ ), obtemos  $T_{Q^*} = 1.5209 < 1.645$ . Assim, a evidência a favor da alternativa não chega a qualificar-se como fraca, pelo que podemos assumir o Cenário 1.

Assumindo então um modelo de efeitos transformados fixos iguais para a meta-análise, obtemos para valor da evidência combinada  $T_{\text{comb}} = 3.7516$ , com um IC a 95% para a evidência combinada esperada de

$$[T_{\text{comb}} - 1.96, T_{\text{comb}} + 1.96] = [1.792, 5.712].$$

Repare-se que o intervalo anterior contempla a possibilidade de uma evidência esperada fraca a forte a favor da hipótese alternativa  $RR > 1$ .

Em relação ao tamanho do efeito transformado combinado, temos  $\hat{\kappa}_{\text{comb}} = T_{\text{comb}}/\sqrt{N^*} = 8.82 \times 10^{-3}$ , com um IC a 95% para o tamanho do efeito transformado comum  $\kappa$  igual a

$$\left[ \frac{T_{\text{comb}} - 1.96}{\sqrt{N^*}}, \frac{T_{\text{comb}} + 1.96}{\sqrt{N^*}} \right] = [0.0042, 0.0134].$$

Atendendo agora a que  $\tilde{p}_{1,\text{comb}} = 0.155$ ,  $\tilde{p}_{2,\text{comb}} = 0.111$ ,  $q_{\text{comb}} = 0.452$ ,  $\tilde{p}_{\text{comb}} = 0.131$  e  $\tilde{\zeta}_{\text{comb}} = 0.460$ , vem

$$\tilde{\theta}_{\text{comb}} = \sqrt{\tilde{\zeta}_{\text{comb}}} \operatorname{sen} \left( \frac{T_{\text{comb}}}{\sqrt{N^*}} \right) = 0.006$$

e, por conseguinte,

$$\widetilde{RR}_{\text{comb}} = 1.054.$$

Assim, estima-se um aumento de 5.4% no risco de mortalidade com a suplementação de antioxidantes, com um IC a 95% para o risco relativo igual a  $[1.026, 1.081]$ . Como podemos constatar, os resultados obtidos com esta abordagem são bastante semelhantes aos que foram obtidos com a abordagem mais clássica para um MEF.

O problema do viés de publicação é encarado em Kulinskaya *et al.* (2008) de outra forma. Estes autores afirmam que deve ser modelado tendo em conta o mecanismo que o gerou por forma a poder-se corrigi-lo.

Suponhamos que  $T_k$  representa a evidência a favor da alternativa no estudo  $k$ , evidência esta calculada com base numa amostra de dimensão  $n_k$ , com  $T_k \sim \mathcal{N}(\sqrt{n_k}\kappa, 1)$ ,  $k = 1, \dots, K$ . Como o viés de publicação resulta frequentemente do facto dos estudos para os quais se observa um  $p_{\text{val}} \geq 0.05$  ( $T_k \leq 1.645$ ) não ser publicados, o modelo correto para a evidência  $T_k$  será a distribuição normal truncada à esquerda do ponto 1.645, ou seja, a distribuição da variável aleatória condicional  $T_k|T_k \geq 1.645$ , com função de distribuição

$$F_{T_k|T_k \geq 1.645}(x) = \begin{cases} 0 & , \text{ se } x < 1.645 \\ \frac{\Phi(x - \sqrt{n_k}\kappa) - \Phi(1.645 - \sqrt{n_k}\kappa)}{1 - \Phi(1.645 - \sqrt{n_k}\kappa)} & , \text{ se } x \geq 1.645 \end{cases}.$$

Como

$$E(T_k|T_k \geq 1.645) = \sqrt{n_k}\kappa + \frac{\phi(1.645 - \sqrt{n_k}\kappa)}{1 - \Phi(1.645 - \sqrt{n_k}\kappa)},$$

o viés associado à evidência  $T_k$  é conhecido e dado por

$$\frac{\phi(1.645 - \sqrt{n_k}\kappa)}{1 - \Phi(1.645 - \sqrt{n_k}\kappa)}.$$

Contudo, Kulinskaya *et al.* (2008) reconhecem que, apesar do modelo normal truncado ser um modelo plausível para o viés de publicação, a abordagem acima pode falhar, porque às vezes os investigadores repetem a experiência até conseguirem uma que origine um valor- $p$  inferior a 0.05, aumentando-lhes assim a probabilidade de publicação dos resultados. Afirmam também que

*A better model is obtained by adding an additional feature, the number of missing or absent studies. The user must specify the number of missing studies. By doing this, one gains control over the bias correction introduced in the meta-analysis. Adjusting the cutoff point would serve a similar purpose.*

Por isso, sugerem uma abordagem alternativa ao problema do viés de publicação, onde se deve considerar que o que está em jogo é na realidade um problema de censura e não um problema de truncatura, sendo necessário para o efeito ter-se em conta o número de estudos em falta.



## Capítulo 8

# Podemos Confiar numa Meta-Análise?

*Meta-Analysis: Statistical Alchemy for the 21st century.*

Feinstein (1995)

A necessidade de combinar, harmonizar ou sintetizar toda a informação que se tem vindo a acumular sobre determinados problemas tem catapultado a Meta-Análise nos últimos anos para as luzes da ribalta. Há, todavia, quem não atribua à Meta-Análise qualquer credibilidade e há quem vendo nela alguma utilidade, também nos procure alertar para os possíveis perigos que pode albergar. Neste capítulo procuraremos chamar a atenção para os principais problemas com que nos podemos deparar numa meta-análise, por forma a tentarmos evitá-los, ou pelo menos termos em conta os seus possíveis efeitos nas conclusões.

As primeiras críticas à Meta-Análise surgem logo na altura do seu “batismo” quando Eysenck, criticando a meta-análise de Glass sobre a eficácia da psicoterapia, chama a atenção para o problema de poder existir estudos selecionados de qualidade inferior. As famosas palavras de Eysenck “*garbage in-garbage out*” são na realidade um sério aviso para a necessidade de haver uma seleção criteriosa dos estudos, pois a

inclusão de estudos de qualidade duvidosa certamente se refletirá sobre as inferências e conclusões.

Um outro aspecto crucial a termos em conta, aquando da realização de uma meta-análise, é que os estudos selecionados sejam comparáveis entre si, isto é, meçam, pelo menos aproximadamente, o mesmo e não dependam, por exemplo, do tipo de delineamento experimental ou da existência de covariáveis diferentes. Por isso, quando em Feinstein (1995) se faz a analogia entre a Meta-Análise e a Alquimia, está-se a chamar a atenção para o problema de poder estarmos a combinar “alhos com bugalhos”, ou seja, a combinarmos estudos que não são combináveis devido às suas características específicas. De facto, a observação de uma heterogeneidade elevada entre os estudos tem um efeito pernicioso numa meta-análise, produz estimativas pouco precisas (para valores médios) e, por conseguinte, conclusões pouco fiáveis. Em Feinstein (1995) afirma-se a propósito que

*The idea of getting something for nothing, while simultaneously ignoring established scientific principles, produces an immediate analogy to the alchemy that existed before modern scientific chemistry. An advantage of alchemy was a principle that might be called the free lunch; the alchemists hoped to convert existing things into something better, such as changing base metals into gold. A scientific disadvantage of alchemy might be called the mixed salad principle. Before the reproducible precision of modern chemistry, the alchemists worked with substances that were heterogeneous, poorly identified mixtures.*

Por outro lado, a existência de viés de publicação é frequentemente considerada a causa da falibilidade de uma meta-análise. Efetivamente, de todos os problemas com que nos podemos deparar numa meta-análise, o viés de publicação é provavelmente o mais grave de todos, uma vez que quase certamente está presente e não se conhecerá o seu verdadeiro impacto.

Um exemplo citado na literatura de uma meta-análise que “falhou”, e com consequências potencialmente graves, é a de Yusuf *et al.* (1993) sobre a eficácia do uso de magnésio na redução do risco de mortalidade após um enfarte agudo do miocárdio. Antes de 1993 eram conhecidos alguns pequenos estudos, controlados e completamente aleatórios, sobre o efeito antiarrítmico do magnésio com resultados geralmente inconclusivos. Porém, quando esses estudos foram meta-analisados, o magnésio revelou ser um tratamento eficaz na redução do risco de mortalidade, razão pela qual passou a ser uma recomendação médica. Mas, em 1995 foram publicados os resultados do mega-estudo ISIS-4 (ISIS-4 *Collaborative Group*) que indicavam não haver qualquer redução no risco de mortalidade com o tratamento. Egger e Davey Smith (1995) avançaram como possível explicação para as conclusões contraditórias a existência de viés de publicação.

Uma área onde a Meta-Análise se tem mostrado ser bastante útil é em Medicina, cujo lema atual é “Uma Medicina baseada em evidência”. Porém, uma Medicina baseada em evidência requer que os investigadores tenham acesso a toda a informação disponível sobre os benefícios e os riscos de intervenções médicas, porque vidas humanas estão em jogo. A importância dada hoje em dia a meta-análises na área da Medicina é tal, que ultimamente os editores de algumas revistas de especialidade médica têm exigido a realização de meta-análises, como condição necessária para que os resultados de um novo estudo venham sequer a ser considerados para publicação. Por exemplo, no artigo “*Putting clinical trials into context*” de Young e Horton (2005) afirma-se que

*In recognition that journal editors have a key part to play in ensuring that published research is presented in a way that clearly illustrates why it is necessary and what impact a particular trial has on the existing state of knowledge, The Lancet has decided to update its policies in this area. From August, 2005, we will require authors of clinical trials submitted to The Lancet to include a clear summary of previous research findings, and to explain how their trials's findings*

*affect this summary. The relation between existing and new evidence should be illustrated by direct reference to an existing systematic review and meta-analysis. When a systematic review or meta-analysis does not exist, authors are encouraged to do their own. If this is not possible, authors should describe in a structured way the qualitative association between their research and previous findings.*

Por conseguinte, a existência de bibliotecas do tipo *Cochrane Library*, e de acesso gratuito, é de extrema importância para todas as áreas científicas que utilizam dados. É também importante que haja repositórios onde se possa ter acesso a resultados de estudos (ainda) não publicados, como forma de eliminar o viés de publicação.

Por outro lado, é igualmente importante que a informação disponibilizada ao público sobre os resultados de estudos seja total e não apenas parcial. Por exemplo, a indústria farmacêutica tem pecado neste capítulo, pois tem sido acusada de ocultar ao longo dos anos dados que descredibilizam a eficácia publicitada para alguns medicamentos. No artigo *“Despite vow, drug makers still withhold data”* de Alex Berenson, publicado a 31 de maio de 2005 no *The New York Times*, afirma-se que

*When the drug industry came under fire last summer for failing to disclose poor results from studies of antidepressants, major drug makers promised to provide more information about their research on new medicines. But nearly a year later, crucial facts about many clinical trials remain hidden, scientists independent of the companies say. [...] As a result, doctors and patients lack critical information about important drugs, academic researchers say, and the companies can hide negative trial results by refusing to publish studies, or by cherry-picking and highlighting the most favorable data from studies they do publish.*

Face ao exposto, impõe-se responder à questão

*Podemos confiar numa meta-análise?*

Apesar dos problemas mencionados que podem minar uma meta-análise, se houver um especial cuidado por parte dos meta-analistas no sentido de procurar evitá-los, ou pelo menos minimizar os seus efeitos, a resposta é sim. Existem muitos exemplos de meta-análises bem feitas que mostram o potencial da Meta-Análise como ferramenta para a construção do conhecimento científico.

É importante que quem realize uma meta-análise tenha a plena consciência do que efetivamente pode inferir a partir dos dados meta-analisados com alguma confiança. Daí que uma boa prática seja sempre indicar as limitações das meta-análises realizadas.



# Bibliografia

- [1] Abramson, H. (1944). Accidental mechanical suffocation in infants. *The Journal of Pediatrics*, **25**, 404-413.
- [2] Allen, I.E., Seaman, J.E. (2017). *Is Heterogeneity your Friend? Using sensitivity analyses to design improved studies*. Statistics Roundtable. Quality Progress, The official publication of ASQ.
- [3] Andrews, D.A., Zinger, I., Hoge, R.D., Bonta, J., Gendreau, P., Cullen, F.T. (1990). Does correctional treatment work?: a clinically relevant and psychologically informed meta-analysis. *Criminology*, **28**, 369-429.
- [4] Baujat, B., Mahé, C., Pignon, J.P., Hill, C. (2002). A graphical method for exploring heterogeneity in meta-analyses: Application to a meta-analysis of 65 trials. *Statistics in Medicine*, **21**(18), 2641-2652.
- [5] Begg, C.B., Mazumbar, M. (1994). Operating Characteristics of a Rank Correlation Test for Publication Bias. *Biometrics*, **50**(4), 1088-1101.
- [6] Behrens, W.V. (1929). Ein Beitrag zur Fehlerberchnung bei Wenigen Beobachtungen. *Landw. Jb.*, **68**, 807-837.
- [7] Biggerstaff, B.J., Tweedie, R.L. (1997). Incorporating variability in estimates of heterogeneity in random effects models in meta-analysis. *Statistics in Medicine*, **16**(7), 753-768.

- [8] Birnbaum, A. (1954). Combining independent tests of significance. *Journal of the American Statistical Association*, **49**, 559-575.
- [9] Bjelakovic, G., Nikolova, D., Gluud, L.L., Simonetti, R.G., Gluud, C. (2007). Mortality in Randomized Trials of Antioxidant Supplements for Primary and Secondary Prevention: Systematic Review and Meta-analysis. *Journal of the American Medical Association*, **297**(8), 842-857.
- [10] Borenstein, M., Hedges, L.V., Rothstein, H. (2009). *Introduction to Meta-Analysis*. Wiley, Chichester.
- [11] Box, G.E.P., Draper, N.R. (1987). *Empirical Model-Building Response Surfaces*. Wiley, New York.
- [12] Brillhante, M.F. (2013). Generalized  $p$  Values and Random  $p$  Values when the Alternative to Uniformity is a Mixture of a Beta(1,2) and Uniform. In Oliveira, P., *et al.* (eds), *Recent Developments in Modeling and Applications in Statistics*, Springer, 159-167.
- [13] Brillhante, M.F., Pestana, D., Rocha, J., Rocha, M.L, Velosa, S. (2011). *Inferência Estatística sobre a Localização usando a Escala*. Instituto Nacional de Estatística, Lisboa.
- [14] Brown, L., Li, X. (2005). Confidence intervals for two sample binomial distribution. *Journal of Statistical Planning and Inference*, **130**, 359-375.
- [15] Cochran, W.G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society (Supplement)*, **4**, 102-118.
- [16] Cochran, W.G. (1954). The combination of estimates from different experiments. *Biometrics*, **10**, 101-129.
- [17] Cohen, J. (1969). *Statistical Power Analysis for Behavioral Sciences*. Academic Press, New York.

- [18] Cooper, H., Hedges, L.V., Valentine, J.C. (2009). *The Handbook of Research Synthesis and Meta-Analysis* (2<sup>a</sup> ed.). Russell Sage Foundation, New York.
- [19] Copas, J. (1999). What Works?: Selectivity Models and Meta-Analysis. *Journal of the Royal Statistical Society, Series A*, **162**(1), 95-109.
- [20] Copas, J., Shi, J.Q. (2000). Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics*, **1**(3), 247-262.
- [21] Copas, J.B., Shi, J.Q. (2001). A sensitivity analysis for publication bias in systematic reviews. *Statistical Methods in Medical Research*, **10**, 251-265.
- [22] Cumming, G. (2012). *Understanding The New Statistics. Effect Sizes, Confidence Intervals, and Meta-Analysis*. Routledge, New York.
- [23] DerSimonian, R., Laird, N. (1986). Meta-Analysis in Clinical Trials. *Controlled Clinical Trials*, **7**(3), 177-188.
- [24] Duval, S., Tweedie, R. (2000a). A Nonparametric “Trim and Fill” Method of Accounting for Publication Bias in Meta-Analysis. *Biometrics*, **56**(2), 455-463.
- [25] Duval, S., Tweedie, R. (2000b). A Simple Funnel-Plot Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Journal of the American Statistical Association*, **95**(449), 89-98.
- [26] Egger, M., Davey Smith, G. (1995). Misleading meta-analysis: Lessons from “an effective, safe, simple” intervention that wasn’t. *British Medical Journal*, **310**, 752-754.
- [27] Egger, M., Smith, G.D., Schneider, M., Minder, C. (1997). Bias in Meta-Analysis Detected by a Simple Graphical Test. *British Medical Journal*, **315**(7109), 629-634.

- [28] Eysenck, H.J. (1978). An exercise in mega-silliness. *American Psychologist*, **33**(5), 517-519.
- [29] Feinstein, A.R. (1995). Meta-analysis: statistical alchemy for the 21st century. *Journal of Clinical Epidemiology*, **48**(1), 71-79.
- [30] Fergusson, D., Glass, K.C., Hutton, B., Shapiro, S. (2005). Randomized controlled trials of aprotinin in cardiac surgery: could clinical equipoise have stopped the bleeding? *Clinical Trials*, **2**(3), 218-232.
- [31] Fisher, R.A. (1932). *Statistical Methods for Research Workers* (4<sup>th</sup> ed.). Oliver and Boyd, London.
- [32] Fisher, R.A. (1939). Samples with Possibly Unequal Variances. *Annals of Eugenics*, **9**, 174-180.
- [33] Galbraith, R. (1988). Graphical display of estimates having differing standard errors. *Technometrics*, **30**(3), 271-281.
- [34] George, E.O. (1977). Combining independent one-sided and two-sided statistical tests – some theory and applications. Doctoral dissertation. University of Rochester.
- [35] Gilbert, R., Salanti, G., Harden, M., See, S. (2005). Infant sleeping position and the sudden infant death syndrome: systematic review of observational studies and historical review of recommendations from 1940 to 2002. *International Journal of Epidemiology*, **34**(4), 874-887.
- [36] Glass, G.V. (1976). Primary, Secondary, and Meta-Analysis of Research. *American Educational Research Association*, **5**(10), 3-8.
- [37] Glass, G.V., McGraw, B., Smith, M.L. (1981). *Meta-Analysis in Social Research*. CA: Sage, Beverly Hills.
- [38] Greenland, S., Robins, J.M. (1985). Estimation of a common effect parameter from sparse follow-up data. *Biometrics*, **41**, 55-68.

- [39] Harbord, R.M., Egger, M., Sterne, J.A. (2006). A modified test for small study effects in meta-analysis of controlled trials with binary endpoints. *Statistics in Medicine*, **25**(20), 3443-3457.
- [40] Hartung, J. (1999). A note on combining dependent tests of significance. *Biometrical Journal*, **41**, 849-855.
- [41] Hartung, J., Knapp, G. (2001). A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine*, **20**, 3875-3889.
- [42] Hartung, J., Knapp, G., Sinha, B.K. (2008). *Statistical Meta-Analysis with Applications*. Wiley, New Jersey.
- [43] Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, **6**, 107-128.
- [44] Hedges, L.V. (1983). A random effects model for effect sizes. *Psychological Bulletin*, **93**, 388-395.
- [45] Hedges, L.V., Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Academic Press, Boston.
- [46] Hedges, L.V., Vevea, J.L. (1998). Fixed-and-Random-Effects Models in Meta-Analysis. *Psychological Methods*, **3**(4), 486-504.
- [47] Higgins, J.P.T., Thompson, S.G. (2002). Quantifying Heterogeneity in Meta-Analysis. *Statistics in Medicine*, **21**(11), 1539-1558.
- [48] Hunter, J.E., Schmidt, F.L. (1990). *Methods of meta-analysis: Correcting error and bias in research synthesis*. Beverly Hills, CA: Sage.
- [49] Iyer, H.K., Patterson, P.L. (2002). *A Recipe for Constructing Generalized Pivotal Quantities and Generalized Confidence Intervals*. Technical Report 2002/10, Colorado State University, Department of Statistics.

- [50] Knapp, G., Hartung, J. (2003). Improved tests for random effects meta-regression with a single covariate. *Statistics in Medicine*, **22**, 2693-2710.
- [51] Kulinskaya, E., Morgenthaler, S., Staudte, R.G. (2008). *Meta-Analysis: A Guide to Calibrating and Combining Evidence*. Wiley, Chichester.
- [52] L'Abbé, K.A., Detsky, A.S., O'Rourke, K. (1987). Meta-analysis in clinical research. *Annals of Internal Medicine*, **107**(2), 224-233.
- [53] Legendre, A.M. (1805). *Nouvelles méthodes pour la détermination des orbites des comètes*. F. Didot.
- [54] Light, R.J., Pillemer, D.B. (1984). *Summing up: The Science of Reviewing Research*. MA: Harvard University Press, Cambridge.
- [55] Makambi, K.H. (2003). Weighted inverse chi-square method for correlated significance tests. *Journal of Applied Statistics*, **30**, 225-234.
- [56] Mantel, N., Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.*, **22**, 719-748.
- [57] Orwin, R.G. (1983). A fail-safe N for effect size in Meta-Analysis. *Journal of Educational Statistics*, **8**(2), 157-159.
- [58] Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *The British Medical Journal*, **2**(2288), 1243-1246.
- [59] Peters, J.L., Sutton, A.J, Jones, D.R, Abrams, K.R., Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association* **295**, 676-680.
- [60] Rosenberg, M.S. (2005). The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*, **59**, 464-468.

- [61] Rosenthal, R. (1979). The “file-drawer problem” and tolerance for null results. *Psychological Bulletin*, **85**, 185-193.
- [62] Rücker, G., Schwarzer, G., Carpenter, J.R. (2008). Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine*, **27**(5), 746-763.
- [63] Schwarzer, G., Antes, G., Schumacher, M. (2007). A test for publication bias in meta-analysis with sparse binary data. *Statistics in Medicine*, **26**, 721-733.
- [64] Schwarzer, G., Carpenter, J.R., Rücker, G. (2015). *Meta-Analysis with R*. Springer, New York.
- [65] Semblano, P. (2016). *Estudo Crítico dos Valores de Prova*. Tese de Doutoramento, Faculdade de Ciências da Universidade de Lisboa.
- [66] Sequeira, F. (2009). *Meta-Análise: Harmonização de Testes usando os Valores de Prova*. Tese de Doutoramento, Faculdade de Ciências da Universidade de Lisboa.
- [67] Shoemaker, P.J., Tankard, J.W., Lasorsa, D.L. (2003). *How to Build Social Science Theories*. Sage, Thousand Oaks, CA.
- [68] Sidik, J., Jonkman, J.N. (2005). Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, **54**, 367-384.
- [69] Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A., Williams, R.M., Jr. (1949). *The American Soldier, Vol. I. Adjustment during Army Life*. NJ: Princeton University Press, Princeton.
- [70] Stigler, S.M. (1986). *The History of Statistics. The Measurement of Uncertainty before 1990*. Harvard University Press, Cambridge.
- [71] Sweeting, M.J., Sutton, A.J., Lambert, P.C. (2004). What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Statistics in Medicine*, **23**, 1351-1375.

- [72] Thompson, S.G., Sharp, S.J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, **18**, 2693-2708.
- [73] Tippett, L.H.C. (1931). *The Methods of Statistics*. Williams & Norgate, London.
- [74] Tsui, K., Weerahandi, S. (1989). Generalized  $p$ -values in significance testing of hypotheses in the presence of nuisance parameters. *Journal of the American Statistical Association*, **84**, 602-607.
- [75] Van Howe, R.S. (1999). Circumcision and HIV infection: review of the literature and meta-analysis. *International Journal of STD and AIDS*, **10**, 8-16.
- [76] Viechtbauer, W. (2005). Bias and Efficiency of Meta-Analytic Variance Estimators in Random-Effects Model. *Journal of Educational and Behavioral Statistics*, **3**(3), 261-293.
- [77] Viechtbauer, W. (2007). Confidence Intervals for the Amount of Heterogeneity in Meta-Analysis. *Statistics in Medicine*, **26**(1), 37-52.
- [78] Viechtbauer, W. (2010). Conducting Meta-Analysis in R with the metafor Package. *Journal of Statistical Software*, **36**(3), 1-48.
- [79] Weerahandi, S. (1993). Generalized Confidence Intervals. *Journal of the American Statistical Association*, **88**(432), 899-905.
- [80] Weerahandi, S. (1995). *Exact Statistical Methods for Data Analysis*. Springer, New York.
- [81] Whitehead, J.T., Lab, S.P. (1989). A meta-analysis of juvenile correctional treatment. *Journal of Research in Crime and Delinquency*, **26**, 276-295.
- [82] Wilkinson, B. (1951). A statistical consideration in psychological research. *Psychological Bulletin*, **48**, 156-158.

- [83] Woolley, P.V. (1945). Mechanical suffocation during infancy. A comment on its relation to the total problem of sudden death. *The Journal of Pediatrics*, **26**, 572-575.
- [84] Yates, F., Cochran, W. (1938). The analysis of group experiments. *Journal of Agricultural Science*, **28**, 556-580.
- [85] Young, C., Horton, R. (2005). Putting clinical trials into context. *The Lancet*, **366**(9480), 107-108.
- [86] Yusuf, S., Peto, R., Lewis, J., Collins, R., Sleight, P. (1985). Beta blockade during after myocardial infarction: an overview of the randomized trials. *Prog. Cardiovasc. Dis.*, **27**, 335-371.
- [87] Yusuf, S., Koon, T., Woods, K. (1993). Intravenous magnesium in acute myocardial infarction: an effective, safe, simple and inexpensive intervention. *Circulation*, **87**, 2043-2046.



# Índice Remissivo

- ajustamento Hartung-Knapp, 28
- análise de subgrupos, 23, 69
  - modelo 1, 70
  - modelo 2, 72
  - modelo 3, 74
- coeficiente de determinação, 81
- efeito síntese, 16
  - modelo de efeitos aleatórios, 21
  - modelo de efeitos fixos, 18
- escala calibrada, 124
- estatística  $Q$  de Cochran, 19, 129
  - generalizada, 29
- evidência, 124
- fator de correção de Sweeting, 15
- função inferência chave, 124
  - modelo  $t$  de Student, 126
  - modelo do qui-quadrado, 130
  - risco relativo, 133
- gráfico de Baujat, 64, 66
- gráfico de bolhas, 86, 87
- gráfico de L'Abbé, 67, 68
- gráfico floresta, 59, 60, 62, 78, 91, 92
- gráfico funil, 94–96, 114
- gráfico radial, 61, 63
- método de Mantel-Haenszel, 22
- métodos de combinação de testes, 49
  - de Fisher, 50
  - de Stouffer, 50
  - de Tippett, 49
  - de Wilkinson, 49
  - Logit, 50
- Meta-Análise, 3, 7
  - modelo de efeitos aleatórios, 20, 131
  - modelo de efeitos fixos, 17, 129
  - modelo de efeitos mistos, 23, 83
- meta-análise cumulativa, 87, 91, 92
- meta-regressão, 23, 32, 81
- número de estudos na gaveta, 103
  - método de Orwin, 104
  - método de Rosenberg, 105
  - método de Rosenthal, 103
- parâmetro de heterogeneidade, 23

- estimador de DerSimonian-Laird, 24  
 estimador de Hedges, 24  
 estimador de Hunter-Schmidt, 25  
 estimador de máxima verosimilhança, 26  
 estimador de máxima verosimilhança restrita, 26  
 estimador de Sidik-Jonkman, 25  
 problema de Behrens-Fisher, 55  
 tamanho do efeito, 7  
   *odds ratio*, 14  
   *odds ratio* de Peto, 22  
 coeficiente de associação  $\phi$ , 44  
 coeficiente de correlação, 12  
 diferença de arcos senos, 15  
 diferença de médias, 9  
 diferença de médias estandarizada, 9  
 diferença de proporções, 14  
 medida de Cohen, 10  
 medida de Glass, 11  
 medida de Hedges, 10  
 risco relativo, 13  
 transformação estabilizadora da variância, 125  
 valor-*p*  
   aleatório, 57  
   generalizado, 52  
   problemático, 52  
 variável  
   de teste generalizada, 54  
   fulcral generalizada, 54  
   moderadora, 82  
   viés de publicação, 94  
   método apara e preenche, 108  
   modelo de seleção de Copas, 113, 119  
   teste de Begg, 97  
   teste de Egger, 98  
   teste de Thompson, 99

Toda a informação disponível em  
[spe2017.iscte-iul.pt](http://spe2017.iscte-iul.pt)

Pode aceder à página do  
XXIII Congresso através do  
QR Code ao lado



**ISCTE IUL**  
Instituto Universitário de Lisboa

**BRU IUL**  
Business Research Unit  
ISCTE-INSTITUTO UNIVERSITÁRIO DE LISBOA

 Sociedade  
Portuguesa de  
Estatística



MUSEU  
DA ÁGUA



**INDEG ISCTE**  
Executive Education

**ISCTE Business School**  
INSTITUTO UNIVERSITÁRIO DE LISBOA

 **EDIÇÕES SÍLABO, Lda.**  
Publicamos conhecimento

