Understanding and Interpreting Systematic Review and Meta-Analysis Results

3

Cristiano Susin, Alex Nogueira Haas,
and Cassiano Kuchenbecker Rösing

Core Message

› Here, we provide an overview of the methods used to combine the results of several studies. Specifically, we discuss the application and interpretation of meta-analytic methods.

3.1 Introduction

Systematic reviews and meta-analysis have become the *de facto* gold standard in evidence-based health care. Nevertheless, most health care providers do not have a clear understanding of how systematic reviews and meta-analyses are conducted and how to interpret their results. This fact greatly hinders the application and dissemination of evidence that could have an important impact on the population health. Frequently, evidence from systematic reviews reaches mainstream health care only when they are adopted or endorsed by professional associations/societies and governmental bodies. In an evidence-based era, it is interesting to note that some of the journals with higher impact in medicine and dentistry are still based on narrative reviews written by invited authorities. This underlines the fact that most health care providers have trouble understanding one of the most important sources of evidence. In this context, the aim of the present chapter is to provide an overview of the methods used to combine the results of several studies. We will focus on the application and interpretation of meta-analytic methods.

First, we would like to acknowledge that systematic reviews and meta-analyses are not easy topics for most readers. This is especially true for health care providers who focused most of their efforts on learning biology-related subjects instead of mathematical concepts. As a consequence, most researchers do not like statistics-related topics, most professionals do not use it in their appraisal of the medical literature, and majority of the students are not willing to learn it. This is an unfortunate truth with known causes and consequences. Our approach to try to explain these concepts will be as intuitive as possible and we will try to avoid the classic mathematical approach whenever possible.

It is beyond the scope of this chapter to review all steps of a systematic review. Thus, we will assume that the necessary steps to carry out a systematic review have been fulfilled (identification of the need for the review, preparation of a review protocol, identification and selection of the studies, quality assessment and data collection, etc.), and we will focus on the analysis and presentation of the results.

C. Susin (✉)
Laboratory for Applied Periodontal and Craniofacial Regeneration, Department of Periodontics and Oral Biology, Medical College of Georgia School of Dentistry,
1459, Laney Walker Blrd,
Augusta, GA 30912, USA
e-mail: csusin@mcg.edu/csusin@me.com

A.N Haas
C.K Rösing
School of Dentistry, Federal University of Rio Grande do Sul,
Rua Ramiro Barcelos, 2492, 90035-003,
Porto Alegre, RS, Brazil
e-mail: alexnhaas@gmail.com
e-mail: ckrosing@hotmail.com

F. Chiappelli et al. (eds.), *Evidence-Based Practice: Toward Optimizing Clinical Outcomes*,
DOI: 10.1007/978-3-642-05025-1_3, © Springer-Verlag Berlin Heidelberg 2010

3.1.1 Example: Studies Characteristics and Descriptive Results

To illustrate this chapter, let us imagine that we are conducting a systematic review about the effect of a new antiviral therapy for recurrent herpes labialis. For simplicity, our main outcome will be reduction in the number of days with pain, i.e., a continuous outcome. Also for simplicity, let us assume that all studies used placebo as the control group.

Most systematic reviews use tables to present the methodological characteristics and outcomes of the selected studies. For example, the success of secondary root canal treatment was investigated in a systematic review published by Ng et al. [7]. The search strategy identified 40 studies, of which 17 were included in the analyses. Table 3.1 describes the methodological characteristics and outcomes of the 17 included studies. The methodological characteristics of the studies facilitate the reader interpretation of the meta-analysis results. Other study characteristics frequently reported are sample characteristics, randomization method, blindness of patients, therapists and examiner, follow-up time, and dropout rate.

In addition to the methodological characteristics, most systematic reviews present the original results in descriptive forms using tables and graphs. Table 3.2 combines study characteristics and results for our systematic review describing 12 studies that tested the effect of our new antiviral therapy. The table presents the year of publication, total sample size, source of funding, sample size in each experimental group (n), estimate of the intervention effect (mean), an estimate of the intervention variability (standard deviation – SD), and the *p*-value.

The overall trend of the studies is used to suggest if a given intervention is better than the standard treatment or no intervention when only descriptive tables are used to present results. This approach is very intuitive and does not need any statistical expertise to be conducted. However, as you will later see in this chapter, it can be misleading for several reasons. An overall assessment of Table 3.2 indicates that between 1997 and 2002, mostly small studies were conducted, with large studies being published only in the last 4 years. This finding is consistent with most new therapy studies since large, costly, time-consuming studies are only conducted after some evidence of positive effect and safety is available. A closer look at the results of the studies shows that within small studies the results are very inconsistent, with few studies showing large positive or negative effects for the therapy when compared to placebo. In contrast, large studies do not show major differences between experimental groups. As expected, variability is larger in smaller studies due to the sample size effect on standard deviation estimates. Only the first two studies reached statistical significance, and in three other studies somewhat borderline results were found ($p \sim 0.10$).

3.2 Main Results: Overall Estimates of Effect

The treatment effect could be estimated by calculating an overall mean of the results simply by summing up the individual results dividing by the number of studies. This approach, although very intuitive, would not take into consideration the studies characteristics, with studies contributing equally to the overall estimate. Looking at the estimates in Table 3.2, it is obvious that some studies have more precise estimates than others. Factors that may affect the precision of the estimates are various, including sample characteristics, sample size, measurement precision, and reliability. In the meta-analysis framework, sample size is often the most important factor to be taken into consideration. Thus, overall estimates should take into account the sample size with larger studies contributing more than small studies. Mathematically, this can be accomplished by multiplying each study estimate by the sample size or, in other words, by weighting the estimates according to sample size. The sum of the estimates can then be divided by the total sample size. Table 3.3 shows the weight of each study of our example according to sample size. Using this approach, the mean reduction in days with pain would be 2.8 days for the treatment group and 3.1 days for the control group. Thus, the placebo treatment reduced in approximately 0.3 days patients' symptoms.

In essence, this is what is done in a meta-analysis to take into consideration the contribution of each study. A similar strategy can be used to account not only for the sample size, but also for the variability in the estimates of the original studies. The overall weighted estimate is calculated multiplying each study estimate by the inverse of the square of the standard error (inverse-variance weighting method), which is highly associated with the sample size of the study. Using this

Table 3.1 Methodological characteristics and outcomes of included studies by Ng et al. [7]

Study authors	Operator	Design	Recall rate (%)	Sample size	Unit of measure	Assessment of success	Radiographic criteria success	≥4 years after treatment	Calibration	Reliability test	Statistical analysis
Grahnen	UG	R	64	502	Ro	C&R	S	✓	–	–	–
Engstrom	UG	R	72	153	T	C&R	L	✓	–	–	X^2
Selden	Sp	R	20	52	T	C&R	L	–	–	–	X^2
Bergneholtz	UG	R	66	556	Ro	C&R	S	–	–	✓	–
Pekruhn	Sp	R	81	36	T	C&R	S	–	–	–	X^2
Molven	UG	R	50	226	Ro	Ra	S	✓	✓	✓	X^2
Allen	–	R	53	315	T	C&R	S	–	✓	–	X^2
Sjogren	UG	R	46	267	Ro	C&R	S	✓	✓	✓	LR
Van Nieuwenhuysen	–	R	–	612	Ro	C&R	S	–	✓	✓	X^2
Friedman	Sp	C	78	128	T	C&R	S	–	–	–	X^2
Danin	Sp	RCT	100	18	T	Ra	L	–	✓	✓	X^2
Sundqvist	UG	C	93	50	T	C&R	S	✓	–	–	X^2
Chugal	PG	R	75	85	Ro	Ra	S	✓	–	–	LR
Hoskinson	Sp	R	78	76	Ro	C&R	S	✓	✓	✓	GEE
Farzaneh	Sp	C	22	103	T	C&R	S	–	✓	✓	LR
Gorni	PG	C	94	452	T	C&R	S	✓	✓	✓	M–W
Çaliskan	Sp	R	96	86	T	C&R	S	–	✓	✓	X^2

"–" missing information; *UG* undergraduate students; *PG* postgraduate students; *Sp* specialist endodontists; *R* retrospective study; *C* prospective cohort study; *RCT* randomized controlled trial; *T* teeth; *Ro* root; *C&R* combined clinical and radiographic examination; *Ra* radiographic examination only; *S* strict criteria; *L* loose criteria; *LR* single level logistic regression; *GEE* generalized estimating equations; X^2 chi-square test; *M-W* Mann-whitney U-test

Table 3.2 Description of study characteristics and original results

Year of publication	Sample size	Source of funding	Treatment			Control			*p*-value
			n	Mean	SD	*n*	Mean	SD	
1997	42	Private	21	2.0	1.7	21	3.9	2.1	0.003
1998	31	Private	16	1.8	2.4	15	3.8	2.7	0.04
1998	44	Private	22	2.1	2.6	22	3.5	2.8	0.09
1999	33	Public	18	3.3	2.7	15	1.8	2.4	0.11
2001	30	Private	14	3	3.2	16	2.1	2.9	0.43
2001	29	Private	13	2.1	2.9	16	3.2	3.2	0.35
2002	27	Public	13	2.9	2.9	14	2.5	2.5	0.70
2002	31	Private	15	2.5	2.5	16	3	2.9	0.61
2005	190	Public	96	2.9	2.1	94	3.1	2.5	0.55
2007	80	Public	39	3.5	2.6	41	3.1	2.2	0.46
2007	145	Public	73	3.2	2.4	72	2.9	1.9	0.41
2008	394	Public	198	2.8	1.7	196	3.1	1.8	0.09

Table 3.3 Study weights according to sample size and inverse-variance methods

Year of publication	Weight based on sample size (%)	Weight based on inverse-variance method (%)
1997	3.9	4.6
1998	2.9	1.9
1998	4.1	2.4
1999	3.1	2.0
2001	2.8	1.3
2001	2.7	1.2
2002	2.5	1.5
2002	2.9	1.7
2005	17.7	14.2
2007	7.4	5.5
2007	13.5	12.4
2008	36.6	51.3
Total	100	100

approach, the weighted mean difference (WMD) between treatments is 0.26 mm in favor of the new antiviral therapy. Table 3.3 also shows the weight attributed to each study according to the inverse-variance method. It is clear that the study published in 2008 dominates the overall estimate not only because it has the largest sample size, but also because it has the greatest precision (smaller standard deviations and confidence intervals). Studies with lower variability receive greater weight and therefore have greater influence in the estimate.

Tables and graphs are popular ways of presenting the results of a meta-analysis. Table 3.4 presents the WMD and the 95% confidence interval for each study. The weighted mean provides an estimate and direction of the effect, and the confidence interval provides an assessment of the variability of the estimates. Confidence intervals also indicate the significance of the results and when it does not include zero (or 1 when the results are presented in odds ratio), the weighted mean is statistically significant.

3.3 Forest Plots

Figure 3.1 is a Forest plot of the results and has essentially the same information presented in Table 3.4. Studies are identified by their year of publication and sample size on the left side of the graph. The WMDs are presented in a graphical form with point estimates being presented as dots or short vertical lines and confidence intervals as horizontal lines. The size of the plotting symbol for the estimate is proportional to the weight of each study in the meta-analysis. The actual estimates are also presented on the right side together with the weight of the study. The overall estimate and

Table 3.4 Meta-analysis result using the inverse-variance method

Year	Sample size	Weighted mean difference	95% CI		Weight (%)
			Lower	Upper	
1997	42	−1.9	−3.056	−0.744	4.6
1998	31	−2	−3.803	−0.197	1.89
1999	44	−1.4	−2.997	0.197	2.41
1998	33	1.5	−0.241	3.241	2.03
2001	30	0.9	−1.297	3.097	1.27
2001	29	−1.1	−3.323	1.123	1.24
2002	27	0.4	−1.649	2.449	1.46
2002	31	−0.5	−2.403	1.403	1.7
2005	190	−0.2	−0.857	0.457	14.21
2008	80	0.4	−0.658	1.458	5.48
2006	145	0.3	−0.404	1.004	12.38
2007	394	−0.3	−0.646	0.046	51.33
Pooled weighted mean difference		−0.257	−0.504	−0.009	100
Significance test of weighted mean difference = 0		$p = 0.042$			

confidence interval are marked by a diamond. A dotted vertical line is used to present the overall estimate.

Since Table 3.4 and Fig. 3.1 have the same information, most publications present only the Forest plot. Looking at Fig. 3.1, it is easier to observe that the first two studies had a significant large effect in favor of the new therapy since both estimates are on the left side and the confidence interval does not include zero. No clear tendency is seen in the next six studies with half of them favoring therapy and the other half favoring control. It is important to note that the confidence intervals include zero for all studies. The overall WMD estimate is clearly dominated by the last study.

The last information in Fig. 3.1 is the I-square (I^2) statistic. The I^2 statistic represents the percentage of heterogeneity that can be attributed to variability between studies. The I^2 statistic varies between 0 and 100% and can be interpreted as follows: low heterogeneity for <50%, moderate heterogeneity for ≥50 – <75%, and high heterogeneity for ≥75%. In this example, the I^2 statistic is approximately 53%, which indicates moderate heterogeneity. This finding can be explained by the inconsistent results of the small studies published between 1997 and 2002. The I^2 statistic is statistically significant with a p-value of 0.016, further indicating that there is heterogeneity in the results. The only information that the Forest plot does not present is the p-value for the overall estimate ($p = 0.042$).

3.4 Exploring Heterogeneity

To further explore heterogeneity, let us try to look into the sample size effect. We stratified the studies into small and large sample sizes. Figure 3.2 presents the Forest plot with estimates for each stratum. Small studies showed a significant effect in favor of the therapy with antiviral treatment reducing the number of days in pain in 0.8 days ($p = 0.01$). In contrast, no significant effect was observed in large studies since the confidence interval includes zero ($p = 0.29$). An overall test for heterogeneity between small and large studies is significant ($p = 0.05$). It is interesting to notice that the I^2 statistic for the small sample size shows moderate heterogeneity (56.8%, $p = 0.02$) indicating that other factors may further explain these results. We will address this finding later on.

Let us try to explore the heterogeneity of the data even more. Figure 3.3 is the Forest plot using a fixed-effect model stratified by funding source: public or private. For public-funded studies, the WMD is 0.10

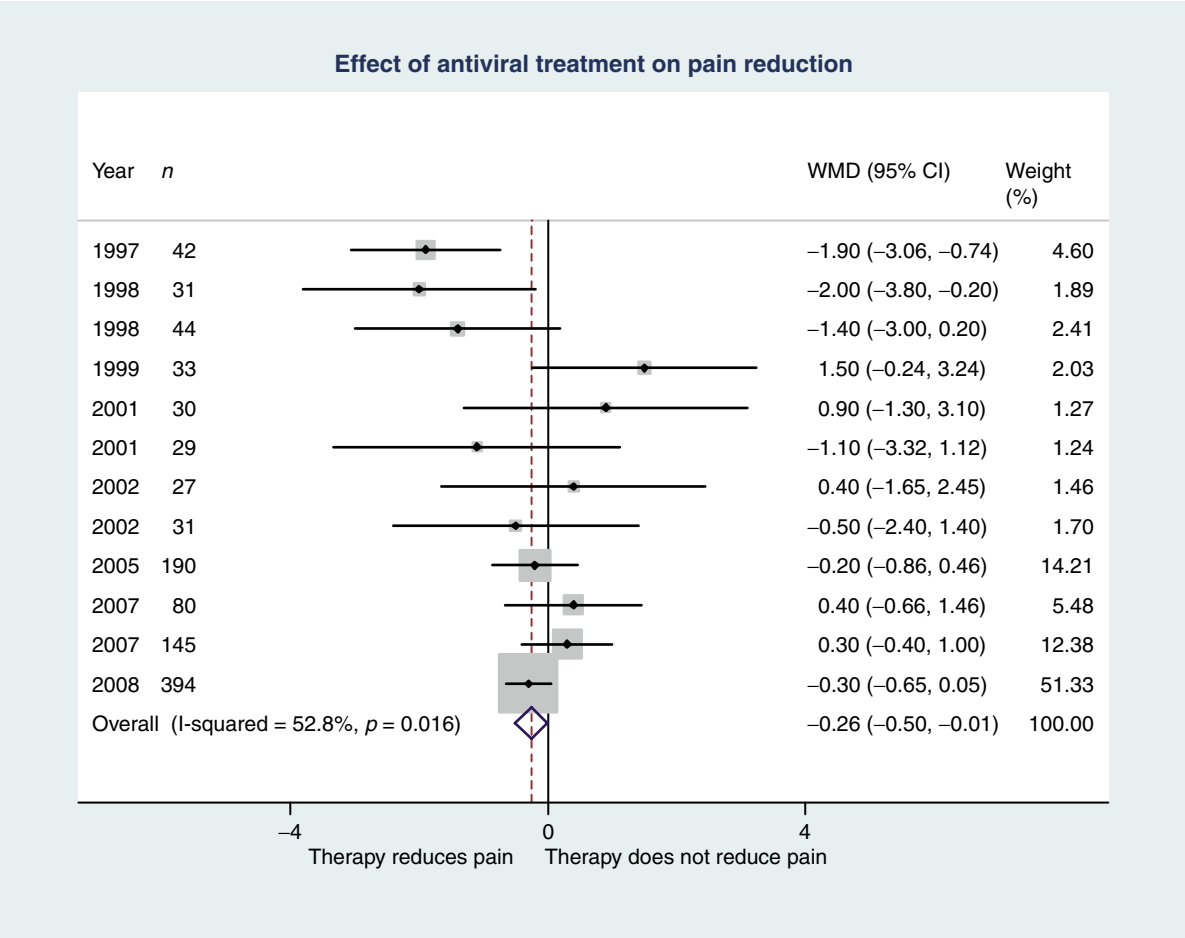

Fig. 3.1 Forest plot showing effect estimates and confidence intervals for individual studies and meta-analysis (fixed-effect model)

($p=0.46$, not reported in the Forest plot) in favor of the antiviral therapy, whereas for private-funded studies the antiviral therapy reduces pain, in average, in 1.29 days ($p<0.001$, not reported in the Forest plot). The heterogeneity test between groups is highly significant, also indicating that funding is an important source of variability.

3.5 Fixed-Effects vs. Random-Effects

So far we have found two possible sources of heterogeneity indicating that these studies may have different characteristics. We have used what is called a fixed-effect model to combine studies in a meta-analysis. When heterogeneity between studies exists, a different approach called random-effects model should be used. The fixed-effect model assumes that the meta-analysis overall estimate represents the same underlying effect and that differences between studies are due to sampling error, i.e., individual studies have the same single effect. The random-effects model includes an estimate of between-study variability assuming that the meta-analysis overall estimate is the mean effect around which individual studies have a normal distribution. In other words, random-effects models assume that the intervention is not the only explanation for the overall estimate allowing for other factors (such as study design, sample characteristics, and treatment differences) to partly explain the results.

In practice, random-effects models yield more conservative estimates with lower p-values and larger confidence intervals than fixed-effect models. Disparities in the overall WMD between treatments can also be seen due to the fact that random-effects models give greater weight to smaller studies than fixed-effect

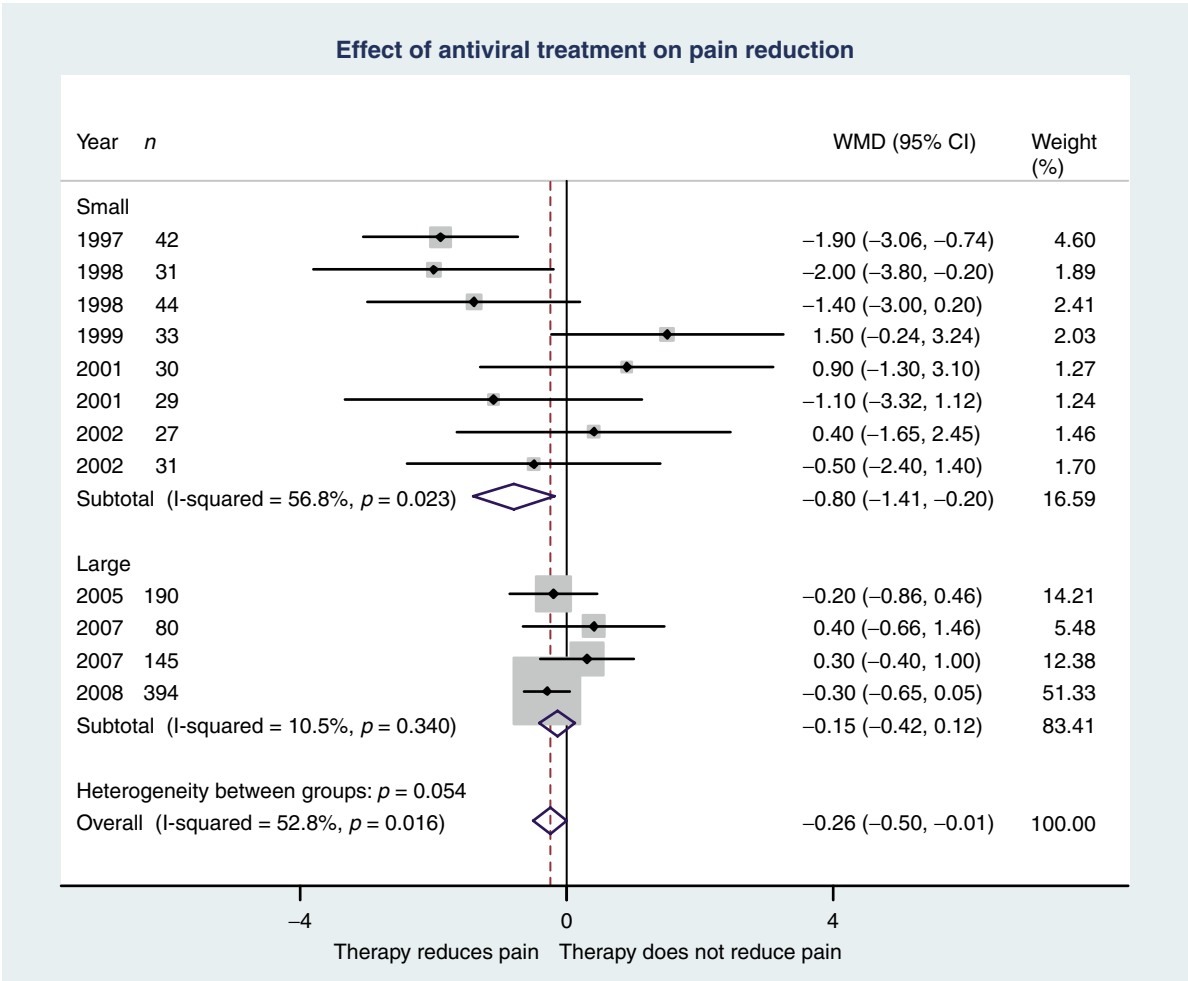

Fig. 3.2 Forest plot showing effect estimates and confidence intervals for individual studies and meta-analysis stratified by study sample size (fixed-effect model)

models (Table 3.5). As can be seen in Fig. 3.4, the overall WMD using the random-effects model is slightly different than the estimate using the fixed-effect model (0.30 vs. 0.26). However, the major difference can be seen in the confidence interval that now includes zero, and therefore, is associated with a non-significant *p*-value ($p = 0.22$). In other words, when the heterogeneity is taken into consideration in the calculation of the estimates, no overall significant differences were observed between treatments with regard to pain reduction. This is in contrast to the conclusion that could be drawn from the fixed-effect model.

Sometimes researchers present the Forest plot of the fixed-effect model and include the random-effects estimate for comparison (Fig. 3.5). This may be confusing for the inexperienced reader because different models may have opposite results. As a rule of thumb, if the I^2 statistic is moderate or high (>50%) and the *p*-value is significant ($p < 0.05$), a random-effects model should be used. In our example, a random-effects model is warranted.

3.6 Meta-Regression

Stratified analysis is an important tool for detecting heterogeneity, but has the same drawbacks of subgroup analysis in clinical trials. A better approach to evaluate between-group difference is to use a meta-regression

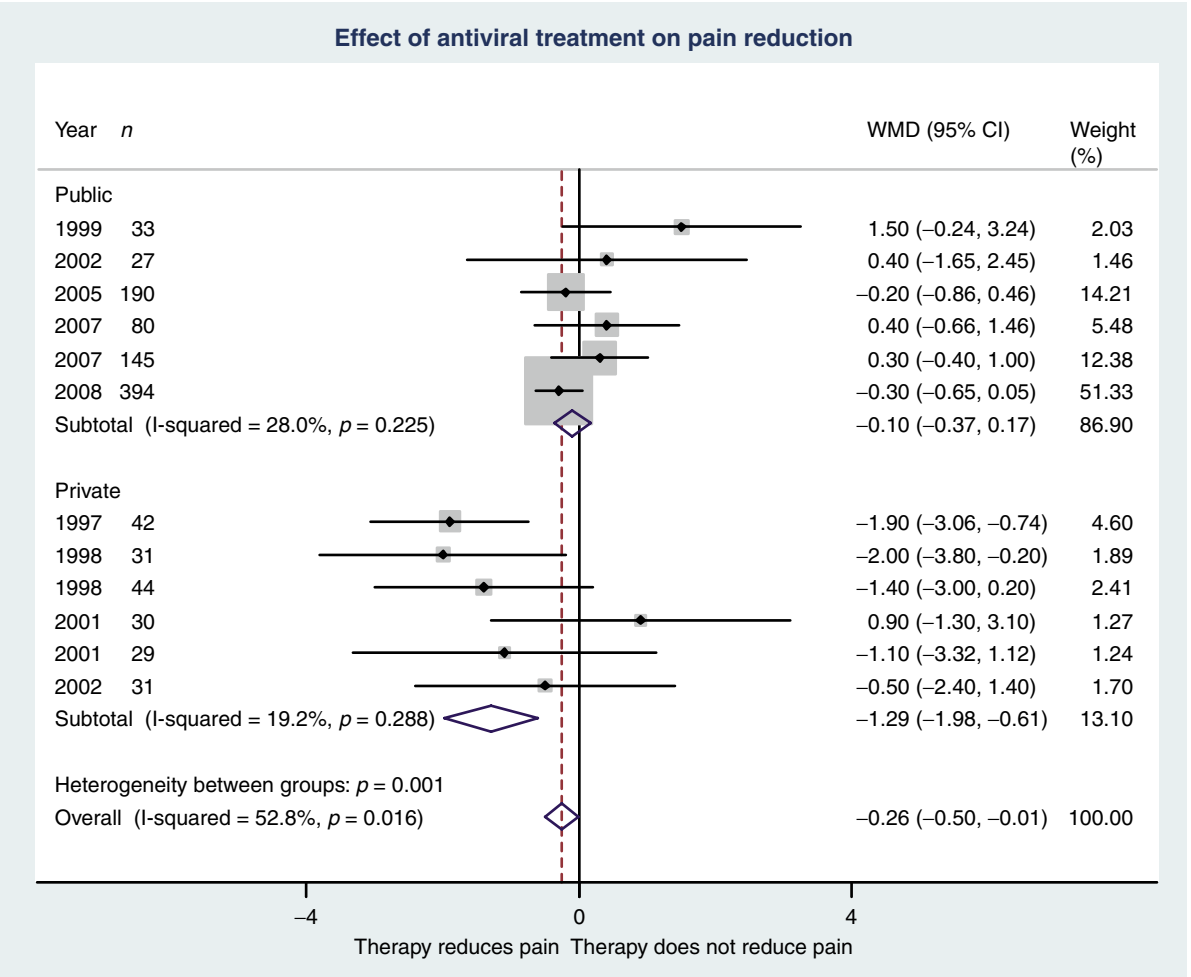

Fig. 3.3 Forest plot showing effect estimates and confidence intervals for individual studies and meta-analysis stratified by source of funding (fixed-effect model)

or meta-analysis regression. For those familiar with regression analysis, a meta-regression could be thought as a regression analysis performed at the study-level, i.e., using study-level data instead of individual-level data. Table 3.6 shows the result of the random-effects meta-regression using sample size and source of funding as explanatory variables. As observed before in the stratified analysis, both factors were significant sources of heterogeneity and funding seems to have the biggest impact on the effect estimates.

Similarly, Pavia et al. [8] conducted a meta-analysis of observational studies about the contribution of fruit and vegetable intakes to the occurrence of oral cancer. They included 16 studies and found that each portion of fruit consumed per day significantly reduced the risk of oral cancer by 49% (pooled odds ratio 0.51 95%CI 0.40–0.65). They found a significant heterogeneity across studies. To additionally explore heterogeneity, a meta-regression analysis was performed. This meta-regression analysis examined the effect of certain variables, such as quality score, type of cancers included, citrus fruit and green vegetable consumption, population studied (men, women, or both), and time interval for dietary recall, on the role of fruit or vegetable consumption in the risk of oral cancer. Table 3.7 shows the results for the meta-regression analysis, demonstrating that the lower risk of oral cancer associated with fruit consumption was significantly influenced by the type of fruit consumed and by the time interval of dietary recall.

Table 3.5 Study weights according to fixed- and random-effects methods

Year of publication	Fixed-effect model (%)	Random-effects model
1997	4.6	9.24
1998	1.9	5.16
1998	2.4	6.15
1999	2.0	5.43
2001	1.3	3.78
2001	1.2	3.71
2002	1.5	4.23
2002	1.7	4.75
2005	14.2	14.73
2007	5.5	10.14
2007	12.4	14.13
2008	51.3	18.54

3.7 Funnel Plots and Publication Bias

Another important issue in meta-analysis is publication bias. Publication bias arises from the fact that studies with statistically significant results are more likely to be reported by authors and accepted for publication. Consequently, there is a risk that meta-analysis estimates are positively biased. It should be remembered that some publication bias might be diminished during the search strategy, looking for grey literature (unpublished data). Graphical and statistical methods have been developed to assist in the identification of publication bias. The Funnel plot is the most commonly used graphic to investigate bias in meta-analysis. Funnel plots are scatterplots of each study treatment effect (i.e., WMD) by a measure of the study precision (i.e., standard error of the treatment effect). Figure 3.6 shows the Funnel plot of the present data. The WMD is plotted in the horizontal axis (*x*-axis) and the standard error

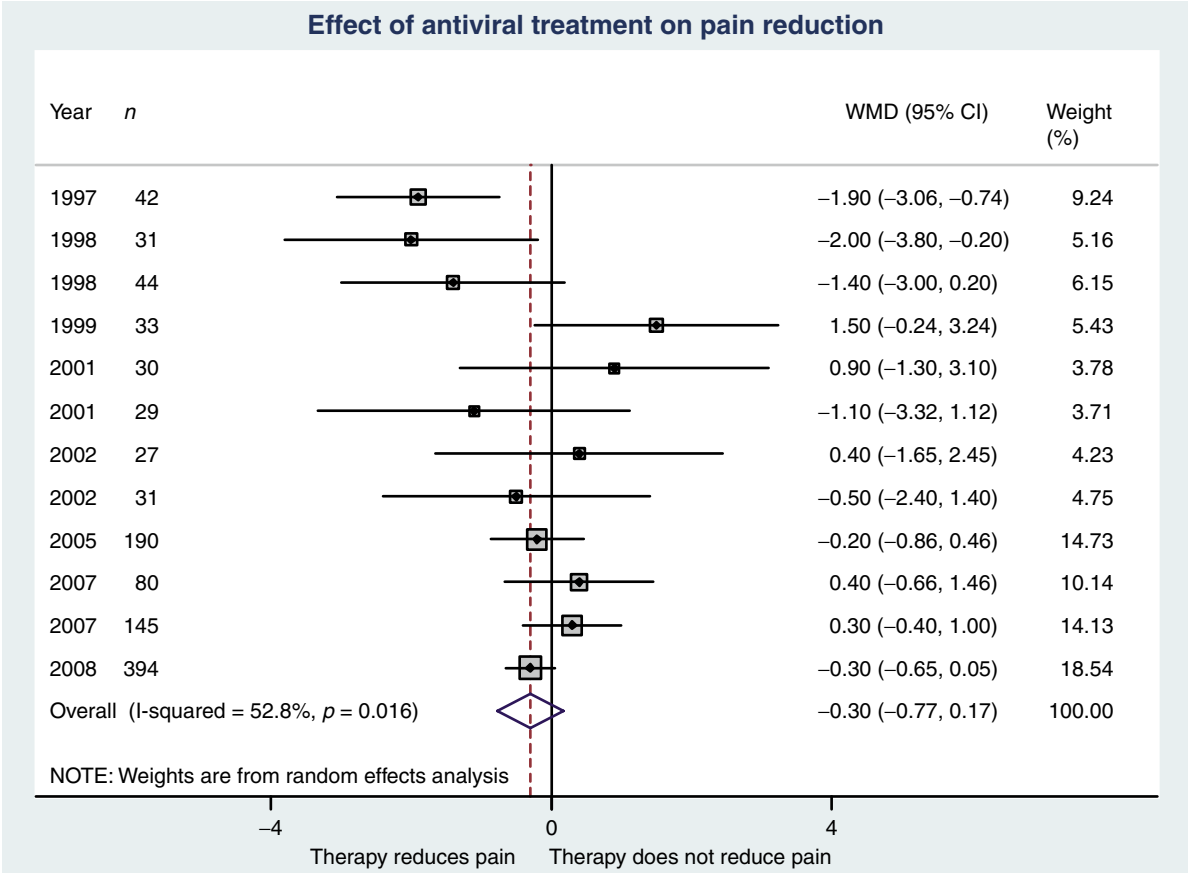

Fig. 3.4 Forest plot showing effect estimates and confidence intervals for individual studies and meta-analysis stratified using a random-effects model

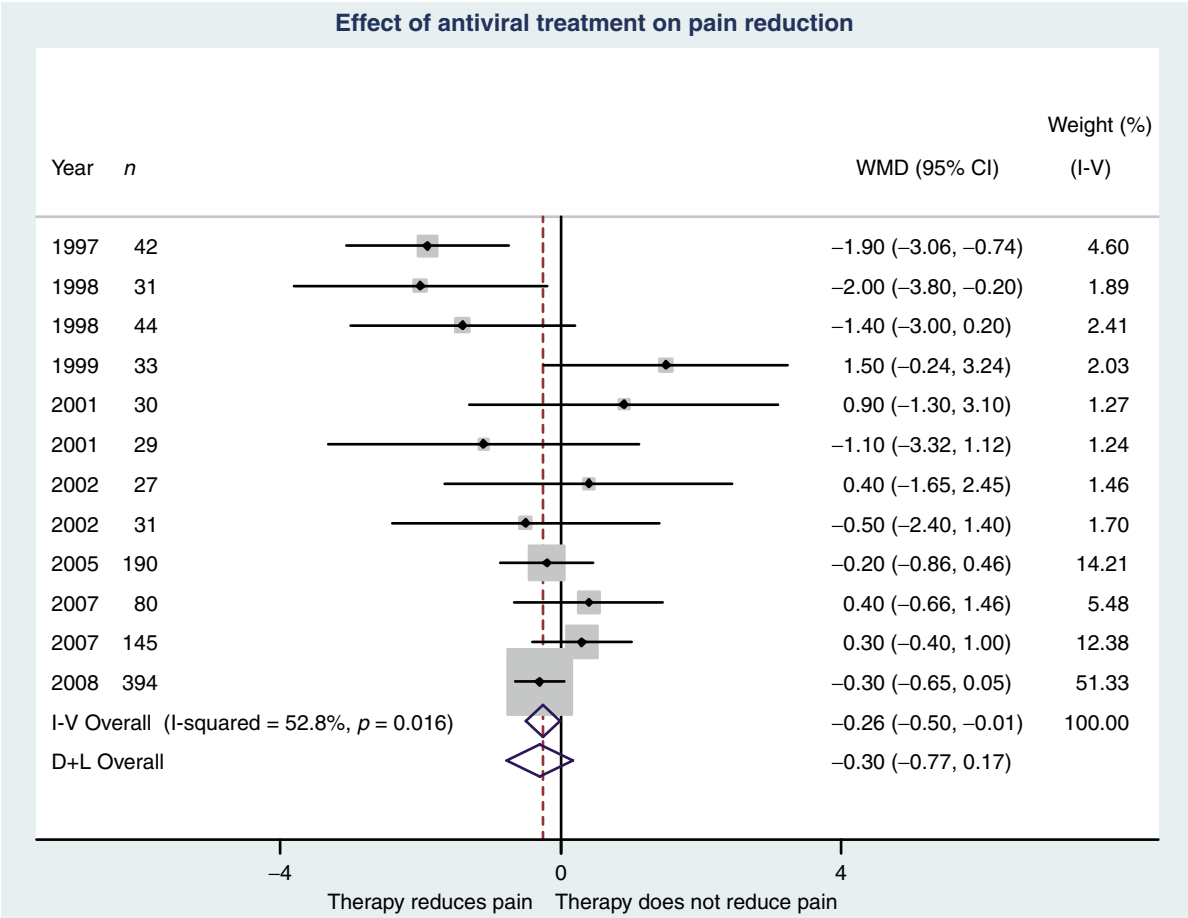

Fig. 3.5 Forest plot showing effect estimates and confidence intervals for individual studies and meta-analysis stratified using a fixed-effect and random-effects model

Table 3.6 Meta-regression analysis using study sample size and source of funding as explanatory variables

Variable	Coefficient	SE	*p*-value
Sample size	−0.66	0.28	0.04
Funding	−2.06	0.52	0.003

is plotted in the vertical axis (*y*-axis). Larger studies will often concentrate in the upper part of the Funnel plot because their standard error is generally smaller than smaller studies. For instance, the standard error for the three largest studies (sample sizes: 145, 190 and 394) ranged between 0.18 and 0.36, whereas that for three smallest (sample sizes: 27, 29 and 30) ranged between 1.05 and 1.13. A vertical solid line representing the overall WMD provides a reference for symmetry. A similar number of studies should be on both sides of this line. In our example, the same number of studies is plotted on the left and right sides of this reference line. The two doted diagonal lines represent the 95% confidence limits for the Funnel plot. In the absence of bias and heterogeneity, 95% of the studies should lie within the confidence limits lines. Two out of 12 (17%) studies are outside the confidence limits, further providing evidence of heterogeneity and perhaps bias.

A clear example of asymmetric Funnel plot using our data could be created by removing four studies with effects favoring the control treatment. In Fig. 3.7, it can be easily seen that small studies (generally shown on the bottom part of the plot) with negative results are missing, which may indicate that they were never reported or accepted for publication.

Formal approaches to test Funnel plot asymmetry have been proposed and implemented in statistical softwares. The Egger test uses a linear regression to draw a straight-line relationship between the WMD

Table 3.7 Meta-regression conducted by Pavia et al. [8]

Variable	Regression coefficient	SE	*p*
Fruit			
Only citrus fruit (no = 0; yes = 1)	−1.53	0.56	0.006
Dietary recall (lifelong = 0, 2 years = 1, 1 year = 2)	0.63	0.3	0.04
Population studied			
Men and women = 0	0	–	–
Only women = 1	−1.06	1.07	0.33
Only men = 2	0.01	0.56	0.99
Study quality score (low = 0, high = 1)	−0.32	0.54	0.56
Vegetables			
Only green vegetables	−0.23	0.43	0.59
Dietary recall (lifelong = 0, 2 years = 1, 1 year = 2)	−0.03	0.21	0.88
Population studied			
Men and women = 0	0	–	–
Only women = 1	1.14	0.73	0.12
Only men = 2	0.25	0.64	0.69
Study quality score (low = 0, high = 1)	0.23	0.47	0.63

and standard errors. When this regression line is plotted in the Funnel plot, it will appear as a vertical line as can be seen in Fig. 3.8. If asymmetry is present, the regression line will be plotted away from the vertical and the slope of the line will indicate the direction of bias (Fig. 3.9). The Egger's bias coefficient provides a measure of the asymmetry. The Egger's bias coefficient and its *p*-value for Fig. 3.7 are small (coefficient: −0.18, SE: 0.75, $p = 0.81$), indicating small chance of bias. On the other hand, the bias coefficient for Fig. 3.8 is larger with a *p*-value approaching significance (coefficient: −1.42, SE: 0.81, $p = 0.13$), indicating some evidence of bias. A negative bias coefficient indicates that the effect estimated from the smaller studies is smaller than the effect estimated from the larger studies. This may be interpreted as evidence that small sample size studies with nonsignificant results were not included in the meta-analysis. In general, bias tests for Funnel plots have lower power; thus, lower *p*-values should be carefully considered especially when less than ten studies are included in the analysis.

Even though we have focused on publication bias, Funnel plot asymmetry can be explained by other reasons such as poor study quality, true study heterogeneity, and chance. As discussed before, study quality can be addressed during study selection, and quality assessment and heterogeneity can be evaluated by stratified analysis and meta-regression.

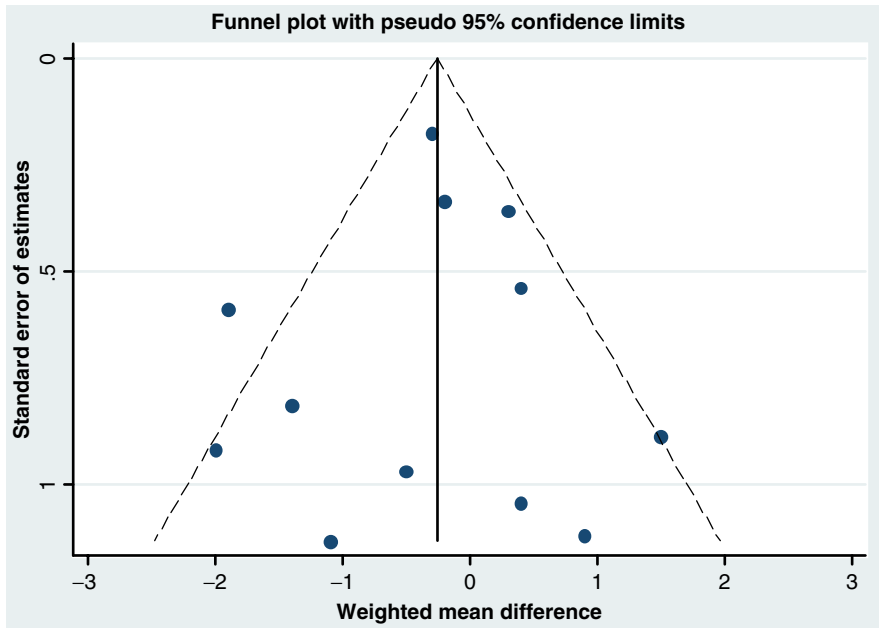

Fig. 3.6 Funnel plot of the weighted mean difference (WMD) against its standard error showing a symmetric distribution of studies

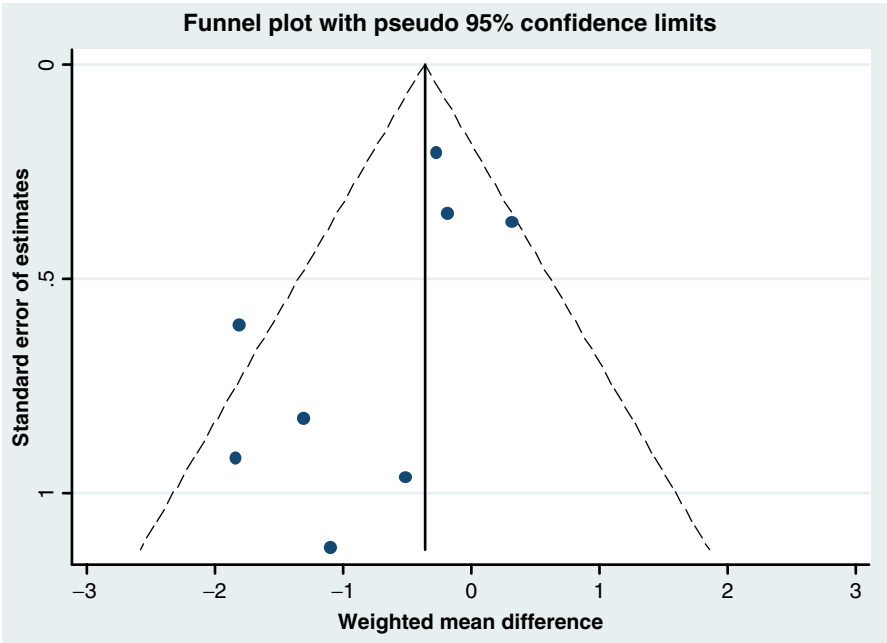

Fig. 3.7 Funnel plot of the WMD against its standard error showing an asymmetric distribution of studies (four studies with effects favoring the control treatment were removed)

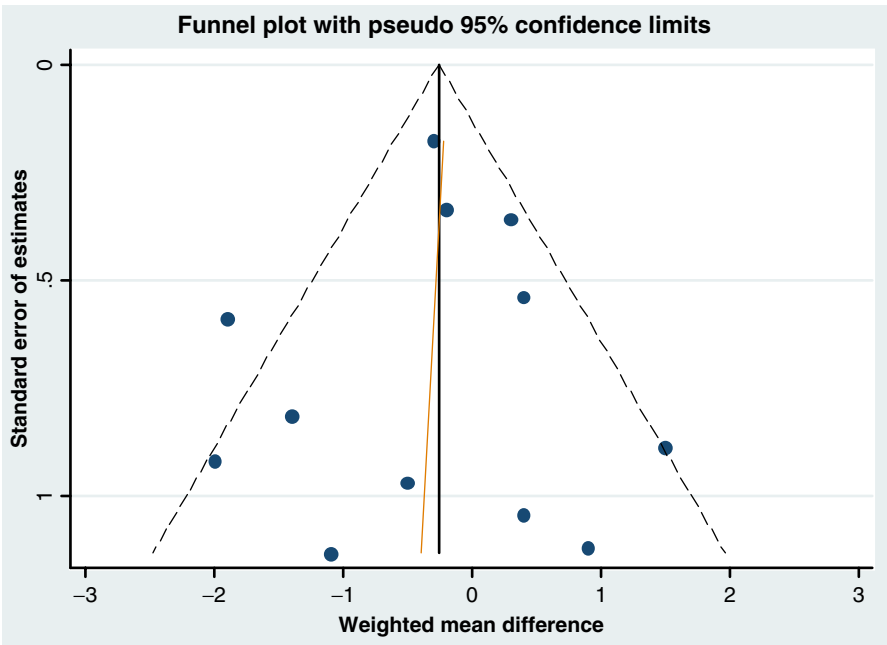

Fig. 3.8 Funnel plot of the WMD against its standard error and Egger regression line

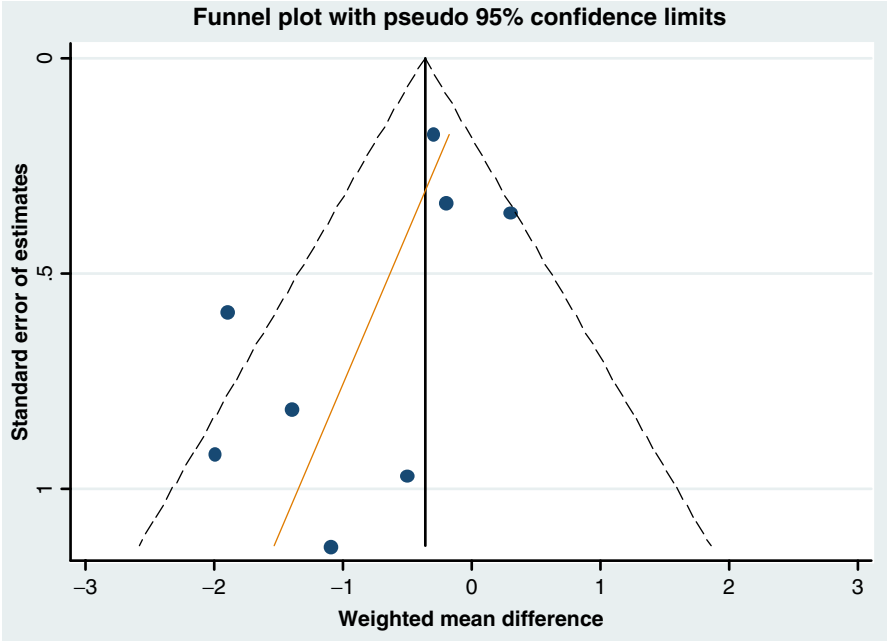

Fig. 3.9 Funnel plot of the WMD against its standard error and Egger regression line. (four studies with effects favoring the control treatment were removed)

3.8 Exploring Influential Studies

Sometimes a single study has a great impact in the estimates. Table 3.8 shows the WMD and 95% confidence intervals for the meta-analysis when one study is omitted at a time. Among the studies that showed large positive effect for the antiviral therapy, the first study published in 1997 has the greatest impact in the WMD. Omitting this study from the meta-analysis would change the WMD from −0.26 days to −0.18 days. A similar but contrary effect would be observed if the 2006 study was omitted. In this case, the WMD would change from −0.26 days to −0.34 days. The impact of a single study in the overall estimate is dependent upon the effect size and sample size. The study with largest influence on the confidence intervals (i.e., precision of the estimate) is the study published in 2007 due to its large sample size. The exclusion of this study would widen the confidence interval in approximately 40%. The search for very influential studies should be done with caution and more attention should be paid to influential small studies.

Table 3.8 Meta-analysis results after omitting one study at a time

Omitted study		Weighted mean difference	95% CI	
Year	Sample size		Lower	Upper
1997	42	−0.18	−0.43	0.08
1998	31	−0.22	−0.47	0.03
1999	44	−0.23	−0.48	0.02
1998	33	−0.29	−0.54	−0.04
2001	30	−0.27	−0.52	−0.02
2001	29	−0.25	−0.50	0.00
2002	27	−0.27	−0.52	−0.02
2002	31	−0.25	−0.50	0.00
2005	190	−0.27	−0.53	0.00
2008	80	−0.29	−0.55	−0.04
2006	145	−0.34	−0.60	−0.07
2007	394	−0.21	−0.57	0.14
Pooled weighted mean difference when all studies are included		−0.26	−0.50	−0.01

3.9 The Cochrane Collaboration Forest Plot

We have used Stata [9] to perform this meta-analysis due to personal preferences, but there are other software and statistical packages that can be used with minor differences in the results. The Cochrane Collaboration has the software Review Manager [5] for preparing systematic reviews and meta-analysis. The Forest plot generated by this software is presented in Fig. 3.10, which is very similar to Fig. 3.1.

3.10 Standardized Mean Differences

We have focused in this chapter on WMDs because it is more intuitive and easy to understand. With respect to continuous outcomes, the standardized mean difference can be used instead of the WMD. The standardized mean difference can be used when studies have measured the outcomes in different units. However, standardized mean differences are usually difficult to interpret because these measures are not directly related to everyday outcomes. In this case, the reader should look for the interpretation given to the results by the authors. Usually, standardized mean differences can be presented as the proportion of patients benefiting from the intervention, or a measure of the minimal important difference can be provided to assist the reader. As a rule of thumb, standardized mean differences ≥0.7 may be considered large effects. For our data, the standard mean difference would be −0.11 (95% confidence interval: −0.23 to 0.01, $p=0.07$) using a fixed-effect model, and −0.12 (95% confidence interval: −0.32 to 0.08, $p=0.24$) using a random-effects model (Table 3.9). These results indicate a small effect of the antiviral therapy, but the interpretation of the results is difficult to translate in practical terms. Several methods to calculate the standardized mean difference have been proposed such as the Glass method, Cohen method, and Hedges method.

3.11 Dichotomous Outcomes

Similar meta-analysis methods can be used for dichotomous (odds ratios and risk ratios), ordinal (indices and scales), counts and rates (number of events), and time-to-event data (survival). We will briefly present below some differences with regard to dichotomous outcomes because they are frequently reported in the medical and dental fields. In addition to the inverse-variance method already discussed for continuous data, three other methods are available for meta-analysis of dichotomous outcomes: Mantel–Haenszel and Peto methods for fixed-effect models and DerSimonian and Laird method for random-effects models. The Mantel–Haenszel is frequently used for fixed-effect models and is the standard method for several statistical programs. The Forest plot is also used to present the results with minor differences. Odds ratios and risk

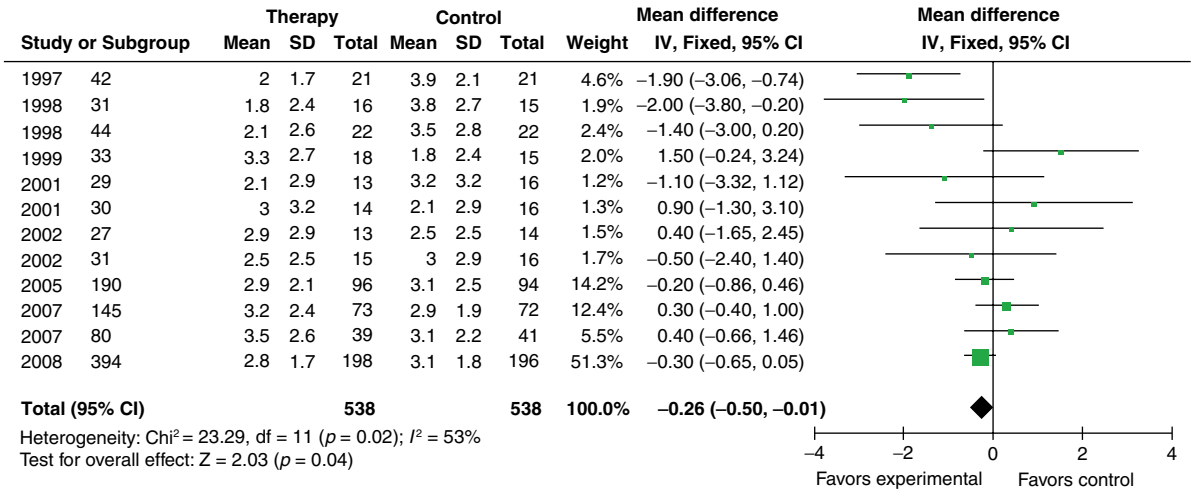

Fig. 3.10 Forest plot using the Cochrane Collaboration software (fixed-effect model)

Table 3.9 Meta-analysis results using the standardized mean difference instead of the weighted mean difference

Year	Sample size	Standardized mean difference	95% CI		Weight (%)
			Lower	Upper	
1997	42	−1.00	−1.64	−0.35	3.50
1998	31	−0.79	−1.52	−0.05	2.70
1999	44	−0.52	−1.12	0.08	4.00
1998	33	0.58	−0.12	1.28	2.95
2001	30	0.30	−0.43	1.02	2.78
2001	29	−0.36	−1.10	0.38	2.66
2002	27	0.15	−0.61	0.90	2.53
2002	31	−0.18	−0.89	0.52	2.90
2005	190	−0.09	−0.37	0.20	17.88
2008	80	0.17	−0.27	0.61	7.50
2006	145	0.14	−0.19	0.46	13.62
2007	394	−0.17	−0.37	0.03	36.97
Fixed-effect model		−0.11	−0.23	0.01	100.00
Random-effects model		−0.12	−0.32	0.08	100.00

ratios are frequently transformed using a natural log scale to facilitate analysis and presentation of the results (this transformation makes the scale symmetric). Thus, the horizontal axis of the Forest plot generally uses this scale, which may be misleading for the inexperienced reader. The same change in scale occurs for the Funnel plot. To test for Funnel plot symmetry in dichotomous data, Harbord et al., Peters et al., and Rücker et al. proposed alternative tests to the Egger test. Nevertheless, the same principles and interpretation of the results are still valid.

Needleman et al. [6] published a Cochrane review about guided tissue regeneration (GTR) for periodontal infra-bony defects compared to open flap debridement (control). The main outcome was clinical attachment gain that was dichotomized using a cut-off point of two sites gaining less than 2 mm of attachment. The Forest plot below was adapted from their study to illustrate an analysis of a dichotomous outcome with the Mantel–Haenszel method to pool the results across studies (Fig. 3.11). Results from 5 out of 6 studies favored GTR, but only one (the study by Tonnetti 1998) found a statistically significant difference compared to the control treatment. The meta-analysis demonstrated a final risk ratio of 0.54 indicating that the use of GTR for periodontal infra-bony defects significantly reduces 46% the chance of having ≥2 sites gaining less than 2 mm. Additionally, it can be seen that they found some heterogeneity ($I^2 = 44\%$) and, consequently, a random-effects model was applied.

3.12 Concluding Remarks

Before concluding this chapter we would like to acknowledge that some of the concepts and statistics presented in this chapter have been simplified in order to improve understanding to a broader audience. Readers with greater statistical background or who are planning on conducting a meta-analysis are encouraged to look for more specialized information on this subject [1–5]. An updated list of books and websites is provided in the references. We also would like to acknowledge that the data sometimes violated some statistical assumptions. These minor violations were necessary in order to build an interesting dataset that could be used to show several important steps in meta-analysis.

Systematic reviews and meta-analyses are an integral part of evidence-based health care practice. In this

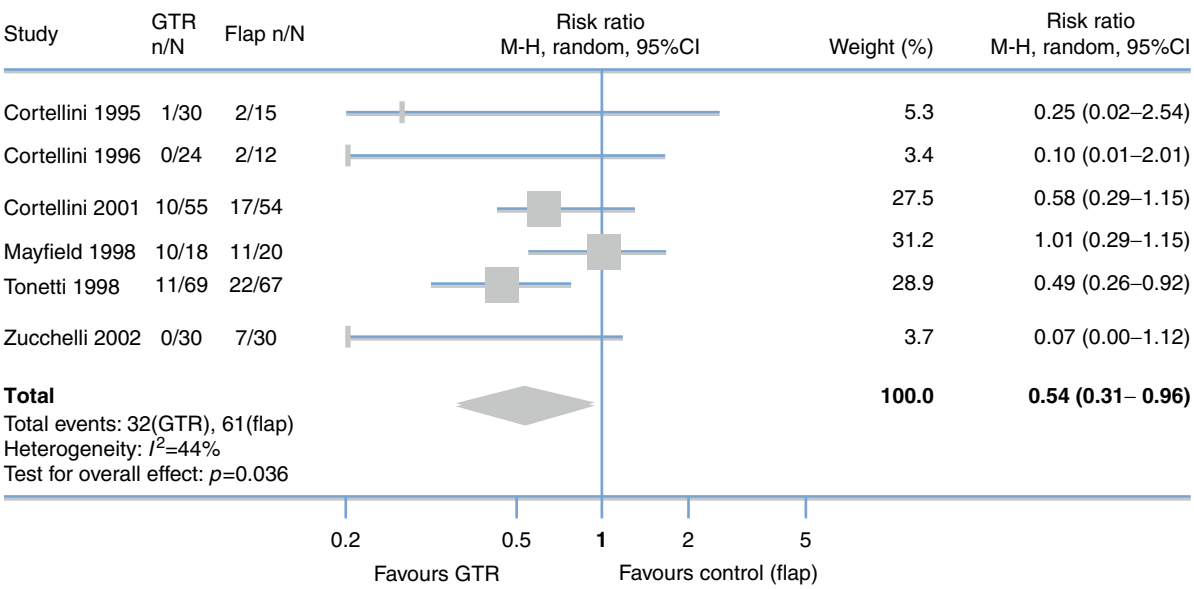

Fig. 3.11 Forest plot adapted from Needleman et al. [6]

context, we hope that this chapter will encourage more health care professionals to read and apply the evidence contained in systematic reviews and meta-analyses in their daily professional lives. Readers are also encouraged to remain updated since new developments over the years are likely to occur.

As a final message, we would like to call the reader's attention to the fact that we are approaching, at least in some areas of medicine and dentistry, a limit of how much information can be extracted from the current body of scientific evidence. Recent systematic reviews and meta-analyses have often been based in few studies of questionable quality yielding inconclusive results. Perhaps it is time to stop being creative with our systematic reviews and time to produce new and better evidence.

References

1. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (eds) (2009) Introduction to meta-analysis. Wiley, New York, 450pp
2. Centre for Reviews and Dissemination at University of York (2009) Systematic reviews – CRD's guidance for undertaking reviews in health care. Centre for reviews and dissemination: York Publishing Services Ltd, York, 3rd edn. 282pp
3. Egger M, Smith GD, Altman D (2001) Systematic reviews in health care: meta-analysis in context. BMJ Books, London, 512pp
4. Hartung J, Knapp G, Sinha BK (2008) Statistical meta-analysis with applications. Wiley, Hoboken, 248 pp
5. Higgins JPT, Green S (eds) (2008) Cochrane handbook for Systematic reviews of intervention. Wiley, Chichester, 672pp
6. Needleman I, Worthington Helen V, Giedrys-Leeper E, Tucker R (2009) Guided tissue regeneration for periodontal infra-bony defects (Cochrane Review). In: The Cochrane Library, Issue 1
7. Ng YL, Mann V, Gulabivala K (2008) Outcome of secondary root canal treatment: a systematic review of the literature. Int Endod J 41(12):1026–1046
8. Pavia M, Pileggi C, Nobile CG, Angelillo IF (2006) Association between fruit and vegetable consumption and oral cancer: a meta-analysis of observational studies. Am J Clin Nutr 83(5):1126–1134
9. Sterne J (ed) (2009) Meta-analysis: an updated collection from the Stata Journal. Stata Press, College Station, 259pp

Websites

The Cochrane Collaboration.http://www.cochrane.org/
The Cochrane Oral Health Group.http://www.ohg.cochrane.org/
The Centre for Reviews and Dissemination.http://www.york.ac.uk/inst/crd/index.htm
Comprehensive meta-analysis.http://www.meta-analysis.com/
The QUOROM statement (Quality of Reporting of Meta-analyses).http://www.consort-statement.org/
The GRADE working group (Grading of Recommendations Assessment, Development and Evaluation).http://www.gradeworkinggroup.org/