2006

Meta-Analysis of Open vs Closed Surgery of Mandibular Condyle Fractures

Marcy Lauren Nussbaum
Virginia Commonwealth University

META-ANALYSIS OF OPEN VS CLOSED TREATMENT OF MANDIBULAR CONDYLE FRACTURES

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

By

MARCY LAUREN NUSSBAUM

Bachelor of Arts, University of Virginia, 2004

Director: Dr. Al M. Best, Associate Professor of Biostatistics, Department of Biostatistics

Virginia Commonwealth University

Richmond, Virginia

May, 2006

Acknowledgement

I would like to take this opportunity to thank my advisor, Dr. Al Best, for all of his guidance and support throughout the process of writing and editing this thesis. I would also like to thank the other members of my committee, Dr. R.K. Elswick and Dr. Daniel Laskin for all of their contributions. I want to acknowledge all of my professors in the department of biostatistics, especially Dr. Kellie Archer for graphical assistance. A special thanks to Ryan Kimes and Amy Herrin, my fellow classmates who endured the pain of graduate school with me. I would not have completed this thesis without the loving support of my parents, Carol and Bruce Nussbaum.

Table of Contents

List of Tables

List of Figures

Abstract

META-ANALYSIS OF OPEN VERSUS CLOSED TREATMENT OF MANDIBULAR CONDYLE FRACTURES

By Marcy L. Nussbaum, M.S.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2006

Major Director: Dr. Al M. Best, Professor of Biostatistics

A review of the literature reveals a difference of opinion regarding whether open or closed reduction of condylar fractures produces the best results. It would be beneficial to critically analyze past studies that have directly compared the two methods in an attempt to answer this question. A Medline search for articles using the key words 'mandibular condyle fractures' and 'mandibular condyle fractures surgery' was performed. The articles chosen for the meta-analysis contained data on at least one of the following: postoperative maximum mouth opening, lateral excursion, protrusion, deviation on opening, asymmetry, and joint pain or muscle pain.

Several common statistical methods were used to test for differences between open and closed surgery, including the weighted average method for fixed and random effects as well as the Mantel-Haenszel method for fixed effects. Some of the outcome variables were found to be statistically significant but were interpreted with caution

because of the poor quality of the studies assessed. There is a need for more standardized data collection as well as patient randomization to treatment groups.

1 Introduction

A systematic review refers to a summary of the medical literature for a particular topic of interest. A meta-analysis is the statistical aspect of such a review that assesses heterogeneity between studies and estimates overall measures of association. The objectives of a meta-analysis include increasing power to detect an overall treatment effect, estimation of the degree of benefit associated with a particular study treatment, assessment of the variability between studies, and identification of study characteristics associated with particularly effective treatments.[1] The concept of combining results from multiple studies dates back a century to Karl Pearson and the effectiveness of inoculation against typhoid fever.[2] Meta-analyses began to enter the social sciences and healthcare fields around 30 years ago, but did not gain popularity until the early 1990s after Yusuf et al[3] published a meta-analysis on beta blockers in myocardial infarction.

A meta-analysis, when feasible, is considered superior to a non data-based systematic review. In an article in *Science*, Mann[4] compared the conclusions drawn from meta-analyses versus literature reviews in five subject areas: psychotherapy, delinquency prevention, school funding, job training, and reducing anxiety in surgical patients. The findings revealed that there is a tendency for literature reviews to underestimate the presence and strength of treatment effects.

However, Egger et al[5] argue that the optimal situation would be to conduct a meta-analysis within the framework of a systematic review. Such a review would include only those studies that meet specific criteria, which should minimize bias. Additionally, studies should meet a minimum level of quality to be included. If the meta-analysis is

comprised of poor quality studies, then it will be of poor quality as well. Maintaining quality is typically done by, among other things, clearly stating the inclusion and exclusion criteria before searching for articles, and by critically evaluating the studies found.

A meta-analysis can potentially resolve conflicting conclusions from multiple studies that address the same research question by increasing statistical power. That is, it may be more difficult to find statistically significant results in smaller individual studies but may become apparent when the effect is pooled across many studies. For example, upon reviewing 1,941 trials of schizophrenia treatment, Thronley and Adams[6] concluded that only 58 (3%) had a large enough sample size to detect a difference. Many clinical questions also cannot be addressed at a single location. In this situation the meta-analysis can become a very useful tool as it combines information across multiple study sites.

Since the meta-analysis includes several studies with different subjects, investigators and locations, the results are more generalizable to the population at large. Bangert-Drowns[7] makes the following claims:

> *"Meta-analysis is not a fad. It is rooted in the fundamental values of the scientific enterprise: replicability, quantification, causal and correlational analysis. Valuable information is needlessly scattered in individual studies. The ability of social scientists to deliver generalizable answers to basic questions of policy is too serious a concern to allow us to treat research integration lightly. The potential benefits of meta-analysis method seem enormous."*

The purpose of this study is to compare open versus closed surgery of mandibular condyle fractures. There is currently a debate in the field of oral and maxillofacial surgery as to which procedure produces the best results in patients. Closed treatment has been used as the standard and involves some form of wiring the jaws shut. Open treatment has not been in practice for as long as closed, and involves surgically exposing the temporomandibular joint to invasively fix the fracture. Many studies have compared these two procedures in terms of patient recovery by looking at the motion of the mandible after surgery. They often measure things like mouth opening, movement of the mandible forward, backward, and side to side, whether the jaw is properly aligned, and presence of joint pain. This study will assess a selection of outcome variables from various studies comparing open and closed treatment.

Section 2 provides the statistical theory and framework behind some available methods for meta-analyses. Procedures for both binary outcomes and continuous outcomes are outlined. This is by no means an exhaustive compilation of all methods available for meta-analyses. However, they have been chosen for this research as the most commonly recurring and popular methods in the medical literature.

Section 3 is comprised of the results discovered from the meta-analyses that were performed. The process by which the studies were chosen is explained. Several outcome variables were assessed, including postoperative maximum mouth opening, lateral excursion, protrusion, deviation on opening, asymmetry, and joint pain.

Section 4 covers the implications and clinical aspects of the results from section 3. Shortcomings of the current research are discussed and suggestions for future research are given.

2 Statistical Methods

This section presents the statistical methods commonly used in meta-analyses within the field of medicine. From the assessment of open and closed jaw surgery from the dental literature it was determined that the reported study outcomes were binary or continuous. Thus, the methods presented in this thesis pertain solely to binary or continuous outcomes. The purpose of this study is to test for the differential effects of open versus closed surgery.

Some methods propose models that use *fixed* effects, while others are categorized as using *random* effects. The choice of which model to use depends on how variable the outcome measures are between studies. If the chosen effect size estimates to be included in a meta-analysis are relatively consistent and have similar standard deviations, then they are deemed as homogeneous and a fixed effects model is used. If there is substantial heterogeneity between the effects from the various studies, then that variability must be accounted for in the meta-analysis by using a random effects model. Typically a statistical test for heterogeneity is used to determine which method to choose.

2.1. Categorical Methods

2.1.1 Binary Outcomes

Binary data have two possible outcomes for a particular dependent variable, e.g. died or survived, increased blood pressure or did not increase blood pressure. Binary data are typically summarized as counts for each of the two outcome categories in order to calculate proportions. A common approach for quantifying risk with binary data is to use the relative risk or the odds ratio. However, when considering how to implement these

outcomes in a meta-analysis, the risk difference may also be assessed. Odds ratios are appropriate for randomized clinical trials and case-control studies, while relative risks are appropriate for cohort studies. In the sections below the calculations for the odds ratios only will be explained.

Odds ratios can be calculated from contingency tables. The two dichotomous variables (group and outcome) combine to form four cells of possible outcomes in a contingency table. Let π_{rc} be the probability that (X,Y) falls into row r (the groups) and column c (the outcomes). π_{rc} can be thought of as the joint probability distribution. We represent the marginal distributions with $\{\pi_{r+}\}$ and $\{\pi_{+c}\}$, where

$$\pi_{r+} = \sum_{c} \pi_{rc} \tag{2.1}$$

and

$$\pi_{+c} = \sum_{r} \pi_{rc} \,. \tag{2.2}$$

These proportions are shown in Table 1 where group and outcome each have two levels.

Table 1. THE JOINT AND MARGINAL PROBABILITIES OF A 2x2 TABLE

	Outcome		
Group	Success	Failure	Total
Treatment	π_{11}	π_{12}	π_{1+}
Control	π_{21}	π_{22}	π_{2+}
Total	π_{+1}	π_{+2}	1.0

The odds that there is a success in the treatment group is

$$\Omega_1 = \pi_{11} / \pi_{12}, \tag{2.3}$$

and the odds that there is a success in the control group is

$$\Omega_2 = \pi_{21} / \pi_{22}. \tag{2.4}$$

The odds ratio of Ω_1 to Ω_2 is

$$\omega = \frac{\Omega_1}{\Omega_2} = \frac{\pi_{11} / \pi_{12}}{\pi_{21} / \pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}. \tag{2.5}$$

An odds ratio of 1 implies that there is no association between the two variables. An odds ratio greater than 1 implies that the treatment was more effective than the control. We can express the 2x2 contingency table in terms of frequencies or number of observations in each cell as shown in Table 2.

Table 2. FREQUENCY COUNTS IN A 2x2 TABLE

	Outcome		
Group	Success	Failure	Total
Treatment	n_{11}	n_{12}	n_{1+}
Control	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n

Here, the sample odds ratio is

$$\hat{\omega} = \frac{n_{11}n_{22}}{n_{12}n_{21}}. \tag{2.6}$$

Often the odds ratios will be positively skewed, indicating that a logarithmic transformation is necessary to normalize the data. It is common to use the natural log of the odds ratio. The estimated variance of $\ln(\hat{\omega})$ is

$$\mathrm{Var}\{\ln(\hat{\omega})\} = \left(\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}} \right). \tag{2.7}$$

The $100(1-\alpha)$% confidence interval around $\ln(\hat{\omega})$ can be shown as

$$\ln\left(\hat{\omega}\right) \pm Z_{\alpha/2} \sqrt{\mathrm{Var}\left\{\ln\left(\hat{\omega}\right)\right\}} \tag{2.8}$$

where $Z_{\alpha/2}$ is the two-sided critical value from the standard normal distribution at a specified α. These values are then exponentiated to get the confidence interval for the odds ratio in its original units.

2.1.2 Methods for Binary Data

The following sections describe three commonly used types of meta-analysis performed on binary data. It should be noted that the first two methods can be used for many types of effect size estimates including odds ratios, relative risks, risk differences, and even estimates from continuous data. The third method covered is used specifically for odds ratios.

2.1.2.1 Weighted Average Method for Fixed Effects

The weighted average method was first described by Birge[8] and Cochran[9] in the 1930s and is conceptually straightforward. Each study effect size estimate is given a weight that is inversely proportional to its variance. This method can be used for many types of estimators. The number of studies included in the meta-analysis is represented by the

variable k. Assuming that the k studies are homogeneous (similar with respect to estimated effect), the weighted average estimator of the population effect size is calculated as follows

$$\overline{T}_+ = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i}, \tag{2.9}$$

where the weights are defined to be $w_i = 1/Var\{T_i\}$. Defining the weights in this manner ensures that the studies with large variability are weighted less overall, while the studies with small variability are weighted more overall. From this point forward, the term $\sum_{i=1}^{k}$ will be abbreviated by $\sum$.

The variance for the estimator $\overline{T}_+$ can be estimated as

$$\text{Var}\left\{\overline{T}_+\right\} = \frac{\sum w_i^2 Var(T_i)}{\left(\sum w_i\right)^2} = \frac{\sum w_i^2 \left(\frac{1}{w_i}\right)}{\left(\sum w_i\right)^2} = \frac{\sum w_i}{\left(\sum w_i\right)^2} = \frac{1}{\sum w_i}. \tag{2.10}$$

The corresponding $100(1-\alpha)\%$ confidence interval for $\overline{T}_+$ is

$$\overline{T}_+ \pm Z_{\alpha/2}\sqrt{\text{Var}\{\overline{T}_+\}} \tag{2.11}$$

where $Z_{\alpha/2}$ is the two-sided critical value from the standard normal distribution at a specified α. When working with categorical outcomes, $\ln(\overline{\omega}_+)$ can be used in place of $\overline{T}_+$.

2.1.2.2 Weighted Average Method for Random Effects

If the studies are found to be heterogeneous (dissimilar with respect to effects) then a random effects model would be more appropriate than a fixed effects model. The Weighted Average Method for Random Effects is actually just a modification of the Weighted Average Method for Fixed Effects. Here, we wish to estimate an extra variance term known as the between study variance. The same weights are used $(w_i = 1/\mathrm{Var}\{T_i\})$ as well as the same effect size estimate $(\overline{T}_+)$. The mean and variance of the weights from the studies are defined as

$$\overline{w} = \sum w_i / k \tag{2.12}$$

and

$$s_w{}^2 = \frac{1}{k-1}\left(\sum w_i{}^2 - k(\overline{w})^2\right) \tag{2.13}$$

The random effects model assumes that the study specific effect sizes come from a random distribution of effect sizes with a fixed mean and variance.[10] If T_i is an estimate of effect size and θ_i is the true effect size in the *i*th study, then

$$T_i = \theta_i + e_i \tag{2.14}$$

where e_i is the error associated with T_i. The variance of T_i is

$$\mathrm{Var}\left(T_i\right) = \tau^2 + v_i\,, \tag{2.15}$$

where τ^2 is the random effects variance and v_i is the variance due to sampling error in the *i*th study.[10] The estimated between study variation in effect size, $\hat{\tau}^2$, is calculated as

$$\hat{\tau}^2 = 0 \text{ if } Q \le k-1 \tag{2.16}$$

or

$$\hat{\tau}^2 = \left(Q-\left(k-1\right)\right)/U \text{ if } Q > k-1 \tag{2.17}$$

where

$$U = \left(k-1\right)\left(\overline{w} - \frac{s_w{}^2}{k\overline{w}}\right). \tag{2.18}$$

Then the adjusted weights w_i* are calculated as

$$w_i* = \frac{1}{\left[\left(1/w_i\right)+\hat{\tau}^2\right]}. \tag{2.19}$$

If the between study variance $\hat{\tau}^2$ is zero then the weights are unaffected and the random effects model reduces to the fixed effects model. The treatment effect is estimated as

$$\overline{T}_{RND} = \frac{\sum w_i* T_i}{\sum w_i*} \tag{2.20}$$

and its variance[10] is expressed as

$$\text{var}(\overline{T}_{RND}) = \frac{1}{\sum w_i*}. \tag{2.21}$$

The $100(1-\alpha)\%$ confidence interval under normality is

$$\overline{T}_{RND} \pm \frac{Z_{\alpha/2}}{\sqrt{\sum w_i*}}. \tag{2.22}$$

2.1.2.3 Mantel-Haenszel Method for Fixed Effects for Odds Ratios

In certain instances there may be small cell counts or even empty cells. Note that in these situations the odds ratios and their variance estimates are undefined. A good remedy for this problem is to use the Mantel-Haenszel estimator. The common odds ratio[11] is calculated as follows

$$\hat{\omega}_{\mathrm{MH}} = \frac{\sum n_{11_i} n_{22_i} / n_i}{\sum n_{12_i} n_{21_i} / n_i}. \tag{2.23}$$

The following equation is a large sample approximation[12, 13] of the variance

$$\mathrm{Var}\left\{\ln\left(\hat{\omega}_{MH}\right)\right\} = \frac{\sum A_i C_i}{2\left(\sum C_i\right)^2} + \frac{\sum\left(A_i D_i + B_i C_i\right)}{2\left(\sum C_i\right)\left(\sum D_i\right)} + \frac{\sum B_i D_i}{2\left(\sum D_i\right)^2} \tag{2.24}$$

where

$$A_i = \frac{n_{11_i} + n_{22_i}}{n_i}, B_i = \frac{n_{12_i} + n_{21_i}}{n_i}, C_i = \frac{n_{11_i} n_{22_i}}{n_i}, D_i = \frac{n_{12_i} n_{21_i}}{n_i}. \tag{2.25}$$

The corresponding $100(1-\alpha)\%$ confidence interval for $\ln(\hat{\omega}_{\mathrm{MH}})$ is

$$\ln\left(\hat{\omega}_{MH}\right) \pm Z_{\alpha/2}\sqrt{\mathrm{Var}\left\{\ln\left(\hat{\omega}_{MH}\right)\right\}}. \tag{2.26}$$

Since there are several methods for combining odds ratios, it would be helpful to know when to use which method. In general, if the number of studies is small, but the within-study sample sizes are large, the Weighted Average method should be used.[14] If there are many studies but the within-study sample sizes are small, then the Mantel-Haenzel method is preferred. Peto's method is a third technique used for combining odds ratios. It has the advantage of effectively dealing with zero cell counts from individual

studies and it is easy to calculate. However, it may produce seriously biased odds ratios and standard errors when there is severe imbalance in sample size between the two groups being compared.[15]

2.1.3 Continuity Corrections

In some instances the outcome of interest for a meta-analysis may be a rare event. For example, in a meta-analysis evaluating a drug intervention, adverse side effects may be rare but serious and therefore important.[16] Thus, there may be zero events in one or both arms of a study. This can become a problem when trying to estimate an effect size such as an odds ratio. A typical method of resolving this situation is to add some constant m to each cell so that no cell is equal to zero. In a standard 2x2 table it has been argued that $m = 1/2$ is a good amount to add to each cell. Research in this area has shown that this continuity correction might not be the optimal choice. Sweeting et al[17] propose several alternatives to use within the context of a meta-analysis.

When only a few studies included in a meta-analysis suffer from zero cell counts, the overall impact in the conclusions will be slight. However, with sparse data throughout a majority of the studies, the impact could be more substantial. Choice of continuity correction then may be an influential factor. It is common practice to remove studies in which there are no events in both treatment arms from a meta-analysis.[17] These studies may be referred to as zero total events. Whitehead and Whitehead[18] believe that these studies should be excluded since they provide no information on the magnitude of the treatment effect. However, others believe such studies should be included in a meta-analysis[19] in order to take into account the sample sizes of these studies.[20] Sweeting et

al[17] outline three different approaches to implementing continuity corrections in a meta-analysis: adding a constant m to each cell in the 2x2 contingency table, using the reciprocal of the opposite treatment arm size, and using an empirical continuity correction.

2.1.3.1 Adding a Constant m

It is most common to use a constant correction factor when faced with zero cell counts in a meta-analysis. The most widespread correction used is $m = 1/2$. The reason for choosing this particular value has arisen from an analysis by Cox[21] on the odds of a single study group. Cox's analysis suggests that when using odds as an effect measure, choosing a correction factor of $1/2$ gives the least biased estimator of the true log odds for a single treatment group. However, other choices of correction factor may be less biased and have better coverage when looking at the odds ratio between two groups. For example, consider the following contingency table (see Table 3).

Table 3. EXAMPLE COUNT DATA

Group	Event	No Event
Treatment	0	100
Control	1	400

With a correction factor of $1/2$ the odds ratio for the data in Table 3 is given by

$$\hat{\omega} = \frac{0.5 * 400.5}{100.5 * 1.5} = 1.33. \qquad (2.27)$$

This suggests that there is an increased risk in the treatment group, yet there were zero events in the treatment group. This paradoxical conclusion prompts the consideration of other types of continuity corrections.

2.1.3.2 The Reciprocal of the Opposite Treatment Arm Size

Choices other than $m = 1/2$ might be more appealing when the groups are unbalanced with respect to sample size. Using a constant correction factor might yield an odds ratios that simply reflects the sample size disparity between the two groups and not the underlying effect size.

Table 4. 1:4 UNBALANCED GROUPS COUNT DATA

Group	Event	No Event
Treatment	0	100
Control	0	400

For example in Table 4, when $k = 1/2$, the odds ratio is approximately 4 (note the ratio of the two sample sizes).

$$\hat{\omega} = \frac{0.5 * 400.5}{100.5 * 0.5} \approx 4 \tag{2.28}$$

If a smaller correction factor that accounts for sample size is used, different results are produced. If $m = 1/400$ for the treatment group and $m = 1/100$ for the control group, the odds ratio is 1.

$$\hat{\omega} = \frac{0.0025 * 400.01}{100.0025 * 0.01} = 1 \tag{2.29}$$

By controlling for the disparity in sample size in the two groups, a more intuitive result is found. As shown above, the reciprocal of the sample size can act as a helpful continuity correction.

2.1.3.3 An Empirical Continuity Correction

The 'treatment' arm continuity correction has the effect of 'pulling' the estimated odds ratio arbitrarily toward no effect, i.e. odds ratio of 1. Perhaps it is more desirable to 'pull' the estimate in the direction of the pooled effect size estimate obtained in the analysis. An empirical approach can be adopted where all of the studies in the meta-analysis without a zero event are used to calculate a pooled odds ratio. Using this estimate, a continuity correction can be calculated which will produce odds ratio estimates close to the pooled odds ratio in the studies with zero events in both arms.[17] We can think of this continuity correction factor as a 'prior' (empirically derived from other studies) added to the observed events.

Suppose that an estimated pooled odds ratio, $\hat{\Omega}$, was obtained using the non-zero studies. Let n_t be the treatment group sample size, n_c be the control group sample size, and R be the group ratio imbalance, where $R = n_c / n_t$ (note that $n_c = n_t R$). Then a total zero event study with continuity corrections k_t and k_c, for the treatment and control groups respectively, is shown in Table 5.

Table 5. TOTAL ZERO EVENT STUDY

Group	Event	No Event
Treatment	$0+k_t$	n_t+k_t
Control	$0+k_c$	n_c+k_c

In order for k_t and k_c to be empirical continuity corrections, it is required that[17]

$$\frac{k_t\left(n_t R+k_c\right)}{k_c\left(n_t+k_t\right)}=\hat{\Omega}. \tag{2.30}$$

The left-hand side of the equation can be approximated by Rk_t/k_c when the group sample sizes are relatively large. If the restriction $k_t+k_c=1$ is imposed, as it is when we use $m=1/2$, then the following equations[17] are obtained:

$$\frac{R\left(1-k_c\right)}{k_c}\approx\hat{\Omega} \tag{2.31}$$

$$\Rightarrow k_c\approx\frac{R}{R+\hat{\Omega}} \tag{2.32}$$

and

$$\Rightarrow k_t\approx\frac{\hat{\Omega}}{R+\hat{\Omega}}. \tag{2.33}$$

After using equations (2.32) and (2.33) to calculate the appropriate empirical continuity corrections, k_c and k_t are added to the control and treatment groups, respectively. The pooled odds ratio $\hat{\Omega}$ can be estimated from all of the non-zero studies in a meta-analysis or even from some known estimate in the literature. This information

along with the sample sizes for both groups will be enough to arrive at empirical continuity corrections. Agresti[22] and others suggest that adding smaller continuity corrections might provide better estimates for the true odds ratio. If we wish to allow $k_t + k_c = 0.01$, the equations (2.32) and (2.33) become

$$k_c \approx \frac{R}{100(R+\Omega)} \tag{2.34}$$

and

$$k_t \approx \frac{\Omega}{100(R+\Omega)}. \tag{2.35}$$

Continuity corrections can be very useful in situations where cell counts are small or zero. It may be beneficial to use the methods proposed by Sweeting et al[17] in addition to the traditional correction of $m = 1/2$ to see how the different methods impact the effect size estimate. Then the preferred method can be chosen for the final analysis.

2.2. Continuous Methods

The following sections describe methods for using continuous data in a meta-analysis. The two meta-analytic methods presented can be seen to be simply variations of two of the methods covered in sections 2.1.2.1 and 2.1.2.2 on categorical data.

2.2.1 Continuous Outcomes

Continuous data are quantitative in nature and are measured on a continuous, numerical scale. Continuous data collected in health studies are usually measured on a positive scale. It is typically of interest to identify some type of difference in average effect between the treatment and control groups in a clinical trial. If studies in a meta-analysis

report an outcome on different scales, they cannot be combined together without some type of transformation to a standardized scale.

If means are given with a range instead of a standard deviation the following transformation is applied.[23] The standard error of the mean is estimated as

$$\widehat{se}(\mu) = \frac{(upper - lower)}{N}, \tag{2.36}$$

where μ is the sample mean, *upper* is the upper range value, *lower* is the lower range value, and N is the sample size. These standard errors can then be converted to standard deviations with the following formula

$$\widehat{SD} = \widehat{se}(\mu) * \sqrt{N}\,. \tag{2.37}$$

One measure of a continuous treatment effect is the standardized mean difference. We define the population standardized mean difference as

$$\delta = \frac{\mu_t - \mu_c}{\sigma} \tag{2.38}$$

where μ_t and μ_c are the population means for the treatment and control groups and σ is the population standard deviation for the mean difference. There are several proposed methods of estimating δ.

2.2.1.1 Standardized Mean Difference Estimates

Glass's Δ: Glass'[24] estimate for δ takes on the following form

$$\hat{\Delta} = \frac{\bar{Y}_t - \bar{Y}_c}{s_c} \tag{2.39}$$

with $\overline{Y}_t$ and $\overline{Y}_c$ as the treatment and control group sample means and s_C as the control group sample standard deviation. The estimated variance of $\hat{\Delta}$ is

$$Var\left(\hat{\Delta}\right) = \frac{n_t + n_c}{n_t n_c} + \frac{\hat{\Delta}^2}{2\left(n_c - 1\right)} \tag{2.40}$$

where n_t and n_c are the respective sample sizes for the treatment and control groups. Glass' estimator is appropriate if more than one treatment group is being compared to a single control group or if the treatment and control group population standard deviations are highly likely to differ.[25] However, it is often reasonable to assume the population variances do not differ even if the sample variances do (especially when only comparing two groups).[26]

Cohen's d : Under the assumption of equal variances, we can pool the variances from the two groups. This should produce more precise estimation than using the control group sample standard deviation alone. Cohen[27] proposed estimating δ with d where

$$d = \frac{\overline{Y}_t - \overline{Y}_c}{\sigma}. \tag{2.41}$$

The maximum likelihood estimator[26] Cohen chose for estimating σ is given by

$$s_m = \sqrt{\frac{\left(n_t - 1\right)\left(s_t\right)^2 + \left(n_c - 1\right)\left(s_c\right)^2}{n_t + n_c}} \tag{2.42}$$

where $\left(s_t\right)^2$ and $\left(s_c\right)^2$ are the respective sample variances for the treatment and control groups. When both the treatment and control group sample sizes are large, the variance of d can be estimated as

$$Var(d) = \left(\frac{n_t + n_c}{n_t n_c} + \frac{d^2}{2(n_t + n_c - 2)} \right). \tag{2.43}$$

Hedges' g : Hedges uses a pooled sample standard deviation given by

$$S_{pooled} = \sqrt{\frac{(n_t - 1)(s_t)^2 + (n_c - 1)(s_c)^2}{n_t + n_c - 2}} \tag{2.44}$$

to estimate δ.[28] The standardized mean difference can be represented as

$$g = \frac{\overline{Y}_t - \overline{Y}_c}{S_{pooled}}. \tag{2.45}$$

When both treatment and control groups have large sample sizes, the estimated variance[28] of g is

$$Var(g) = \frac{n_t + n_c}{n_t n_c} + \frac{g^2}{2(n_t + n_c - 2)}. \tag{2.46}$$

All three estimators of δ are biased.[25] The bias may become a serious problem if the group sample sizes are small. As long as the equal variance assumption holds, it would be unwise to choose Glass's Δ because it only uses the variance estimate from the control group. Hedges' g has smaller sample variance than Cohen's d.[25] Hedges gives an exact correction factor for the sample bias

$$C(\upsilon) = \frac{\Gamma\left(\frac{\upsilon}{2}\right)}{\sqrt{\frac{\upsilon}{2}}\Gamma\left(\frac{\upsilon - 1}{2}\right)} \tag{2.47}$$

where $\upsilon = n_t + n_c - 2$ are the degrees of freedom and $\Gamma(.)$ is the gamma function.[29] This gives an unbiased estimate of δ stated as

$$g_u = C(\upsilon) * g \tag{2.48}$$

When the sample sizes in the treatment and control groups are equal, g_u is the unique minimum variance unbiased estimator of δ.[28]

When the treatment and control group sample sizes are large, the estimated variance[26] of g_u is

$$Var(g_u) = \frac{n_t + n_c}{n_t n_c} + \frac{{g_u}^2}{2(n_t + n_c)} \tag{2.49}$$

It appears that g_u proposed by Hedges[28] is the preferential estimator of δ due to its favorable statistical properties.[25] Once a particular estimator has been chosen, the meta-analysis can be performed.

2.2.2 Methods for Standardized Mean Differences

The following sections discuss two types of meta-analytic methods for a continuous outcome measure in terms of a standardized mean difference.

2.2.2.1 Weighted Average Method for Fixed Effects

This method of combining effect sizes for continuous data is the same as the method for categorical data for fixed effects with the only differences being the actual effect size and its variance. It is essentially a modification of section 2.1.2.1. Let δ_i and $Var(\delta_i)$ represent the effect size and its corresponding large sample variance estimator in the *i*th study. We may choose to use any of the previously mentioned estimators of the standardized mean difference: Δ_i, d_i, g_i, or g_{u_i}. Assuming that the studies are

homogeneous, the weighted average estimator of the population standardized mean difference is calculated as follows

$$\delta_+ = \frac{\sum w_i \hat{\delta}_i}{\sum w_i} \tag{2.50}$$

where the weights are defined as $w_i = 1/Var\{\hat{\delta}_i\}$.[30]

The variance for the estimator $\hat{\delta}_+$ is

$$\text{Var}\left\{\hat{\delta}_+\right\} = \frac{1}{\sum w_i} \tag{2.51}$$

and the corresponding $100(1-\alpha)\%$ confidence interval for δ is

$$\hat{\delta}_+ \pm Z_{\alpha/2}\sqrt{\text{Var}\{\hat{\delta}_+\}}\,. \tag{2.52}$$

2.2.2.2 Weighted Average Method for Random Effects

This method for random effects for continuous data is also the same as the method for categorical data that was previously presented. It is essentially a modification of section 2.1.2.2. The calculations are done in the same way with the standardized mean difference and its variance. The same weights are used where $w_i = 1/\text{Var}\{\hat{\delta}_i\}$.

2.2.3 Converting Continuous Data

There may be instances where different studies report outcomes in different formats. Often in medical research data collected on a continuous scale is reported as categorical. It will be categorized at some chosen point or points along a continuum. It is desirable to choose a cutoff point that has been established as clinically meaningful in the literature.

For instance, one might classify those with low diastolic blood pressure as <=90 mm Hg and those with high blood pressure (hypertension) as >90 mm Hg. This is an easy way to make categories within a population in order to diagnose and establish a standard of care. When conducting a meta-analysis we usually have to use the outcomes in the metric in which they are reported. Thus, we may end up with some continuous measures and some categorical measures for the same variable. In an effort to use all possible data available for analysis the continuous data is converted to categorical equivalents. By choosing a method of treatment difference such as the log-odds ratio, which can be estimated from both binary and continuous data, the number of trials included in the meta-analysis is increased and better representation is ensured.[18]

The classical method of estimating the risk or prevalence of the continuous data is to dichotomize the outcome variable at the cutoff value. Then the statistical analysis uses methods developed specifically for binary data, usually based on the binomial distribution.[31] Suissa proposes a method that is based on the assumption of a Gaussian (normal) distribution which does not resort to dichotomization. He found that the binomial approach was less efficient than his method by up to 67%. His method was also very accurate for small sample sizes.

Let Y be a continuous outcome variable normally distributed with mean μ and standard deviation σ. The conventional approach to estimating the probability of an event of interest is to choose a cutoff value C such that the probability of the event of interest is $R = pr\{Y > C\}$. The approach used to estimate this probability from a random sample of n observations of Y is to count the number of observation, say a, that are

larger than the cutoff C. It is assumed these data follow a binomial distribution with parameters n and R. The probability of the event of interest (R) is estimated by maximum likelihood estimation as $r = a/n$ and its variance as $\widehat{\text{var}}(r) = r(1-r)/n$.

Suissa's method[31] can be described as follows. Using the continuous data values and their assumed normal distribution, the maximum likelihood estimator of R is

$$r = pr\{Z > c^*\} = 1 - \Phi(c^*) \tag{2.53}$$

where

$$c^* = \frac{(C - \bar{y})}{s}. \tag{2.54}$$

The terms $\bar{y}$ and s are the sample mean and standard deviation and Z follows a standard normal distribution with cumulative distribution function Φ. The MLE of the variance of r is given by

$$\widehat{\text{var}}(r) = \left[1 + \left(c^{*2}/2\right)\right]\left[f(c^*)\right]^2 n, \tag{2.55}$$

where f is the standard Gaussian density function, i.e.

$$f(x) = \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}}. \tag{2.56}$$

In the case of two independent samples, R_t and R_c are estimated separately by r_t and r_c. The estimated odds ratio, $OR = R_c(1-R_t)/[R_t(1-R_c)]$ may be calculated. The variance is simply a function of the one-sample estimate of the risk and its variance. It is is approximated by[31]

$$\widehat{\text{var}}(\log OR) = [r_c(1-r_c)]^2 \widehat{\text{var}}(r_c) + [r_t(1-r_t)]^2 \widehat{\text{var}}(r_t)\,. \qquad (2.57)$$

Suissa[31] showed that his method of risk estimation from the normal distribution is more efficient than the method based on binary data under a binomial distribution. At equal sample size, the variance from the normal model is roughly two-thirds of that from the binomial model when the risk is between 10%-90%. The normal model performs even better when the risk is less than 10% or greater than 90% (the variances decrease rapidly to nearly zero). Also, the binomial model encounters problems when zero events are observed in the sample while the normal model does not.

Finally, dichotomization carries with it subjectivity with respect to classification. A slight difference in cutoff could cause a large difference in results when many points are at or near the cutoff. The method based on the continuous model is not as sensitive to minute shifts in cutoff. It must be noted, however, that the continuous model is completely reliant on the underlying normal distributional assumption for the data.

Before conducting any of the previously mentioned analyses in this chapter it is necessary first to test for heterogeneity among the studies. How different the studies are from one another will determine what type of model will be used for a particular outcome measure.

2.3. Heterogeneity

Heterogeneity refers to a state of dissimilarity among outcome measures from a group of studies. In this situation it is assumed that there are other sources of variability present among outcome measures than just sampling error. Additional sources of variability could arise when the studies are not conducted in the same way. For example, the

variability could be caused by things like differences in intervention methods, study design, measurement, or subject profile (age, race, etc.). This extra variation needs to be accounted for when performing a meta-analysis. A random effects model is used when the outcomes are considered to be heterogeneous, while a fixed effects model is used when they are considered to be homogeneous. Random effects models have greater generalizability but less statistical power.[25]

Thus, it is essential to test for heterogeneity before conducting the meta-analysis. The Q statistic of heterogeneity is commonly used and is presented next. First, the outcome measures from the various studies (odds ratios, standardized mean differences, etc.) are weighted and summed to yield $\bar{y}_w$, a weighted estimator of the treatment effect. This is calculated as

$$\bar{y}_w = \frac{\sum w_i y_i}{\sum w_i}, \tag{2.58}$$

where y_i is the treatment effect estimate for the *i*th study and w_i is the inverse of the *i*th sampling variance.[32] Next, the difference between this average and each outcome is calculated, weighted, and summed as follows

$$Q = \sum w_i (y_i - \bar{y}_w)^2 . \tag{2.59}$$

These calculations yield Q, a sum of squares that indicates how spread out the outcome measures are from the weighted average estimator. It follows a chi-square distribution with $k-1$ degrees of freedom. Q is known as the commonly used large sample test statistic for heterogeneity.

Heterogeneity has both advantages and disadvantages: exploring reasons for its presence can lead to useful insights, accounting for it can make modeling problematic and failure to allow for it may lead to inappropriate results. Whether or not a group of studies is identified as heterogeneous can greatly determine the overall effect size estimate and subsequent conclusions.

The concepts covered in this section provide the theoretical background for the analyses used in this study. Both categorical and continuous data were assessed using the previously discussed methods for meta-analysis. Section 3 explains the steps by which the analysis was done and the subsequent findings. Several different approaches were taken in order to present a broad prospective on the studies of interest.

3 Results

3.1. Overview

The initial literature search was conducted in PubMed. Two searches were run with the following keyword phrases, 'mandibular condyle fractures' and 'mandibular condyle fractures surgery.' The searches were restricted to the English language only. These articles were reviewed first from their abstracts and those that were identified as off-subject or without data were excluded. The remaining articles were collected and assessed for certain outcome variables. Articles that did not contain any of the variables of interest were discarded. Any article that appeared to have duplicate data was also excluded. However, those with duplicate *subjects* were not excluded. Ellis and Throckmorton, for example, might have used the same patients for multiple studies but each study measured a different outcome variable. [33, 34] Additional studies were added after reviewing reference sections of the articles initially chosen.

The outcome variables included in this study were chosen on the basis of the following criteria. The statistical summary of results had to be available from both open and closed treatment groups. For example, scarring and facial paralysis were not included because they only occurred in the open group. Outcome measures that were objective were given preference. Subjective responses from the patients and subjective judgments from physicians were kept to a minimum. Outcomes that occurred in at least several of the selected papers were considered for inclusion in this study. Each outcome variable in the meta-analysis was analyzed independently.

3.2. Studies

Of the 32 articles identified, 13 met the final selection criteria (see Table 6). These contained data on at least one of the following postoperative measures: maximum mouth opening, deviation on opening, lateral excursion, protrusion, facial asymmetry, or joint or muscle pain. It was decided to use measurements that were taken at least 6 months after treatment, with preference to those obtained 1 year after treatment. It was the interest of this study to look at the outcome measures after the patients have healed from the procedures, not at time points during the healing process. The long-term effects are most important in the recovery of the patients. Note that only one[35] of the 13 studies randomized subjects to a treatment group. In the next sections, results for the following outcome measures will be given: maximum mouth opening (section 3.3), deviation (section 3.4), lateral excursion (section 3.5), protrusion (section 3.6), facial asymmetry (section 3.7), and joint or muscle pain (section 3.8).

Table 6. RAW DATA FROM 13 STUDIES (FINAL SELECTION)

2	Yang et al[41]	1we, 2we, 1mo, 2mo, 3mo, 4mo, 6mo, 1yr	36	30	mean=41.57, range=29-44	mean=46, range=31-53	count=8	count=12								count=0	count=0	count=2	count=5
3	Santler et al[42]	mean=2.5yr, minimum 6mo	37	113	mean=45.5, SD=7.3, range=34-67	mean=47, SD=6.8, range=26-70				mean=8.5, SD=3.3, range=3-17		mean=8.7, SD=3.4, range=0-15	mean=5.9, SD=2.3, range=0-10	mean=6.2, SD=2.7, range=0-13	count=1	count=1	count ~1	count ~4
4	Konstantinovic and Dimitrijevic[43]	mean=2.5yr	26	54	mean=39 mm, range=23-50	mean=39 mm,range =10-60							count=2	count=3				
5	Takenoshita et al[36]	2yr (mean=11.6mo)	16	20	mean=39	mean=50			mean=8.7, SD=3	mean=7.9, SD=2.1	mean=6, SD=3.8	mean=6, SD=3.8	9.5mm (SD=2.1)					
6	Hidding et al[38]	1yr - 5yr	20	14	count=0 (<30mm)	count=0 (<30mm)	count=2 (>3mm)	count=9 (>3mm)	count=0 (<5mm)	count=0 (<5mm)	count=1 (<5mm)		count=1 (<5mm)	count=1 (<5mm)				
7	Oezman et al[39]	6mo - 24mo	20	10	count=0 (<40mm)	count=0 (<40mm)	count=0 (>2mm)	count=0 (>2mm)	count=0 (<6mm)	count=0 (<6mm)	count=0 (<6mm)		count=0 (<6mm)	count=0 (<6mm)				
8	Throckmorton and Ellis[33]	6we, 6mo, 1yr, 2yr, 3yr	62	74	mean=45.7, SD=9.4	mean=46, SD=12.9	mean=0.4, SD=6.8	mean=4.2, SD=6.6	mean=10.9, SD=2.5	mean=10.1, SD=2.8	mean=10.3, SD=3.6	mean=9.4, SD=3.5	mean=8.3, SD=2.8	mean=7.2, SD=2.8				
9	Widmark et al[40]	1yr	19	13	count=0 (<40mm)	count=0 (<30mm), count=3 (30-40)	count=6 (>2mm)			count=2 (<4mm), count=6 (4-6mm)							count=3	count=5
10	Villarreal et al[44]	mean=8.45mo, range=0mo - 33mo	10	74	mean=38.8, SD=5.71	mean=40.95, SD=4.13	count=8	count=15							count=3	count=2		
11	Haug and Assael[45]	minimum 6mo, range for Open=3.4mo - 52.4mo, range for Closed=34.8mo - 70.2mo	10	10	mean=46.9, SD=9.7	mean=42.5, SD=9.92	mean=.5, SD=1.08	mean=0.8, SD=0.92		**reported left and right**			mean=6.4, SD=3.31	mean=5.1, SD=2.42				
12	De Riu et al[46]	range for Open=5yr - 6yr, range for Closed=8yr - 12yr	20	19	mean=43.7, SD=5.9	mean=46, SD=7	count=6 (<3mm), count=2 (>3mm)	4 (<3mm), 2 (>3mm)	mean=8.6, SD=2.2	mean=8.6, SD=1.8	mean=8.5, SD=3.5	mean=7.5, SD=2.9	mean=7.4, SD=2.2	mean=6.3, SD=2.5			count=0	count=0
13	Joos and Kleinheinz[37]	10d, 6we, 3mo, 6mo, 12mo	25	26	mean=45	mean=41	0.2mm	1.2mm		**reported left and right**			mean=3.1	mean=5.1				

Abbreviations: O = open, C = closed, d = day, we = week, mo = month, yr = year, SD = standard deviation, frac = fractured, non = nonfractured

3.3. Maximum Mouth Opening (MMO)

MMO can be defined as a measurement between the upper and lower central incisor teeth when the patient opens the mouth as wide as possible. The raw data from each study are displayed in Table 7. Of the 13 studies considered, two contained only means of MMO but no measure of deviation so they could not be used in the meta-analysis.[36, 37] Three studies reported MMO as counts.[38-40] These studies used cutoffs of 30mm, 40mm, or both. Hidding and Wolf[38] reported that no patients had a MMO of less than 30mm, while Oezman and Mischkowski[39] reported that no patients had a MMO of less than 40mm. Widmark and Bagenholm[40] indicated that none were less than 40mm in the open group and none were less than 30mm in the closed group, but 3 patients were between 30mm and 40mm in the closed group. Thus, all three of these studies had no patients with a MMO less than 30mm in either group. The remaining 8 studies all reported continuous means with either a standard deviation, a range, or both.[34, 35, 41-46] Note that the study numbers in Table 7 correspond to the first column of study numbers in Table 6. In an effort to use the most data possible in the meta-analysis, studies that reported a range were given an estimated standard deviation (see equations (2.36) and (2.37)).

Table 7. MAXIMUM MOUTH OPENING RAW DATA (means)

Study	Maximum Mouth Opening	
	Open	Closed
1	46mm (range = 34-61)	50mm (range = 34-65)
2	41.57mm (range = 29-44)	46mm (range = 31-53)
3	45.5mm (SD = 7.3)	47mm (SD = 6.8)
4	39mm (range = 23-50)	39mm (range = 10-60)
5	39mm	50mm
6	count = 0 (<30mm)	count = 0 (<30mm)
7	count = 0 (<40mm)	count = 0 (<40mm)
8	45.7mm (SD = 9.4)	46mm (SD = 12.9)
9	count = 0 (<40mm)	count = 0 (<30mm), 3 (30mm-40mm)
10	38.8mm (SD = 5.71)	40.95mm (SD = 4.13)
11	46.9mm (SD = 9.7)	42.5mm (SD = 9.92)
12	43.7mm (SD = 5.9)	46mm (SD = 7)
13	45mm	41mm

Heterogeneity can be visually assessed by looking at the various effect size estimates with respect to the study sample size. It is expected that larger studies will tend toward the mean, while smaller studies may be more to the left or more to the right of the mean if the effect size is plotted on the horizontal axis. Figure 1 shows the effect sizes of the eight studies used in the analysis of mean MMO. The effect sizes are Hedges' *g*

standardized mean differences with the correction factor (see equation (2.48)). Figure 1 is referred to as a funnel plot; if the studies are relatively homogeneous it is expected that the smaller studies will be more spread out along the bottom of the plot and the larger studies will be close to the mean toward the top of the plot. If the studies are homogeneous this plot produces a funnel-like shape. If all the studies come from a single population then it makes sense to average the sample effect sizes to estimate the true population effect size.

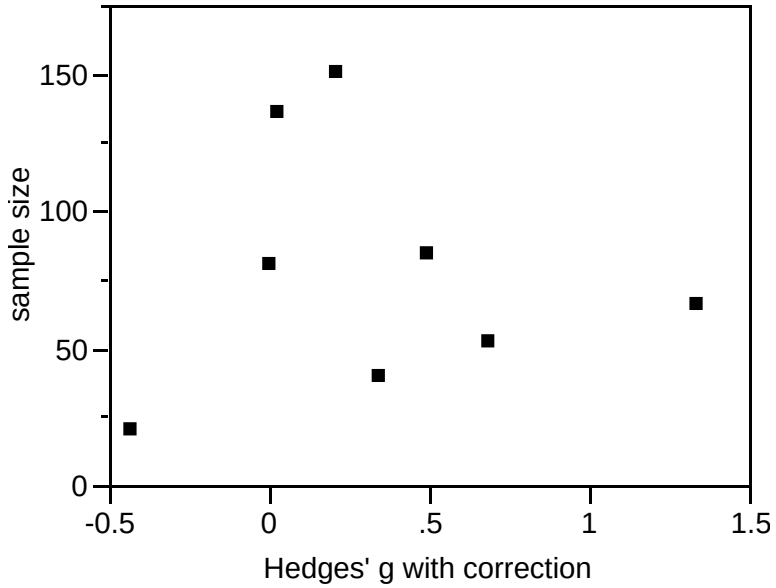

Figure 1. FUNNEL PLOT FOR MMO

The results shown in Figure 1 do not clearly indicate whether these effect sizes are homogeneous or not. With so few studies, it is difficult to get a true funnel shape. Plots can be very useful as exploratory tools, but a formal test of heterogeneity is necessary. The test for the Hedges' *g* estimators with the correction factor yields $Q = 23.511$ ($p < 0.0014$). Thus, we reject the null hypothesis of homogeneity, and proceed with the meta-analysis using random effects that adjusts for heterogeneity.

The weighted average method for random effects from section 2.2.2.2 was used to combine the Hedges' *g* estimates from each study. The overall treatment effect is $\overline{T}_{RND} = 0.349$ with a variance of $\text{var}(\overline{T}_{RND}) = 0.029$. A 95% confidence interval around $\overline{T}_{RND}$ is (0.015, 0.684). This interval does not contain zero, suggesting that the effect size is statistically significant. Cohen[47] has offered conventional values for "small", "medium", and "large" effects. For the standardized mean difference, these values are 0.2, 0.5, and 0.8, respectively. Thus, patients that underwent closed treatment had a moderately greater MMO on average than patients that underwent open surgery. These results can be seen in Figure 2. This plot shows the Hedges' *g* estimators for the individual studies as well as the overall weighted average with their 95% confidence intervals.

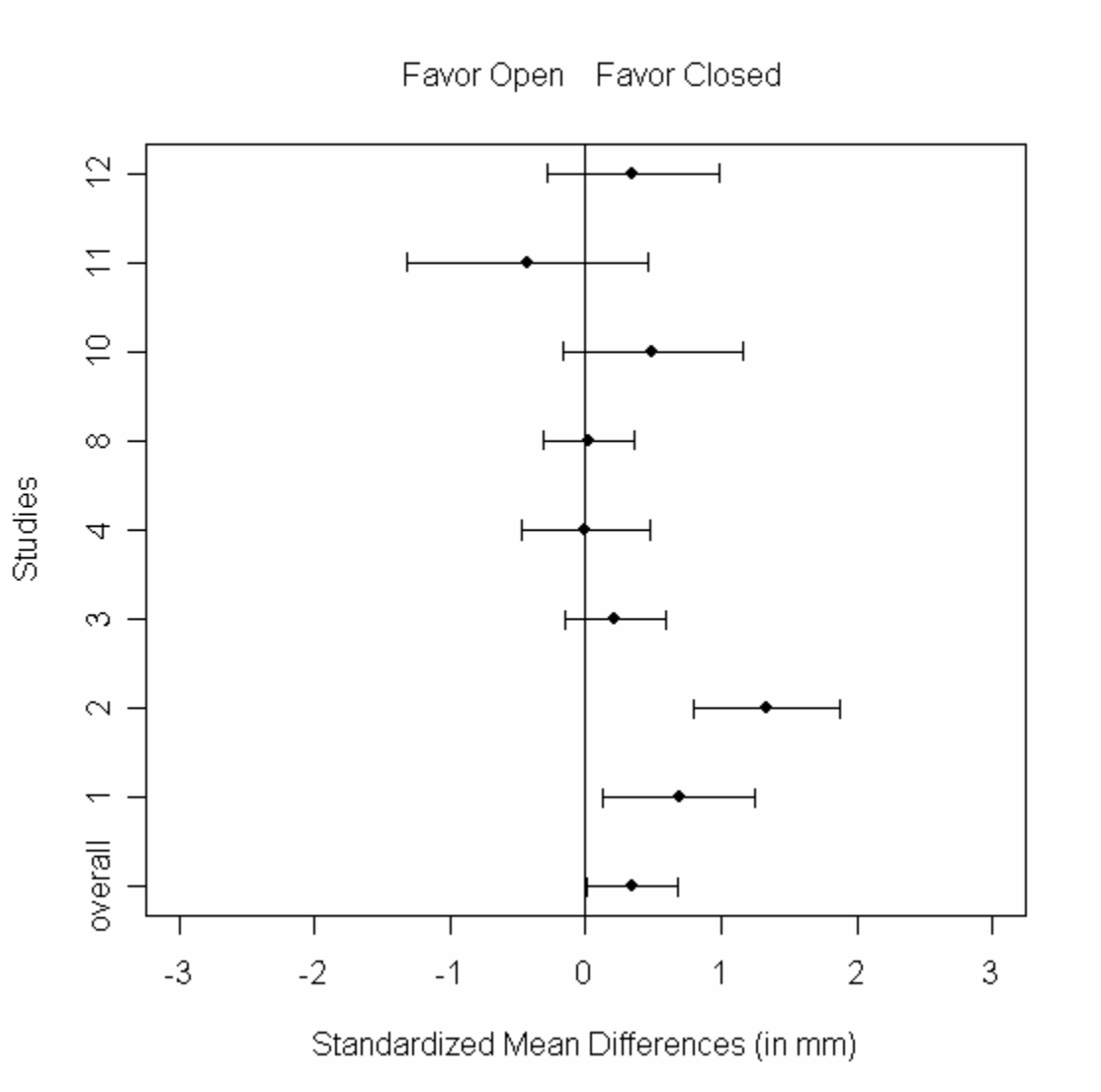

Figure 2. CONTINUOUS STANDARDIZED MEAN DIFFERENCES (HEDGES' *g*) FOR MMO

Because standardized mean differences are somewhat difficult to interpret, a separate figure was produced for mean differences. The mean differences along with their 95% confidence intervals are given in Table 8 and graphically displayed in Figure 3. The mean differences can be directly interpreted by the mm scale on the x-axis. For example, study 2 has a mean difference of 4.43 mm. This can be seen in Figure 3 and

interpreted as ‘closed treatment produced a MMO 4.43 mm greater on average than open treatment in study 2.’

Table 8. MEAN DIFFERENCES FOR MMO (in mm)

Study	Mean Difference (Closed – Open)	95% Confidence Interval	Hedges’ *g* (Standardized)	95% Confidence Interval
1	4.00	(2.698, 5.302)	0.69	(0.130, 1.252)
2	4.43	(3.553, 5.307)	1.34	(0.801, 1.872)
3	1.50	(0.523, 2.477)	0.22	(-0.157, 0.588)
4	0	(-1.180, 1.180)	0	(-0.468, 0.468)
8	0.30	(-0.841, 1.441)	0.03	(-0.311, 0.364)
10	2.15	(0.776, 3.524)	0.49	(-0.173, 1.156)
11	-4.40	(-7.146, -1.655)	-0.43	(-1.316, 0.457)
12	2.30	(0.704, 3.896)	0.35	(-0.284, 0.982)
Overall	1.48	(-0.018, 2.972)	0.35	(0.015, 0.684)

The pooled estimate for the mean differences suggests that, on average, closed treatment produces a MMO of about 1.48mm greater than open treatment.

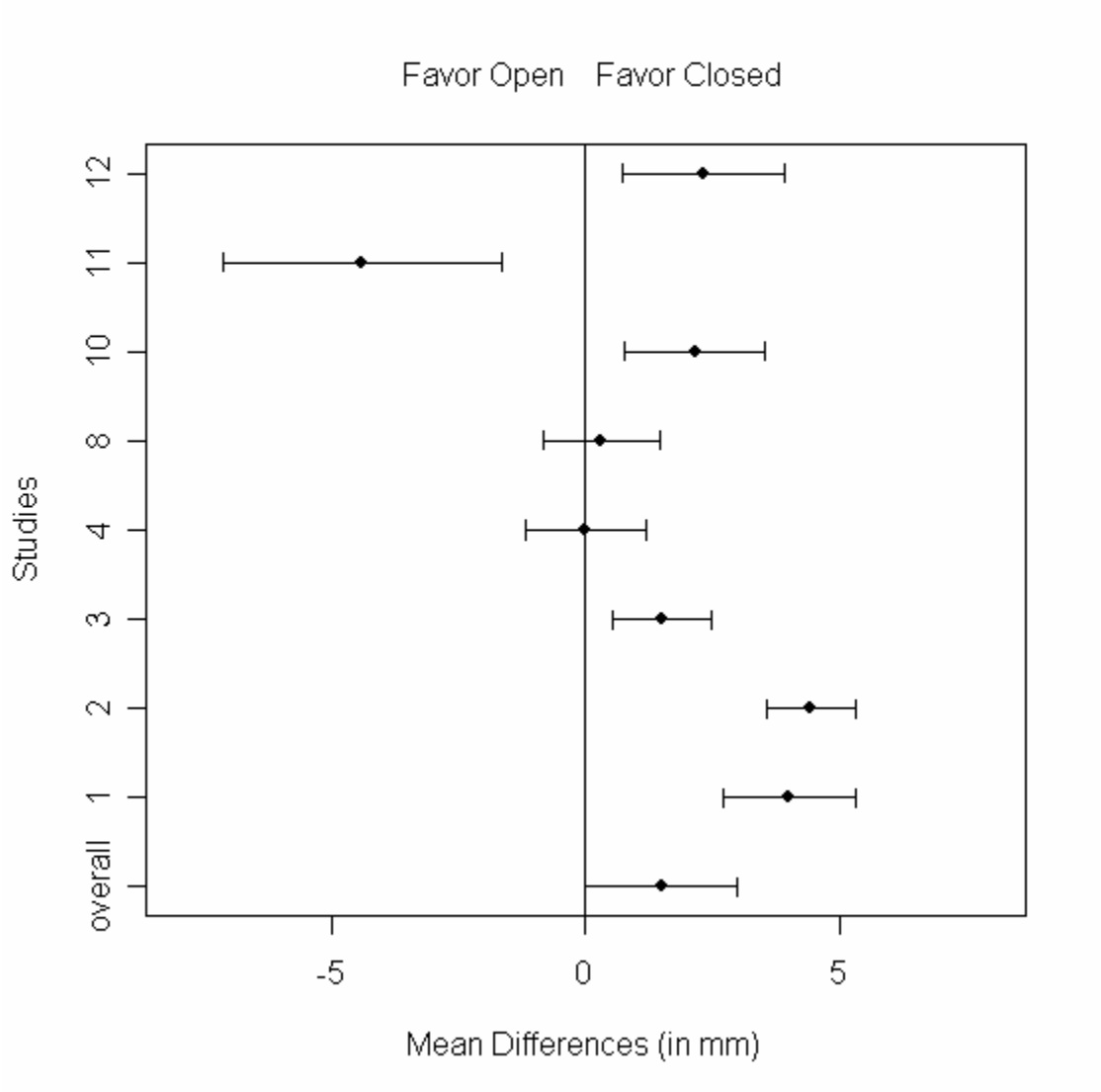

Figure 3. MEAN DIFFERENCES WITH 95% CONFIDENCE INTERVALS FOR MMO

Note that only 8 of the 13 studies were used in the continuous analyses. Three more studies could be added if frequency counts were included. There are methods available for converting continuous data into proportions or counts. Since the number of studies at hand is relatively small, it was decided to convert the continuous outcomes in an effort to use all of the data available. The eight studies that reported continuous

outcomes were converted into dichotomous variables using Suissa's method[31] described in section (2.2.3). The three studies that reported counts for both open and closed groups used a cutoff of 30mm. Thus, the cutoff value C was chosen to be 30. This value along with the means and standard deviations of the continuous outcomes were used in equations (2.53) and (2.54) to approximate proportions of patients in both groups that had a MMO greater than and less than 30mm. These proportions were multiplied by their respective sample sizes to give the cell counts. The cell counts and the resulting odds ratios are displayed in Table 9.

Table 9. CONVERTED AND UNCONVERTED COUNT DATA FOR MMO

	Count <30mm		Count ≥ 30mm		
Study	Open	Closed	Open	Closed	Odds Ratio
1	0.0443	0.0090	23.9557	27.991	5.778
2	0.0001	0.0010	35.9999	29.999	0.054
3	0.6240	0.7017	36.3760	112.298	2.745
4	1.1595	5.0200	24.8405	48.980	0.455
8	2.9412	7.9499	59.0588	66.050	0.414
10	0.6164	0.2967	9.3836	73.703	16.321
11	0.4073	1.0382	9.5927	8.962	0.367
12	0.2023	0.2116	19.7977	18.788	0.908
6	0	0	20	14	NA
7	0	0	20	10	NA
9	0	0	19	13	NA

The last three rows of Table 9 are the studies that were originally given as counts, and thus were not converted. The odds ratios for these three groups are undefined because there are zero cell counts. In addition, the studies with converted data have some cell counts that are close to zero. This is an obvious problem for the undefined odds ratios but also a problem for some of the others. For example, study 10 gives an odds ratio of 16.321. The counts for patients with an MMO less than 30mm are almost the same in both groups (and less than 1), but the counts for those greater than 30mm vary a great deal between the open and closed groups simply because the groups have unbalanced sample sizes. Thus, the odds ratio would be expected to be close to 1, but instead it is an inflated estimate due to the sample size disparity among the groups.

A solution to undefined and markedly inflated odds ratios is to add a continuity correction to boost zero and close to zero cell counts. All three of the suggested types of continuity corrections discussed in section 2.1.3 were imposed on the data for MMO (see Table 10). For each method, the total amount added to each treatment group typically sums to 1, such that $k_O + k_C = 1$. But more recently Agresti[22] and others suggest that adding smaller continuity corrections might provide better estimates for the true odds ratio. They suggest adding 0.01 to each treatment group instead of 1. Thus, each type of continuity correction was imposed under both of these conditions, where the amount added to each treatment group was 1 or 0.01. The six sets of odds ratios are displayed in Table 10.

Table 10. ODDS RATIOS FOR ALL 3 METHODS AFTER CONTINUITY CORRECTIONS FOR MMO

	$k_O + k_C = 1$			$k_O + k_C = 0.01$		
Study	Constant	Treatment	Empirical	Constant	Treatment	Empirical
1	1.246	1.080	1.080	4.128	3.986	3.986
2	0.834	0.998	0.998	0.702	0.827	0.827
3	2.861	1.847	1.847	2.748	2.727	2.727
4	0.587	0.514	0.514	0.457	0.456	0.456
6	0.707	1.000	1.000	0.700	1.000	1.000
7	0.512	1.000	1.000	0.500	1.000	1.000
8	0.455	0.448	0.448	0.414	0.414	0.414
9	0.692	1.000	1.000	0.684	1.000	1.000
10	10.521	4.902	4.902	16.173	15.880	15.880
11	0.553	0.553	0.553	0.369	0.369	0.369
12	0.938	0.971	0.971	0.908	0.910	0.910

The treatment method and empirical method odds ratios are identical. This is due to the use of the prior as 1 for the empirical distribution. Typically the prior is a pooled odds ratio of all of the studies without zero cell counts, but because most of the cells for counts less than 30mm were zero or very close to zero, the prior was set as 1. The treatment method continuity correction under the constraint $k_O + k_C = 1$ was used in the final analysis. Considering how small the cell counts were for the cells less than 30mm,

it appeared that the odds ratios from this method would be the best fit. Using a constant correction factor may produce an odds ratio that is simply a reflection of the sample size difference among groups, as shown in section 2.1.3.2. Since the treatment and empirical methods were identical for these data, the treatment method was chosen for simplicity.

The odds ratios were log transformed to approximate a normal distribution. The Q statistic for testing heterogeneity was calculated from the log-odds ratios from the treatment method corrected values, such that $Q = 2.886$ (p = 0.984). The null hypothesis of homogeneity is therefore not rejected. A Mantel-Haenszel test for fixed effects was used to test the cumulative effect size estimate. The common log-odds ratio is $\ln(\hat{\omega}_{MH}) = -0.345$ with a variance of $Var\{\ln(\hat{\omega}_{MH})\} = 0.151$. The corresponding 95% confidence interval is (-1.107, 0.418). Figure 4 shows the individual study results as well as the overall estimator. The interval around the common log-odds ratio contains zero, indicating that the effect size estimate is not different from zero, and a log-odds ratio of zero implies no difference among treatment groups. There is no statistically significant difference between the open and closed treatment groups from looking at the variable MMO dichotomized at 30mm.

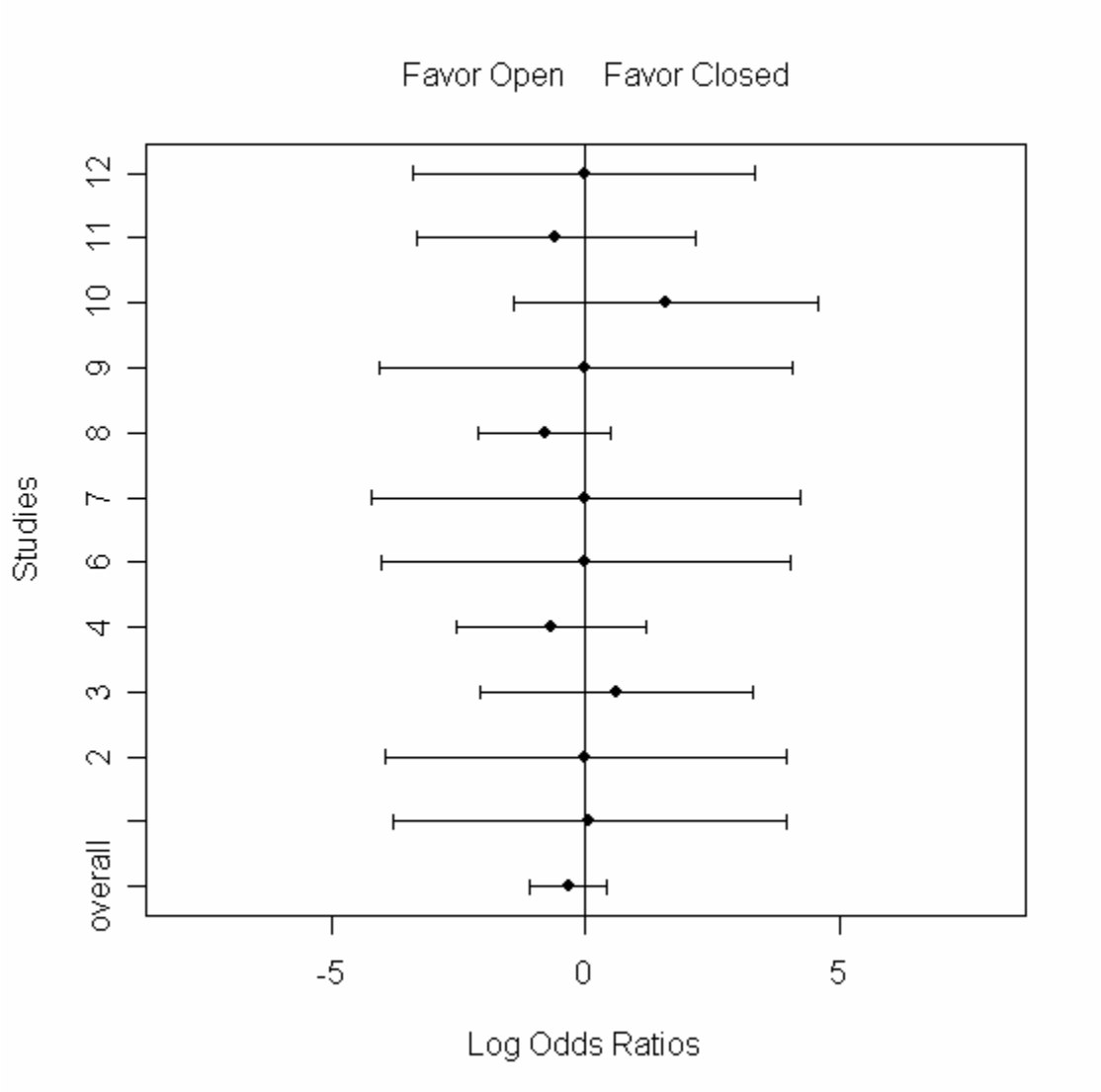

Figure 4. LOG-ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR MMO

Since odds ratios are easier to interpret than log-odds ratios, the odds ratios with 95% confidence intervals were calculated for each study included in the MMO categorical meta-analysis (see Table 11). They are further displayed in Figure 5. The x-axis is on a log-scale since some of the upper confidence limits were extremely large.

Every interval contains the value 1, suggesting that there is no difference between open and closed treatment for the individual studies.

Table 11. ODDS RATIOS AND LOG-ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR MMO

	Odds Ratio		Log-Odds Ratio	
Study	Estimate	95% CI	Estimate	95% CI
1	1.080	(0.023, 51.275)	0.077	(-3.784, 3.937)
2	0.998	(0.019, 52.506)	-0.002	(-3.965, 3.961)
3	1.847	(0.0126, 26.986)	0.614	(-2.068, 3.295)
4	0.514	(0.079, 3.333)	-0.665	(-2.534, 1.204)
6	1	(0.018, 56.769)	0	(-4.039, 4.039)
7	1	(0.015, 68.328)	0	(-4.224, 4.224)
8	0.448	(0.121, 1.652)	-0.804	(-2.110, 0.502)
9	1	(0.017, 57.443)	0	(-4.051, 4.051)
10	4.902	(0.246, 97.499)	1.590	(-1.401, 4.580)
11	0.553	(0.036, 8.581)	-0.592	(-3.334, 2.150)
12	0.971	(0.034, 27.832)	-0.029	(-3.384, 3.326)
overall	0.709	(0.331, 1.519)	-0.345	(-1.107, 0.418)

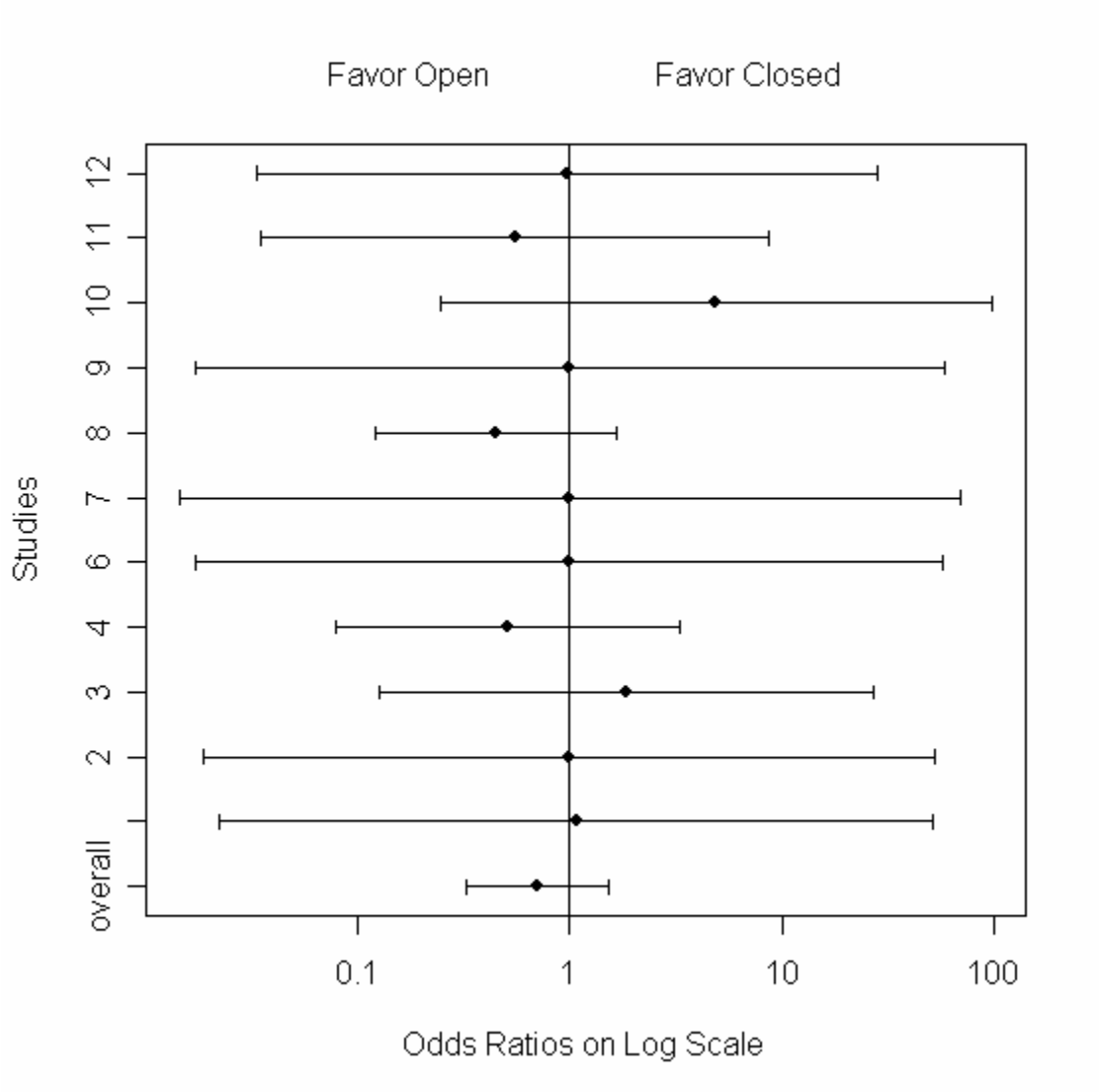

Figure 5. ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR MMO

It should be noted that substantial manipulation was performed on the data for the categorical analyses. Eight of the studies were converted to counts and continuity corrections were added to all of the studies to avoid zero cell counts. Because of all the alterations to the data, the differences between the treatment groups that were detected in the continuous analysis were minimized.

3.4. Deviation

Deviation can be defined as a shift of the lower jaw to one side as the mouth is opened. It is measured as the maximum distance that the jaw shifts away from the middle during opening. The studies did not indicate if deviation was observed on the fractured or non-fractured side of the face. It was assumed that deviation was on the fractured side. The raw data for the outcome variable deviation are displayed in Table 12. A total of eight studies contained measures for both groups, but the means given by Joos and Kleinheinz[37] could not be used because they didn't give standard deviations. Two of the studies gave means with standard deviations and the other 5 contained counts. It was decided to use a cutoff of 0mm for the studies that gave counts greater than a certain amount, since some studies gave no cutoff at all. Therefore, the proportions used in the meta-analysis were defined as those with any amount of deviation.

Table 12. **DEVIATION RAW DATA (means)**

Study	Deviation	
	Open	Closed
2	count = 8	count = 12
6	count = 2 (>3mm)	count = 9 (>3mm)
7	count = 0 (>2mm)	count = 0 (>2mm)
8	0.4mm (SD = 6.8)	4.2mm (SD = 6.6)
9	count = 6 (>2mm)	
10	count = 8	count = 15
11	0.5mm (SD = 1.08)	0.8mm (SD = 0.92)
12	count = 6 (<3mm), 2 (>3mm)	count = 4 (<3mm), 2 (>3mm)
13	0.2mm	1.2mm

Since only two studies reported continuous data, a meta-analysis for continuous data was not performed. A meta-analysis was done on the log-odds ratios of the data for this outcome variable. The two studies that reported continuous outcomes were converted into dichotomous variables using Suissa's method.[31] The cutoff value C was chosen to be 0. This value along with the means and standard deviations of the continuous outcomes were used in equations (2.53) and (2.54) to approximate proportions of patients in both groups that had a deviation greater than and less than 0mm. These proportions were multiplied by their respective sample sizes to give the cell counts. The cell counts and the resulting odds ratios are displayed in Table 13. The first

4 studies are the original count data, while the last 2 studies are the converted count data. The study that contained two zero cells counts was first excluded from the analysis since it showed no difference between treatment groups. If the following meta-analysis shows a treatment difference, study 7 could be included in a sensitivity analysis. However, if the meta-analysis excluding study 7 shows no difference, then it will be unnecessary to go to the trouble of continuity corrections to further prove the same result.

Table 13. **CONVERTED AND UNCONVERTED COUNT DATA FOR DEVIATION**

	Count > 0mm (had deviation)		Count = 0mm (had no deviation)			Log-Odds
Study	Open	Closed	Open	Closed	Odds Ratio	Ratios
2	8	12	28	18	2.333	0.847
6	2	9	18	5	16.200	2.785
10	8	15	2	59	0.064	-2.756
12	8	6	12	13	0.692	-0.368
8	32.451	54.592	29.546	19.408	2.561	0.940
11	6.783	8.077	3.217	1.923	1.992	0.689

The odds ratios were log transformed to approximate a normal distribution. Next, the Q statistic for testing heterogeneity was calculated from the log-odds ratios such that $Q = 24.494$ ($p < 0.0002$). The null hypothesis of homogeneity was therefore rejected. The funnel plot is displayed in Figure 6.

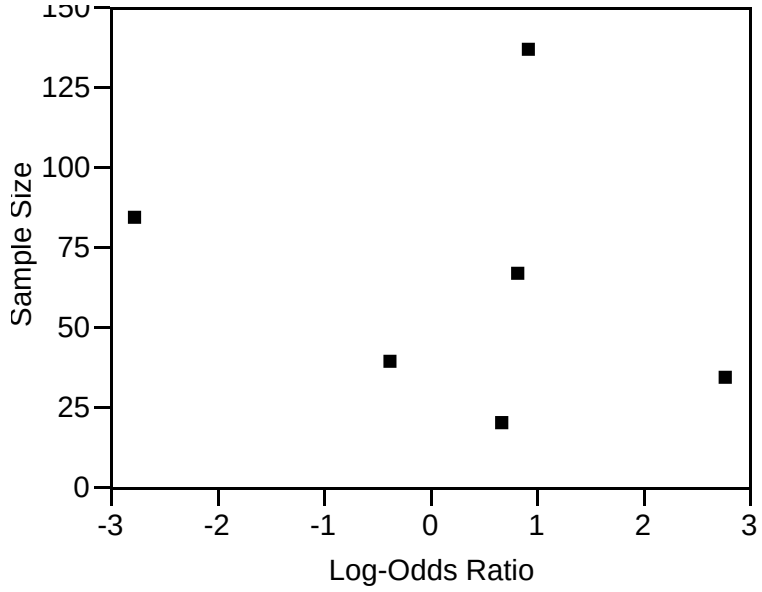

Figure 6. FUNNEL PLOT FOR DEVIATION

The weighted average method for random effects was used to test the cumulative effect size estimate. The estimated log-odds ratio is $\overline{T}_{RND} = 0.357$ with a 95% confidence interval of (-0.826, 1.540). This interval contains zero, indicating that the effect size estimate is not statistically significant. The odds that a patient who had closed treatment will have deviation are the same as the odds that a patient who had open treatment will have deviation. Although study 7 was not included in this analysis, it further supports the conclusion of no difference between groups and it would not have changed the results if it had been included in the analysis. Figure 7 shows the individual log-odds ratios and the common log-odds ratio with their corresponding 95% confidence intervals.

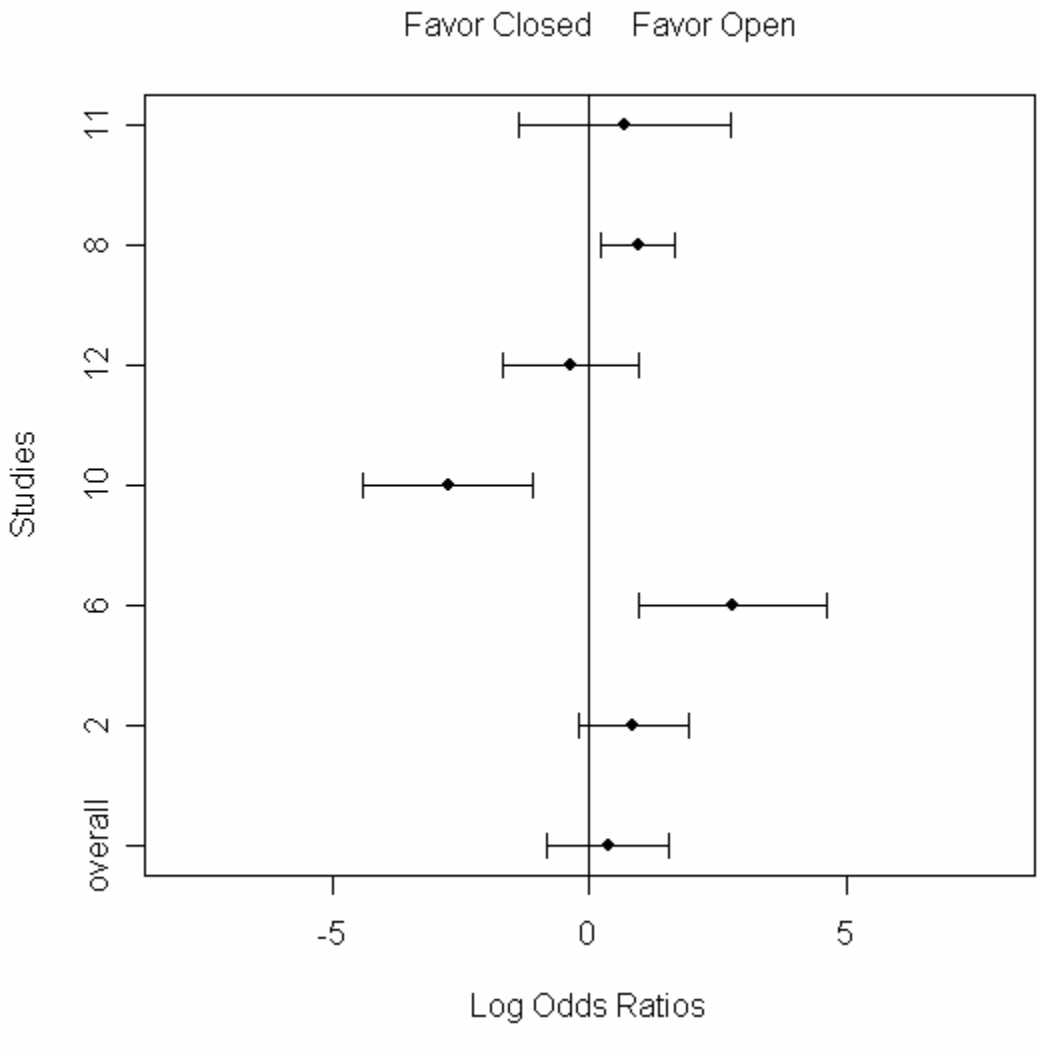

Figure 7. LOG-ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR DEVIATION

The odds ratios were also calculated for each study and they are given along with the log-odds ratios in Table 14. The odds ratios and their 95% confidence intervals are displayed in Figure 8.

Table 14. ODDS RATIOS AND LOG-ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR DEVIATION

	Odds Ratio		Log-Odds Ratio	
Study	Estimate	95% CI	Estimate	95% CI
2	2.333	(0.798, 6.822)	0.847	(-0.226, 1.920)
6	16.200	(2.613, 100.451)	2.785	(0.960, 4.610)
8	2.561	(1.248, 5.255)	0.940	(0.222, 1.659)
10	0.064	(0.012, 0.331)	-2.756	(-4.406, -1.106)
11	1.992	(0.255, 15.597)	0.689	(-1.368, 2.747)
12	0.692	(0.185, 2.585)	-0.369	(-1.685, 0.950)
Overall	1.429	(0.438, 4.665)	0.357	(-0.826, 1.540)

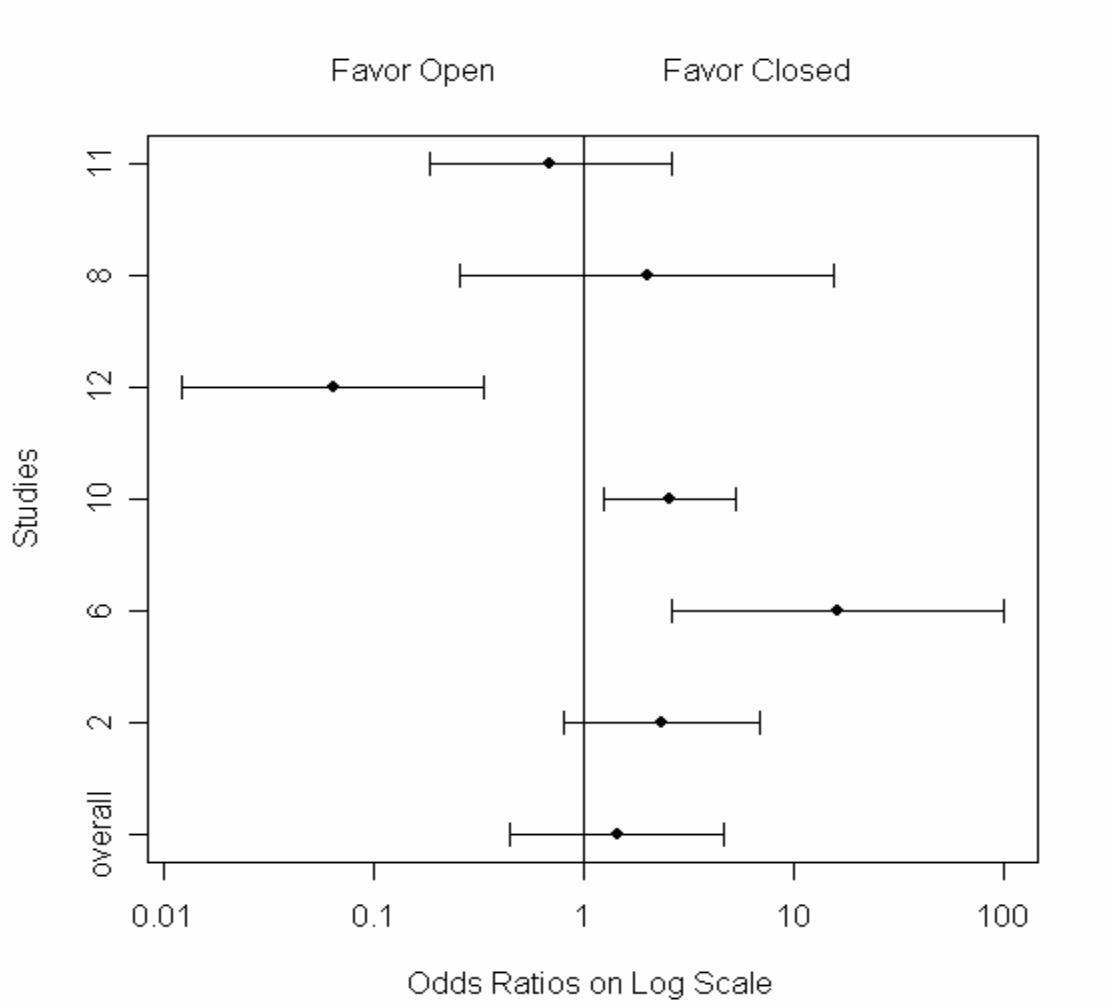

Figure 8. ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR DEVIATION

3.5. Lateral Excursion

Lateral excursion refers to movement of the lower jaw in a lateral direction. The raw data for the outcome variable excursion are given in Table 15. Most studies reported excursion to the fractured side of the face and to the non-fractured side of the face for each patient. However, two of the studies[37, 45] reported excursion as measurements taken from the left and right side of the face instead. These studies cannot be used in the meta-

analysis because it is unclear which side (left or right) was the fractured side. It is clinically meaningful to compare excursion to the fractured side among patients in the open and closed groups, as well as excursion to the non-fractured side. It is not helpful for a meta-analysis to have excursion recorded as “right side” and “left side”.

Study 3 only gave measurements to the non-fractured side. Thus, it was only included in the meta-analysis for the non-fractured side data. Studies 6 and 7 gave combined fractured and non-fractured side counts for the closed treatment group. In study 7, since the count was zero for the combined groups, it will also be zero for the individual groups. However, in study 6, the count is 1 and it is unknown if that one measurement of less than 5mm was taken toward the fractured or non-fractured side of the patient’s face. Therefore, study 6 had to be excluded from the analysis.

After the previously mentioned exclusions, the following studies could be used in a meta-analysis. Studies 1, 8, and 12 contained continuous means and standard deviations or ranges of excursion to both fractured and non-fractured sides. Study 3 contained means and standard deviations for excursion to the non-fractured side only. Study 7 contained counts for excursion to both sides, but they were all zero, which doesn’t lend much information. Two meta-analyses were performed on the continuous outcomes. The first analysis was for excursion to the fractured side data (including studies 1, 8, and 12) and the second was for excursion to the non-fractured side data (including studies 1, 3, 8, 12).

Table 15. **LATERAL EXCURSION RAW DATA (in mm)**

	Lateral Excursion			
	Open		Closed	
Study	Fractured	Non-fractured	Fractured	Non-fractured
1	mean = 10 range = 5-15	mean = 9 range = 4-18	mean = 9 range = 4-14	mean = 7 range = 3-12
3		mean = 8.5 sd = 3.3		mean = 8.7 sd = 3.4
6	count = 0 (<5mm)	count = 0 (<5mm)	count = 1 (<5mm)	
7	count = 0 (<6mm)	count = 0 (<6mm)	count = 0 (<6mm)	
8	mean = 10.9 sd= 2.5	mean = 10.1 sd = 2.8	mean = 10.3 sd = 3.6	mean = 9.4 sd = 3.5
9		count = 2 (<4mm), count = 6 (4-6mm)		
11	Reported as left side and right side.			
12	mean = 8.6 sd = 2.2	mean = 8.6 sd= 1.8	mean = 8.5 sd= 3.5	mean = 7.5 sd= 2.9
13	Reported as left side and right side.			

Before conducting the meta-analysis for excursion to the fractured side data, the range values for study 1 were converted to approximate standard deviations. Hedges' *g* estimators were calculated for the three studies. They were found to be homogeneous

($Q = 1.359$, $p = 0.715$). The weighted average method for fixed effects was used to test the overall effect size. The treatment effect was $\overline{T}_+ = 0.349$ with a variance of $\text{var}(\overline{T}_+) = 0.018$. The 95% confidence interval around $\overline{T}+$ is (-0.030, 0.495). This interval contains zero, suggesting that the effect size is not statistically significant. There is no difference between the open and closed treatment with respect to excursion to the fractured side. The individual Hedges' *g* estimators as well as the overall estimator and their 95% confidence intervals can be seen in Figure 9.

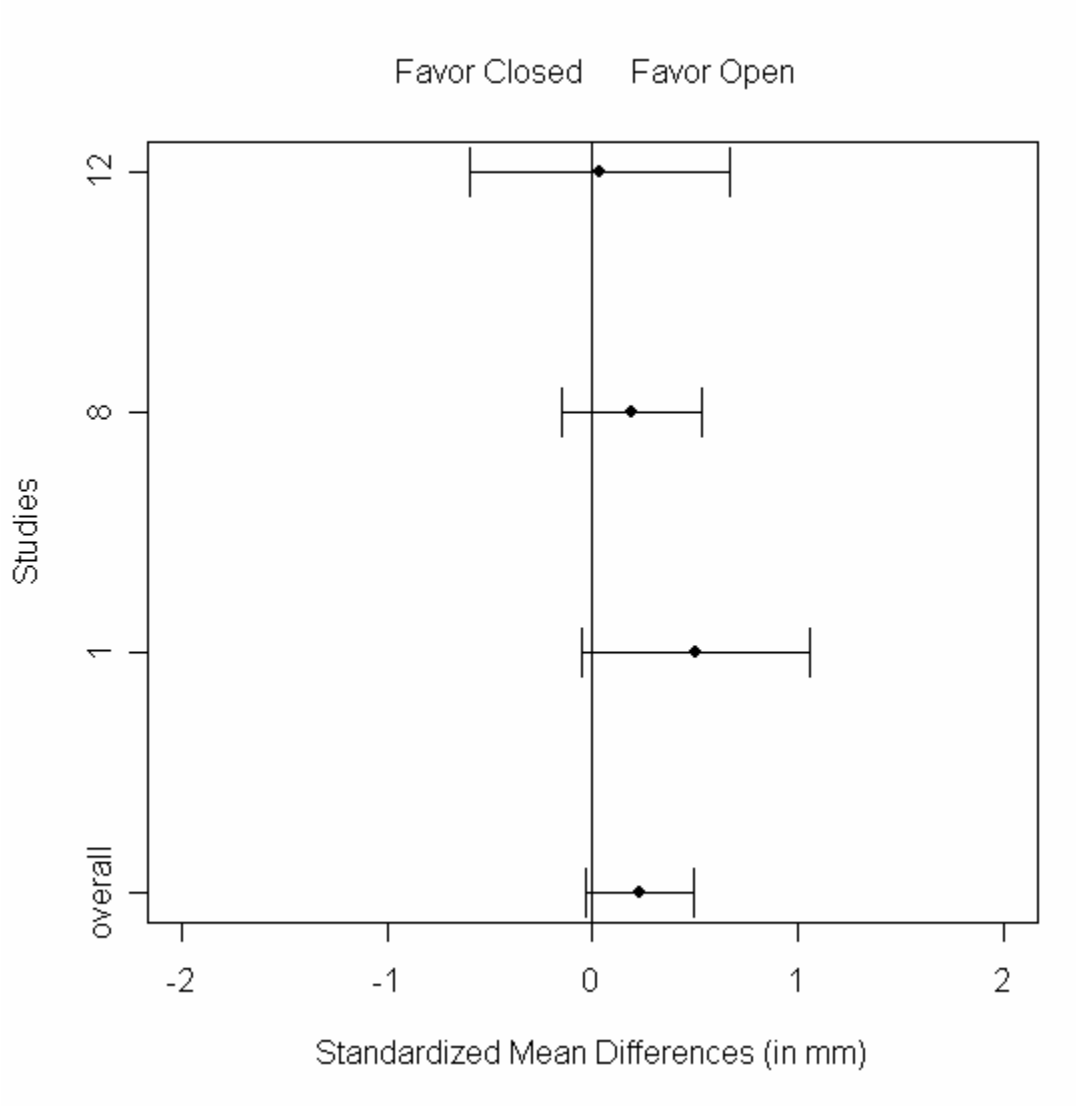

Figure 9. HEDGES' *g* WITH 95% CONFIDENCE INTERVALS FOR FRACTURED SIDE EXCURSION

Because standardized mean differences are somewhat difficult to interpret, a separate display was produced for mean differences. The mean differences along with their variances and 95% confidence intervals are given in Table 16 and graphically displayed in Figure 10. The mean differences can be interpreted directly on the mm scale.

Table 16. MEAN DIFFERENCES FOR FRACTURED SIDE EXCURSION (in mm)

Study	Mean Difference*	95% CI	Hedges' g	95% CI
1	1.0	(0.237, 1.764)	0.50	(0.051, 1.056)
8	0.6	(0.001, 1.199)	0.19	(-0.149, 0.528)
12	0.1	(-0.970, 1.170)	0.03	(-0.594, 0.662)
overall	0.65	(0.215, 1.078)	0.23	(-0.03, 0.50)

* Open - Closed

The individual and pooled mean differences with the corresponding confidence bounds are shown in Figure 10. On average, patients who receive open treatment had excursion to the fractured side of about 0.65mm greater than patients who receive closed treatment. This is not a large difference from a clinical point of view.

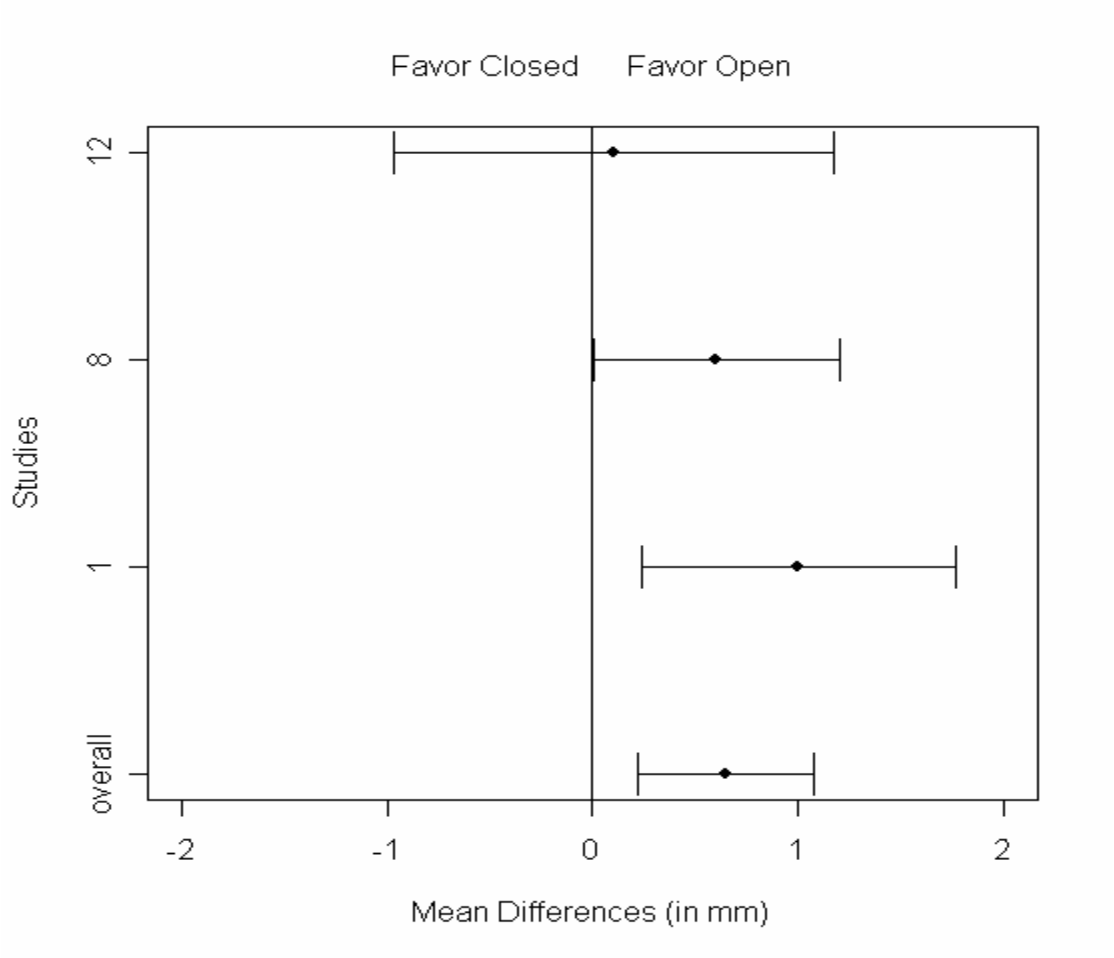

Figure 10. MEAN DIFFERENCES WITH 95% CONFIDENCE INTERVALS FOR FRACTURED SIDE EXCURSION

All data points for the non-fractured side were converted to Hedges' g standardized mean differences. They were found to be homogeneous $(Q = 7.391, p = 0.06)$ although Q was close to the critical value $Q_{CV} = 7.815$. The weighted average meta-analysis for fixed effects yielded a statistically significant effect. The treatment effect is $\overline{T}_+ = 0.242$ with a variance of $\text{var}(\overline{T}_+) = 0.012$. The corresponding 95% confidence interval is (0.027, 0.458). This interval does not contain

zero implying a statistically significant result. However, $\overline{T}_+ = 0.242$ suggests a relatively small effect size for a standardized mean difference, as indicated by Cohen.[47] Thus, patients that underwent open treatment had a slightly greater lateral excursion to the non-fractured side on average than patients that underwent closed surgery. The individual Hedges' *g* estimators as well as the overall estimator and their 95% confidence intervals can be seen in Figure 11.

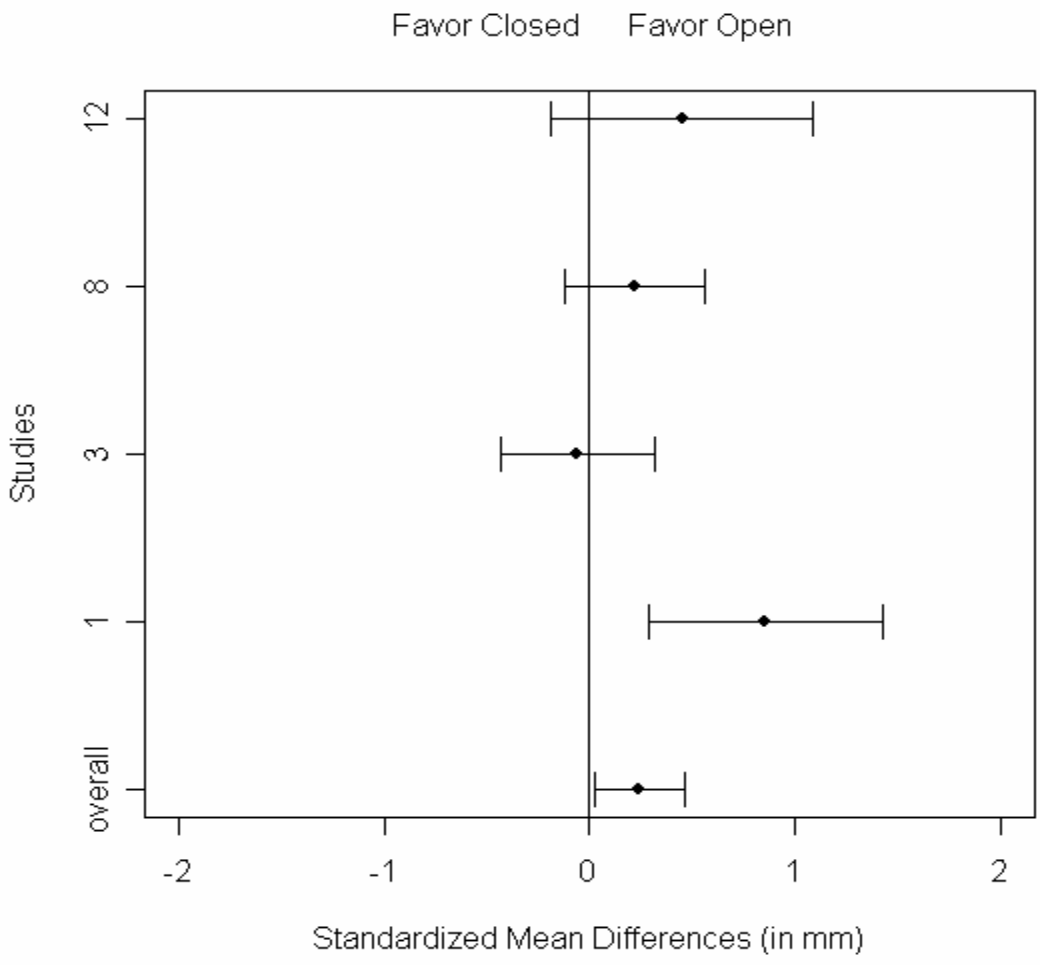

Figure 11. HEDGES' *g* WITH 95% CONFIDENCE INTERVALS FOR NON-FRACTURED SIDE EXCURSION

Because standardized mean differences are somewhat difficult to interpret, a separate analysis was done for mean differences. The mean differences along with their variances and 95% confidence intervals are given in Table 17 and graphically displayed in Figure 12. The mean differences can be directly interpreted on the mm scale. For example, a difference of 2 mm for study 1 indicates that open treatment produced an excursion 2 mm greater on average than closed treatment.

Table 17. MEAN DIFFERENCES FOR NON-FRACTURED SIDE EXCURSION (in mm)

Study	Mean Difference*	95% CI	Hedges' *g*	95% CI
1	2.0	(1.172, 2.828)	11.04	(8.848, 13.228)
3	-0.2	(-0.882, 0.482)	-1.64	(-2.058, -1.227)
8	0.7	(0.096, 1.304)	7.34	(6.402, 8.272)
12	1.1	(0.128, 2.073)	4.38	(3.220, 5.533)
overall	0.88	(-0.196, 1.957)	0.24	(0.027, 0.458)

* Open - Closed

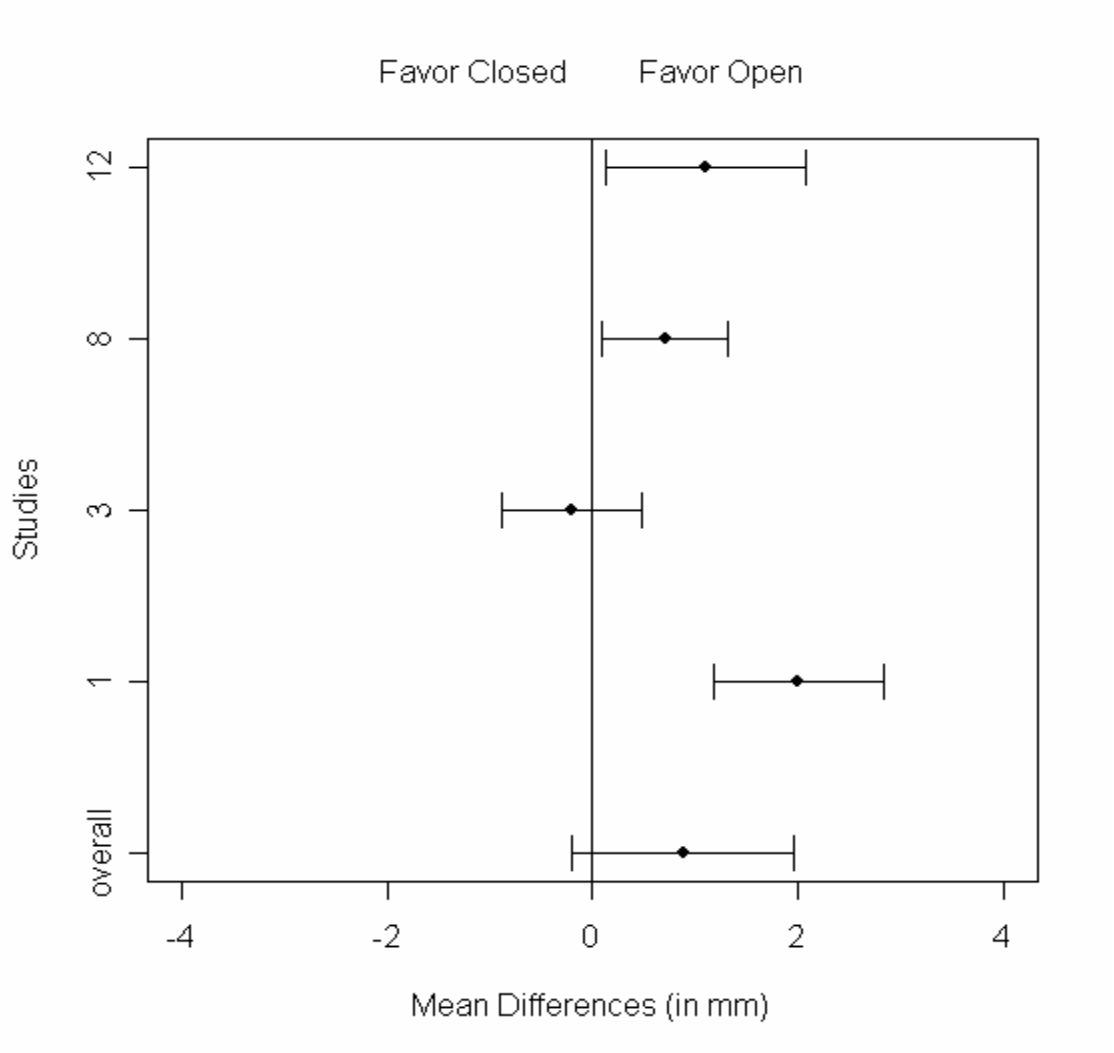

Figure 12. MEAN DIFFERENCES WITH 95% CONFIDENCE INTERVALS FOR NON-FRACTURED SIDE EXCURSION

3.6. Protrusion

Protrusion refers to movement of the lower jaw in a forward motion. This variable is measured as the maximum distance that the patient can move the jaw forward. The original data for this outcome variable are displayed in Table 18. Study 5 only gave data

for the open group and study 13 only gave means without standard deviations.[36, 37] These two studies were not included in the analysis.

Table 18. PROTRUSION RAW DATA (means)

	Protrusion	
Study	Open	Closed
1	7mm (range = 4-13)	7mm (range = 3-12)
3	5.9mm (SD = 2.3)	6.2mm (SD = 2.7)
4	count = 2	count = 3
5	9.5mm (SD = 2.1)	
6	count = 1 (<5mm)	count = 1 (<5mm)
7	count = 0 (<6mm)	count = 0 (<6mm)
8	8.3mm (SD = 2.8)	7.2mm (SD = 2.8)
11	6.4mm (SD = 3.31)	5.1mm (SD = 2.42)
12	7.4mm (SD = 2.2)	6.3mm (SD = 2.5)
13	3.1mm	5.1mm

The range values for study 1 were converted to approximate standard deviations. The five continuous means were evaluated first using Hedges' *g* estimators. They were found to be homogeneous ($Q = 5.340$,p=0.254). The weighted average method for fixed effects did not yield a statistically significant result. The treatment effect is $\overline{T}_+ = 0.184$ with a variance of $\text{var}(\overline{T}_+) = 0.011$. The 95% confidence interval around $\overline{T}+$ is (-0.025, 0.392). This interval contains zero, suggesting that the effect size is not statistically

significant. There was no difference in protrusion between the open and closed treatment groups (see Figure 13).

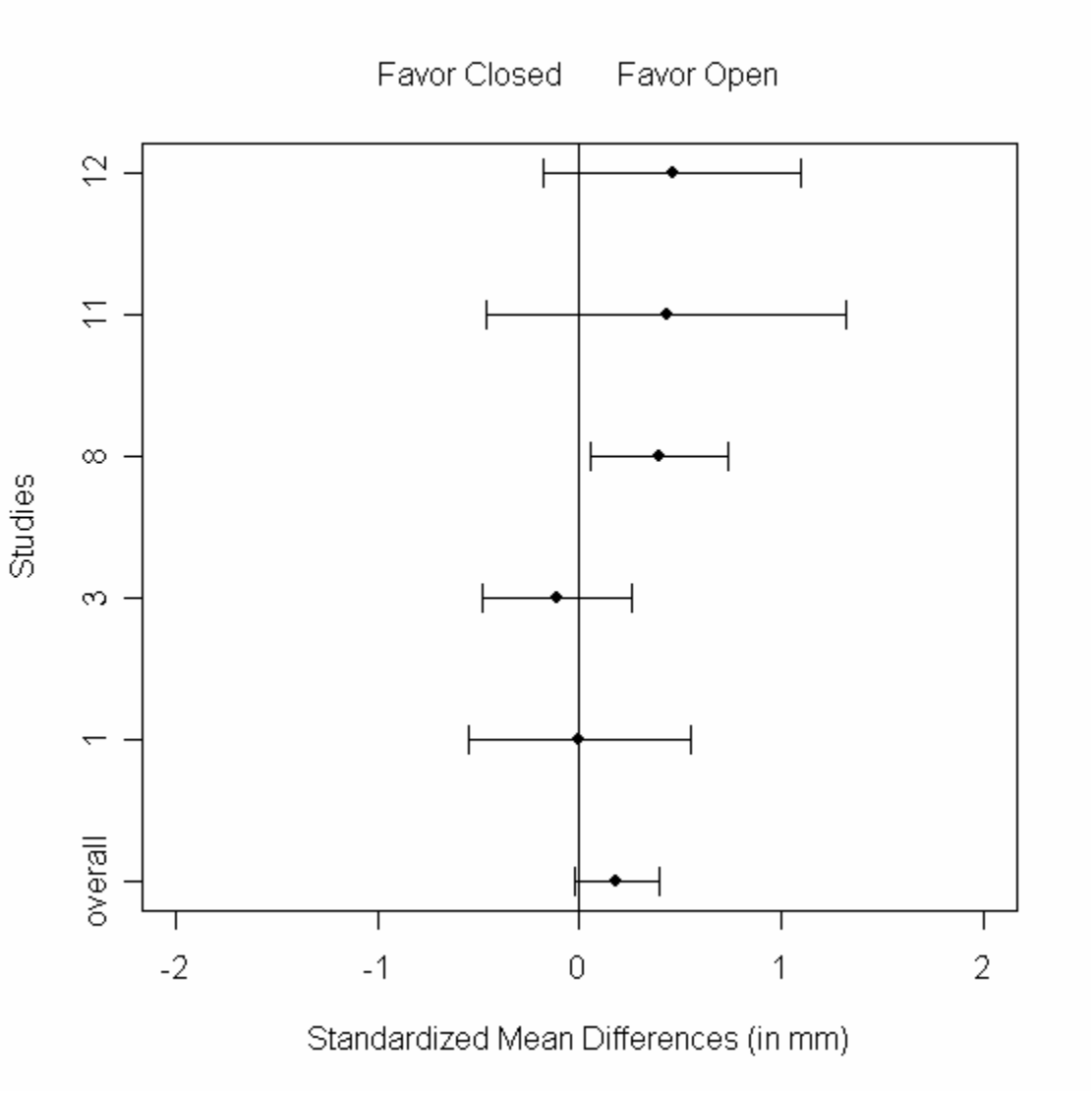

Figure 13. HEDGES' *g* WITH 95% CONFIDENCE INTERVALS FOR PROTRUSION

Because standardized mean differences are somewhat difficult to interpret, a separate analysis was done for mean differences. The mean differences along with their

variances and 95% confidence intervals are given in Table 19 and graphically displayed in Figure 14.

Table 19. MEAN DIFFERENCES FOR PROTRUSION (in mm)

Study	Mean Difference*	95% CI	Hedges' *g*	95% CI
1	0	(-0.724, 0.724)	0	(-0.545, 0.545)
3	-0.3	(-0.900, 0.300)	-3.19	(-3.707, -2.672)
8	1.1	(0.535, 1.665)	13.18	(11.577, 14.781)
11	1.3	(-0.193, 2.793)	2.15	(1.047, 3.247)
12	1.1	(0.137, 2.063)	4.47	(3.293, 5.640)
overall	0.58	(-0.289, 1.449)	0.18	(-0.025, 0.392)

* Open - Closed

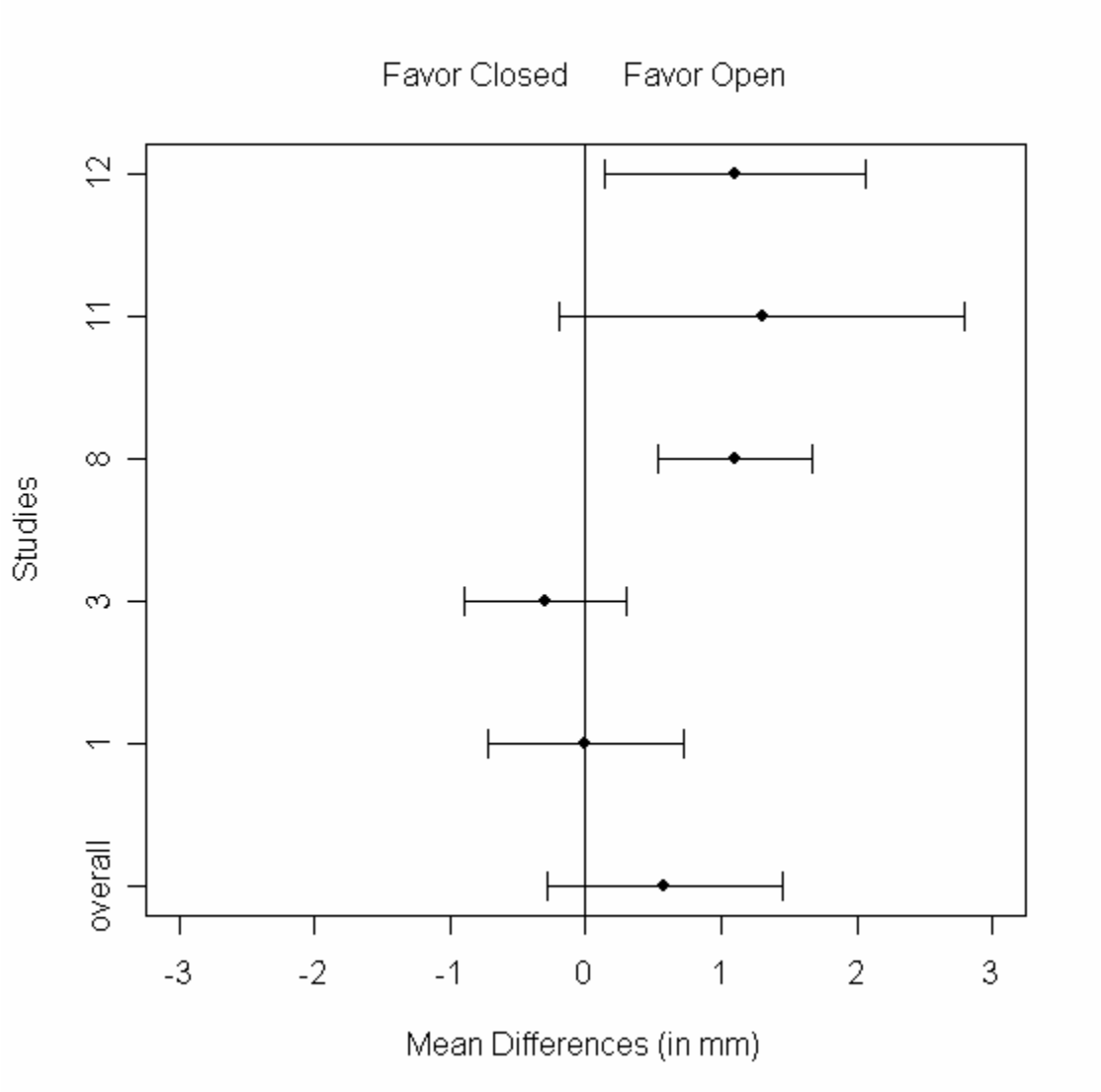

Figure 14. MEAN DIFFERENCES WITH 95% CONFIDENCE INTERVALS FOR PROTRUSION

Next, the continuous data was converted to cell counts using Suissa's method. Study 4 did not indicate a cutoff for protrusion so it could not be used in this analysis. Study 6 used a cutoff of 5mm and study 7 used a cutoff of 6mm. For study 7, if no patients had protrusion less than 6mm, then no patients had protrusion less than 5mm. Therefore, the cutoff was set to 5mm. The log-odds ratios were found to be

homogeneous $(Q = 2.811, p=0.729)$. The weighted average method for fixed effects did not yield a statistically significant result. The common log-odds ratio is $\overline{T}_+ = 0.307$ with a 95% confidence interval of (-0.197, 0.810). This interval contains zero, suggesting that the log-odds ratio is not statistically significant. There is no difference in protrusion between the open and closed treatment groups. These findings are shown in Figure 15. As was done with deviation, study 7 was excluded from the analysis at first to avoid continuity corrections to the entire dataset. Since the results showed no difference between open and closed groups, study 7 further supports this result and would not have made an impact if it had been included in the analysis.

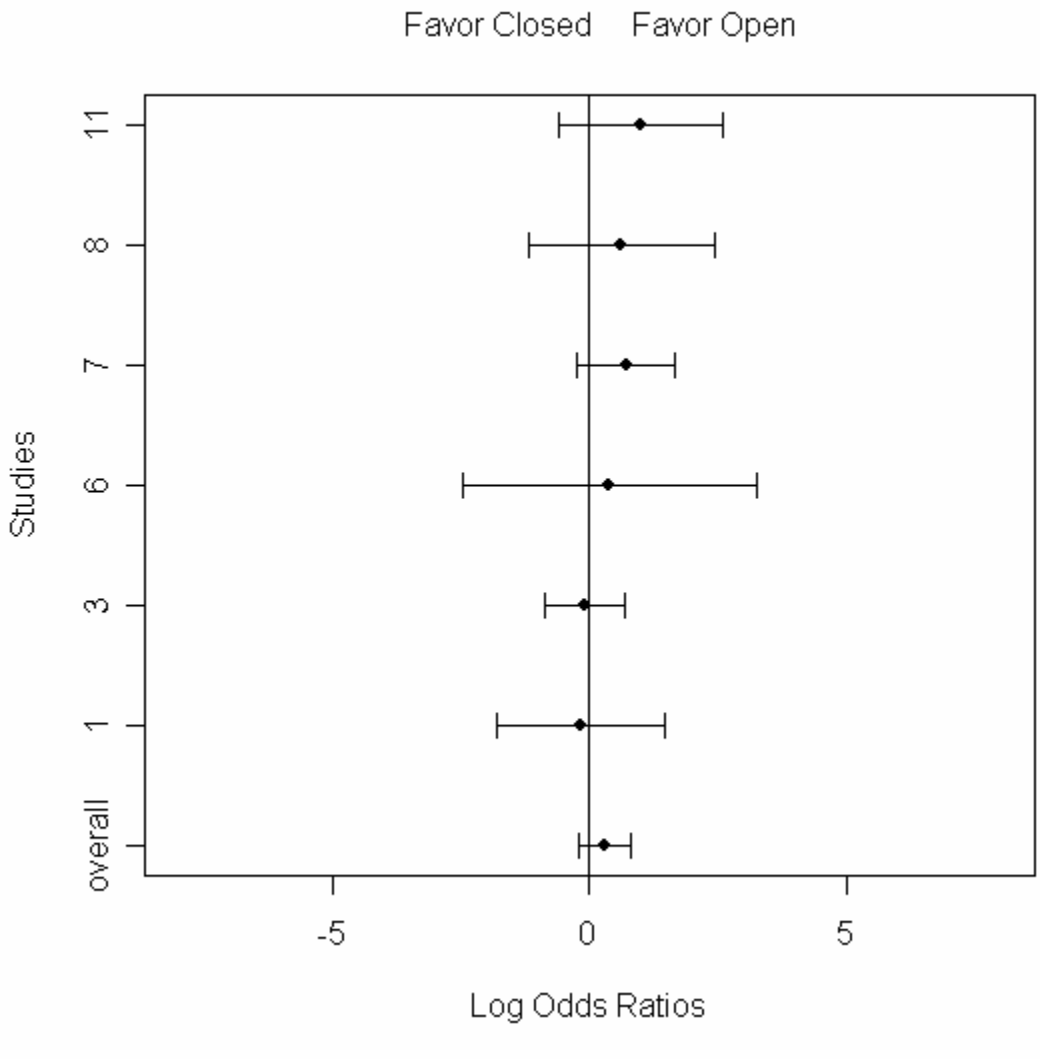

Figure 15. LOG-ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR PROTRUSION

The odds ratios were also calculated for each study and they are given along with the log-odds ratios in Table 14. The odds ratios with 95% confidence intervals are displayed in Figure 16. The odds ratios can be more easily interpreted than the log-odds ratios. For example, study 8 has an odds ratios of 2.034. This means that the odds of a

patient who had open surgery having protrusion greater than 5 mm are about twice the odds of a patient who had closed surgery having protrusion greater than 5mm.

Table 20. ODDS RATIOS AND LOG-ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR PROTRUSION

Study	Odds Ratio	Odds Confidence Interval	Log-Odds Ratio	Log-Odds Confidence Interval
1	0.849	(0.167, 4.319)	-0.163	(-1.790, 1.463)
3	0.917	(0.419, 2.005)	-0.087	(-0.869, 0.695)
6	1.462	(0.084, 25.527)	0.380	(-2.481, 3.240)
8	2.034	(0.789, 5.243)	0.710	(-0.237, 1.657)
11	1.849	(0.304, 11.246)	0.615	(-1.191, 2.420)
12	2.704	(0.543, 13.471)	0.995	(-0.611, 2.601)
Overall	1.359	(0.821, 2.248)	0.307	(-0.197, 0.810)

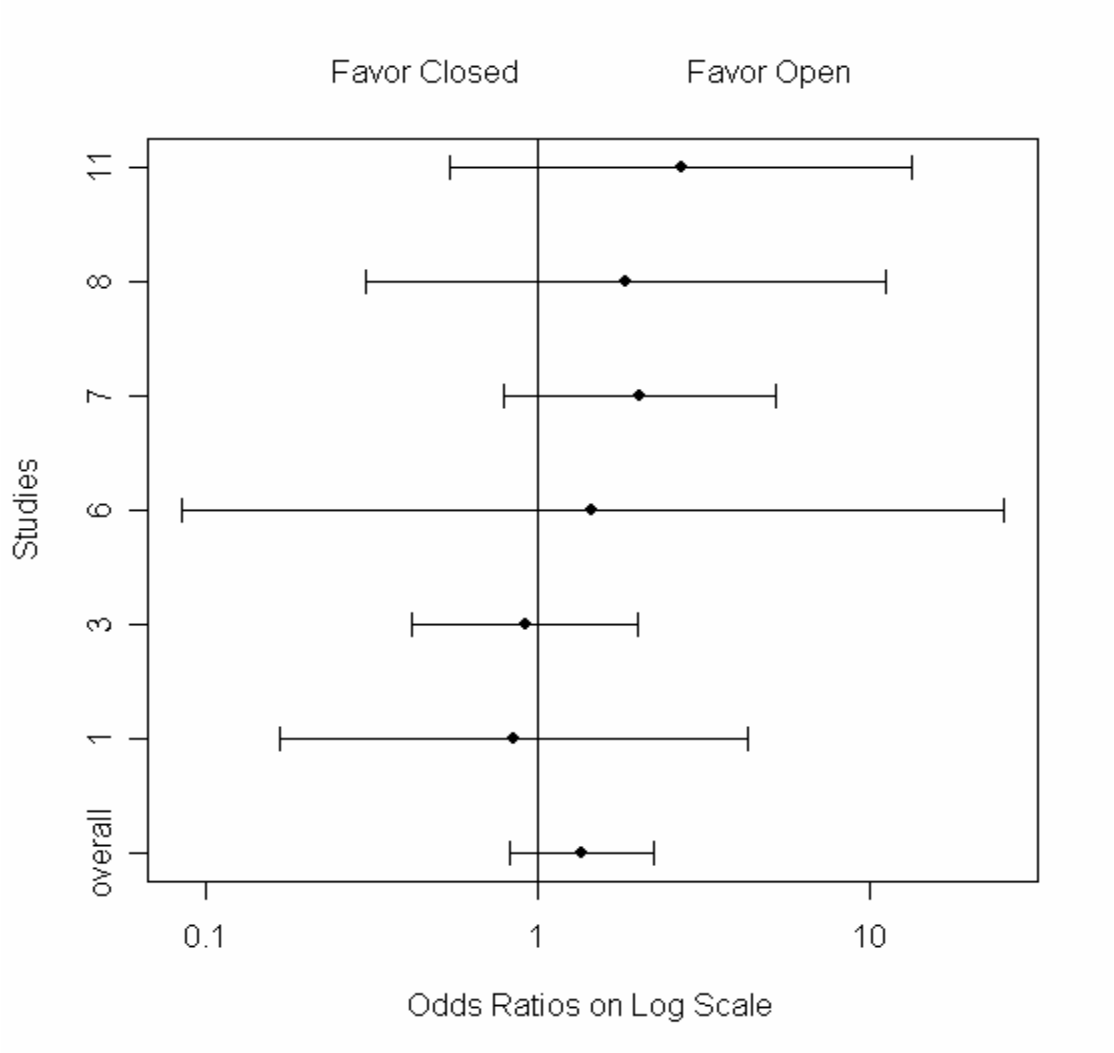

Figure 16. ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR PROTRUSION

3.7. Facial Asymmetry

Facial asymmetry is a difference in the shape of the face on the two sides. In this study it refers to the lower jaw not being centered beneath the upper jaw.

Table 21. ASYMMETRY RAW DATA

Study	Asymmetry	
	Open	Closed
1	count = 0	count = 3
2	count = 0	count = 0
3	count = 1	count = 1
10	count = 3	count = 2

All of the data available for asymmetry were given as counts. This variable is subjective because it wasn't quantitatively measured. Log-odds ratios were calculated for each study. Three out of the eight cells are zero, so two of the four studies had odds ratios that were undefined. Thus, continuity corrections were applied to the data.

Table 22. ODDS RATIOS FOR ALL 3 METHODS AFTER CONTINUITY CORRECTIONS FOR ASYMMETRY

	$k_O + k_C = 1$			$k_O + k_C = 0.01$		
Study	Constant	Treatment	Empirical	Constant	Treatment	Empirical
1	6.726	6.196	6.196	576.965	535.701	535.701
2	1.197	1.432	1.432	1.200	1.440	1.440
3	0.324	0.233	0.233	0.321	0.320	0.320
10	0.074	0.060	0.060	0.065	0.065	0.065

Table 22 shows highly inflated estimates for study 1 when the amount of correction per row was set to 0.01. As seen with MMO, the treatment and empirical values are identical because the priors were set to 1. The actual odds ratios before continuity corrections for the studies without zero cell counts (studies 3 and 10) were 0.321 and 0.065, respectively. These values most closely match the values from using the constant correction when $k_O + k_C = 1$. Thus, this continuity correction was used for the meta-analysis.

The log-odds ratio estimates were found to be heterogeneous $(Q = 9.707, p = 0.021)$. The weighted average method for random effects produced a common log-odds ratio of $\overline{T}_+ = -0.743$ with a 95% confidence interval of (-2.559, 1.072). This interval contains zero, indicating that there is no difference in asymmetry between open and close treatment groups. Again, there were only four studies in this analysis giving it low power. The individual log-odds ratios and the overall common log-odds ratio are displayed in Figure 17.

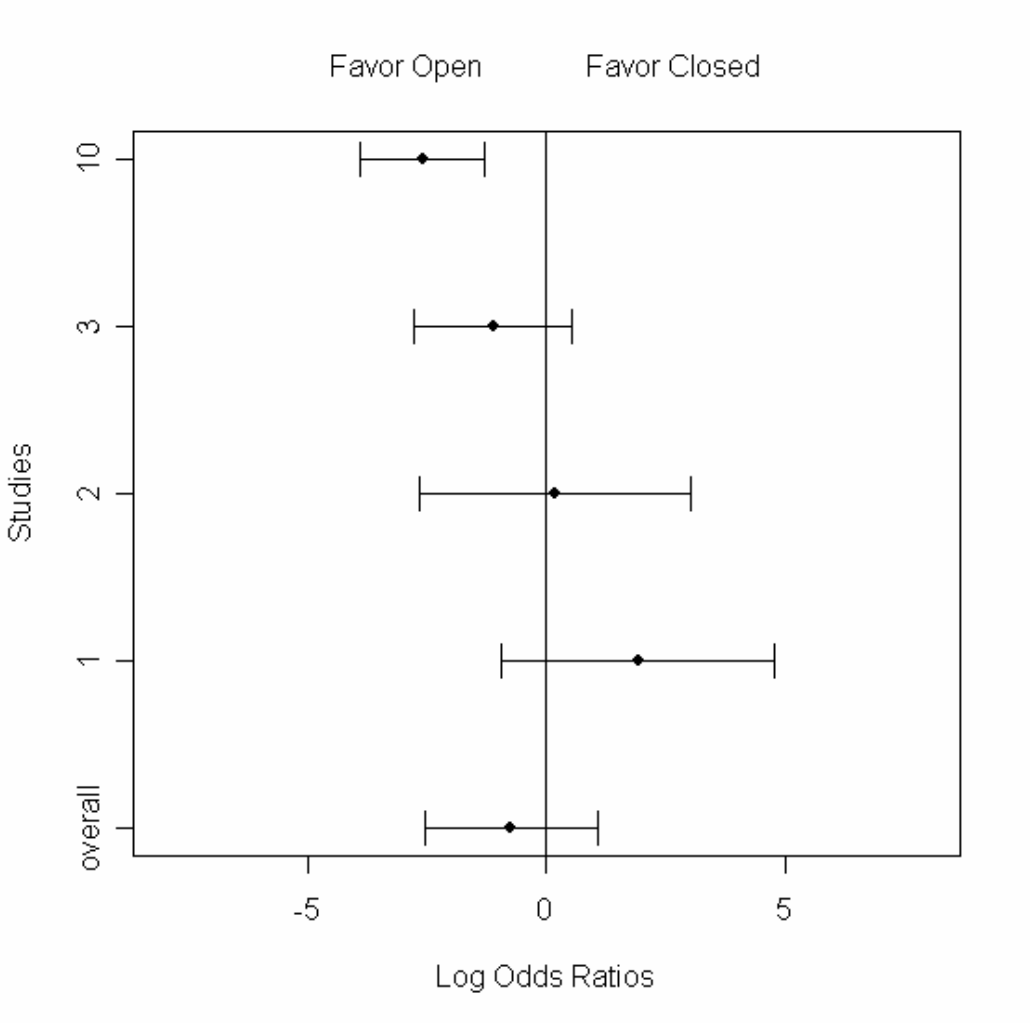

Figure 17. LOG-ODDS RATIOS AND 95% CONFIDENCE INTERVALS FOR ASYMMETRY

The odds ratios were also calculated for each study and they are given along with the log-odds ratios in Table 23. The odds ratios and 95% confidence intervals are displayed in Figure 18.

Table 23. ODDS RATIOS AND LOG-ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR ASYMMETRY

	Odds Ratios		Log-Odds Ratios	
Study	Estimate	95% CI	Estimate	95% CI
1	6.73	(0.388, 116.653)	1.91	(0.947, 4.759)
2	1.20	(0.070, 20.392)	0.18	(-2.656, 3.015)
3	0.32	(0.062, 1.696)	-1.13	(-2.779, 0.528)
10	0.07	(0.020, 0.274)	-2.61	(-3.915, -1.295)
overall	0.48	(0.077, 2.903)	-0.74	(-2.559, 1.072)

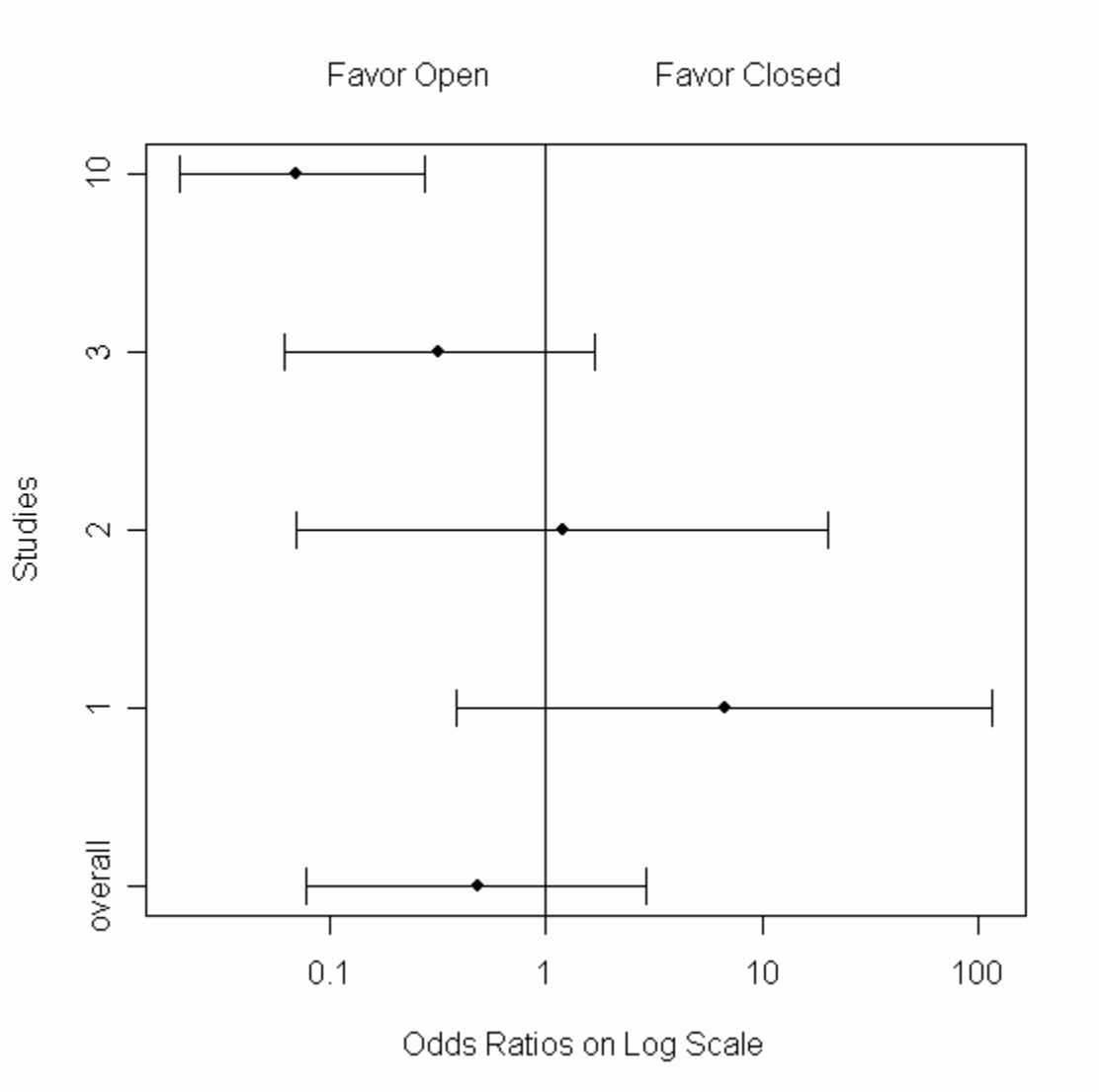

Figure 18. ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR ASYMMETRY

3.8. Joint or Muscle Pain

Temporomandibular joint pain is pain in the jaw joint that generally increases with jaw function. In some studies, joint or muscle pain was verified by reaction on palpation of the masticatory muscles. In others, the patients verbally expressed as whether or not they had joint pain since their surgery. Joint or muscle pain data was given in counts (see

Table 24). In study 3, counts were determined from given percentages. For the open group, Santler et al[42] reported that 2.7% of the patients had muscle pain. The total number of patients in the open group was 37, implying that 0.999 patients had pain. It assumed that this figure should be one and that this small difference is due to rounding. However, it was reported that 3.3% of patients in the closed group had muscle pain. The total number of patients in this group is 113, implying that 3.729 patients had pain. For this analysis, this figure was assumed to be 4. However, it is not entirely clear that this is a correct assumption.

Table 24. JOINT OR MUSCLE PAIN RAW DATA

	Joint or Muscle Pain	
Study	Open	Closed
1	count = 1	count = 6
2	count = 2	count = 5
3	count = 1	count ~ 4
9	count = 3	count = 5

Log-odds ratios were calculated for each study. The study estimates were found to be homogeneous $(Q = 0.976, p = 0.807)$. The weighted average method for fixed effects produced a common log-odds ratio of 1.16 with a 95% confidence interval of (0.210, 2.109). This interval does not contain zero, indicating that there is a difference in joint pain between open and closed treatment. The common odds ratio is 3.19 with a 95% confidence interval of (1.234, 8.240). Specifically, the odds of patients in the closed

group having joint or muscle pain are over 3 times as great as the odds of patients in the open group having joint or muscle pain. The individual and common log-odds ratios are displayed in Figure 19.

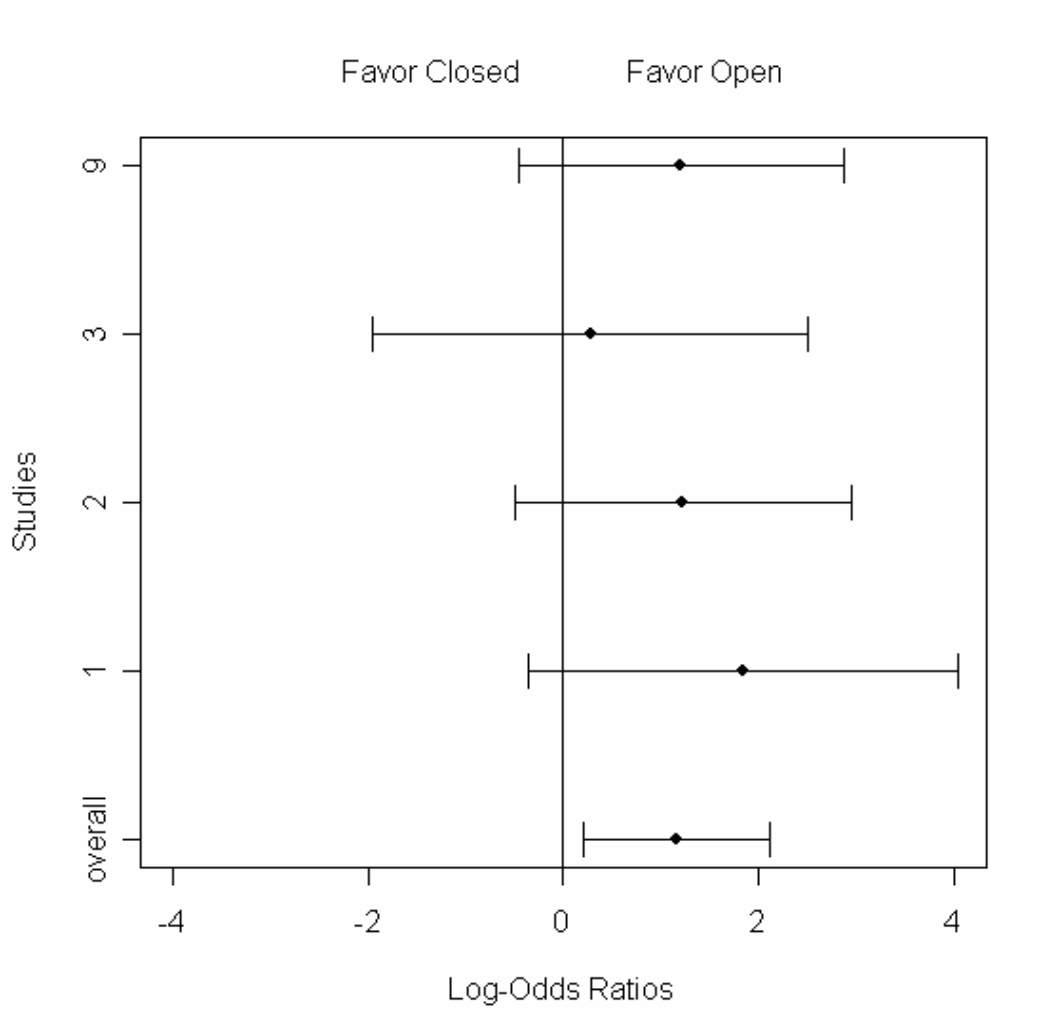

Figure 19. LOG-ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR JOINT OR MUSCLE PAIN

The odds ratios were also calculated for each study and they are given along with the log-odds ratios in Table 25. The odds ratios and 95% confidence intervals are displayed in Figure 20.

Table 25. ODDS RATIOS AND LOG-ODDS RATIOS WITH 95% CONFIDENCE INTERVALS FOR JOINT OR MUSCLE PAIN

Study	Odds Ratio	95% CI	Log-Odds Ratio	95% CI
1	6.27	(2.046, 19.235)	1.836	(0.716, 2.957)
2	3.40	(1.414, 8.174)	1.224	(0.347, 2.101)
3	1.32	(0.425, 4.108)	0.279	(-0.856, 1.413)
9	3.33	(1.426, 7.791)	1.204	(0.355, 2.053)
Overall	3.19	(1.234, 8.240)	1.159	(0.210, 2.109)

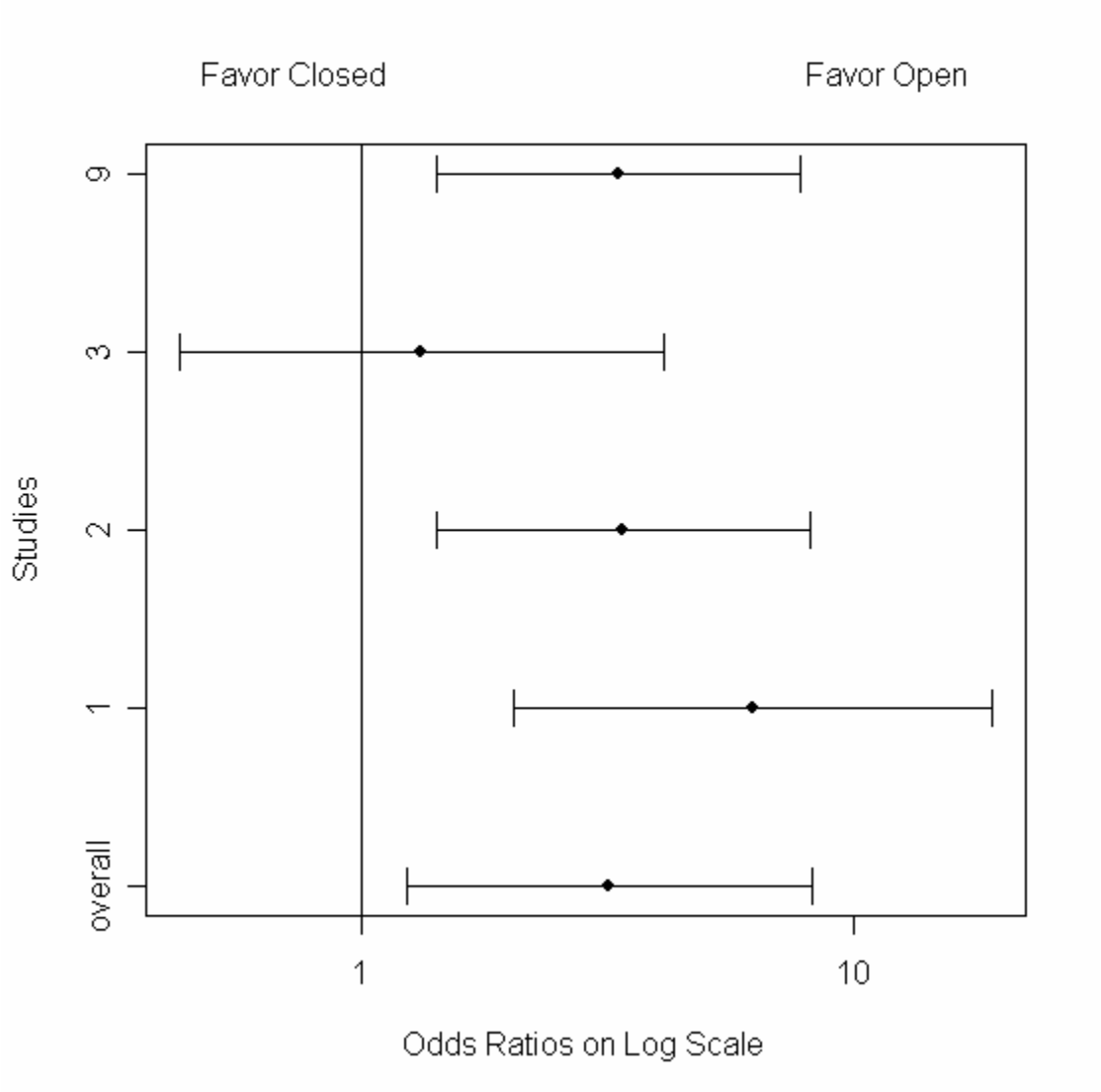

Figure 20. ODDS RATIOS AND 95% CONFIDENCE INTERVALS FOR JOINT OR MUSCLE PAIN

3.9. Summary of Results

In summary, the majority of the variables did not show differences among patients treated with open surgery versus patients treated with closed surgery. Table 26 gives the overall results from all of the meta-analyses performed. Three of the analyses yielded a

statistically significant result. Two of the three favored open treatment and one favored closed treatment. The rest of the results were inconclusive.

Table 26. SUMMARY OF RESULTS FROM META-ANALYSES

Outcome Measure		Method	Summary Statistic	Result
MMO		WA random	Hedges' $g = 0.349$	Favor Closed
		M-H fixed	$OR_{MH} = 0.709$	Neither
Deviation		WA random	log OR = 0.357	Neither
Excursion	Fractured	WA fixed	Hedges' $g = 0.349$	Neither
	Non-fractured	WA fixed	Hedges' $g = 0.242$	Favor Open
Protrusion		WA fixed	Hedges' $g = 0.011$	Neither
		WA fixed	log OR = 0.307	Neither
Asymmetry		WA random	log OR = -0.743	Neither
Joint or Muscle Pain		WA fixed	log OR = 1.159	Favor Open

Abbreviations: WA = weighted average, OR = odds ratios, and M-H = Mantel-Haenszel

4 Discussion

4.1. Conclusions

It is increasingly popular to use meta-analyses in all aspects of medical literature. The benefits from conducting a meta-analysis are potentially large. Individual studies may present conflicting views on a specific medical treatment and a meta-analysis can serve to combine the information from pertinent studies to form an overall conclusion. Therefore, meta-analysis techniques appear to be especially useful in a clinical setting where there is no clear evidence-based treatment. By including a group of studies into a meta-analysis, power for detecting an overall treatment effect is increased, making it easier to detect a difference that may exist in the population.

The available statistical methods tend to be somewhat rigid and underdeveloped. A great deal of manipulation was performed on the data for this series of meta-analyses. From converting continuous outcomes to continuity corrections, it is quite possible that some information was lost in this lengthy process. The actual results from the analyses are perhaps a good indicator for the existing population effects, but certainly not conclusive or well substantiated. The analyses were performed on all variables in an attempt to show how a meta-analysis could be applied for this research question. The findings are merely to give direction for future research and not to be taken as validated findings because the number of limitations is large. However, the results will be explored in a general sense.

It was discovered that the closed treatment group had a higher MMO than the open treatment group when the outcome measure was evaluated continuously as a

standardized mean difference (Hedges' *g*). However, it appeared that the effect size was not large and there were only 8 studies included in the analysis. When looking at excursion to the non-fractured side of the face, it was discovered that patients who had open treatment performed slightly better than patients who had closed treatment. These results should be interpreted with caution, as there were only 4 studies in this analysis. The same caution should be noted for the results from the analysis of joint or muscle pain. While the effect size seems substantial in favor of open treatment, the number of studies was only 4 as well.

A meta-analysis is only as good as the studies that comprise it. If the individual studies are flawed, the findings from a review of these studies will also be flawed. The trials included in a meta-analysis should ideally be of high methodological quality and free from bias, such that the differences in outcomes observed between groups of patients can be confidently attributed to the intervention under investigation.[5] There are several types of potential biases including: systematic differences in the patients' characteristics as baseline (selection bias), unequal provision of care apart from the treatment under evaluation (performance bias), biased assessment of outcomes (detection bias), and bias due to exclusion of patients after they have been allocated to treatment groups (attrition bias).[5]

4.2. Limitations

The single most important limitation with the data for this research was that only one study actually randomized patients to treatment groups. This introduces selection bias since the manner in which patients are assigned to a treatment group is not random.

Ideally, randomization would tend to make the two treatment groups relatively equal with respect to all baseline characteristics. Since only one of the studies used random assignment, the generalizability of these individual observational studies is, at best, suspect. Thus, a meta-analysis of suspect studies should be interpreted with caution. That is, these results should not be taken as definitive.

Some studies indicated that they classified type of fracture before treatment, but not all of the studies did. Thus, this information could not be included. If all future studies would include the type of fracture in a systematic classification scheme, then this information could be included (accounted for) in a meta-analysis. In addition, many pre-operative variables could be included in an analysis if researchers gathered this information in their studies. For example, it would be helpful to know if a patient with asymmetry after treatment may have had asymmetry before treatment as well. It is important to know the state of the patient before treatment in order to determine if the outcomes were due to treatment.

With regard to performance bias, several studies reported different ways of conducting open and closed surgery. There were differences in type of materials used as well as surgical protocols. Not all of the studies even gave a detailed description of what the surgical protocol was. There was variation in whether or not studies used elastics to fix the jaws postoperatively and whether surgeons had their patients perform jaw exercises after surgery and for how long. The length of time for closed treatment, known as maxillomandibular fixation (MMF), was also variable (ranging from 0 to 6 weeks).

Detection bias (bias in outcome assessment) was almost certainly present in several studies. For the most part, outcome measures were given in millimeters, but asymmetry and joint pain data were collected subjectively. Perhaps the amount of asymmetry could be measured. It also may be informative to assess these outcomes on an ordinal scale.

Attrition bias was also present in nearly all of the studies included in the meta-analyses. Many patients were lost to follow-up. Getting patients to come back for subsequent visits after surgery can be difficult, especially since the time between surgery and follow-up can be lengthy in order to see long-term results.

Publication bias is always a limitation when conducting a systematic review. Interesting or favorable results are more likely to be submitted and published. The funnel plot can be a useful tool to detect publication bias. If publication bias is present there will be a lack of small studies with negative results. The funnel plots for the previous analyses did not appear to have publication bias, but because there were so few studies, it was difficult to tell. It is not entirely clear how to best proceed when publication bias is present. Rosenthal presented a method of estimating how many 'typical' unpublished non-significant studies would have to exist to overturn the current pooled results.[48]

4.3. Future Suggestions

Future studies need to randomize patients to treatment groups and report results measured by examiners blind to treatment group. These two aspects of a clinical trial are fundamental to ever being able to conduct a reliable meta-analysis. Certain surgeons likely favor either open or closed treatment and so they may (even unintentionally) give

good ratings to a patient who had their treatment of preference. It is necessary that the person who measures the outcome variables on the patients at follow-up does not know which patients are in which treatment group.

Data reporting was rather varied among studies. Some studies reported an outcome as continuous while others reported it as a count or proportion. In many situations it is straightforward to report certain measures as proportions or counts, such as number of deaths out of total number treated. But often the outcome variable is measured on a continuous scale. When dichotomizing data, a cutoff point is chosen and the data are then reported as being either below or above the cutoff. Perhaps it is only of interest to divide the data into these two groups because they diverge so greatly at the cutoff point. The argument here is that the risk of a defined event is a clinically more meaningful measure of the extent of disease than the mean of the continuous outcome variable from which the disease is defined.[31]

However, there is a cost associated with this process. First, there is a loss of information when data is collapsed over a continuous scale. Dichotomizing over a continuous scale is essentially lumping together data into a group that has variation. Data that are not really the same are treated to be the same and the chance of finding factors to be significant is lessened. Second, the misclassification related to measurement error is increased. Lastly, the choice of the cutoff point can greatly influence the results if many points lie near the cutoff or if the selection of the cutoff is made after the data is collected.

Many of the studies in this research dichotomized MMO at 30mm. That is, they had counts or proportions of patients with a MMO great than 30mm and less than 30mm. Nearly every patient in every study had an MMO greater than 30mm. When a variable is dichotomized so that nearly all of the observations fall into one category, i.e. MMO greater than 30 mm, there is no way to identify which treatment group is better. Although a MMO greater than 30 mm may be considered minimally acceptable, by setting a higher cutoff (in the middle of the data), it may be possible to show if one treatment is outperforming the other. Better yet, reporting the data with means and standard deviations provides the most information and allows for the most accurate comparison of groups. It is strongly recommended that all studies in this field of literature report the sample size of each treatment group along with the means and standard deviations for each group or the proportions, including the numerator and denominator. If possible, it is also helpful to include means and standard deviations along with proportions if the data are available.

5 List of References

List of References

1. Normand SL: Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med 18(3):321-59, 1999

2. Pearson K: Report on certain enteric fever inoculation statistics. Br Med J 3:1243-6, 1904

3. Yusuf S, Peto R, Lewis J, Collins R, Sleight P: Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog Cardiovasc Dis 27(5):335-71, 1985

4. Mann C: Can meta-analysis make policy. Science 266:960-2, 1994

5. Egger M, Smith GD, Sterne JA: Uses and abuses of meta-analysis. Clin Med 1(6):478-84, 2001

6. Thornley B, Adams C: Content and quality of 2000 controlled trials in schizophrenia over 50 years. Bmj 317(7167):1181-4, 1998

7. Bangert-Drowns RL: Review of developments in meta-analytic method. Psychological Bulletin 99:388-99, 1986

8. Birge R: The calculation of errors by the method of least squares. Phys Rev 16:1-32, 1932

9. Cochran W: Problems arising in the analysis of a series of similar experiments. J Roy Statist Soc 4((Supplement)):102-18, 1937

10. Sutton AJ, Keith, R.A., Jones, D.R., Sheldon, T.A., Song, F.: Methods for Meta-Analysis in Medical Research. 2000

11. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22(4):719-48, 1959

12. Robins J, Breslow N, Greenland S: Estimators of the Mantel-Haenszel variance consistent in both sparse data and large-strata limiting models. Biometrics 42(2):311-23, 1986

13. Robins J, Greenland S, Breslow NE: A general estimator for the variance of the Mantel-Haenszel odds ratio. Am J Epidemiol 124(5):719-23, 1986

14. Fleiss JL: Statistical Methods for Rate and Proportions. 2nd ed., 1981

15. Greenland S, Salvan A: Bias in the one-step method for pooling study results. Stat Med 9(3):247-52, 1990

16. Sutton AJ CN, Lambert PC, Jones DR, Abrams KR, Sweeting MJ: Meta-analysis of rare and adverse event data. Expert Review in Pharmacoeconomics Outcomes Research 2:367-79, 2002

17. Sweeting MJ, Sutton AJ, Lambert PC: What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med 23(9):1351-75, 2004

18. Whitehead A, Whitehead J: A general parametric approach to the meta-analysis of randomized clinical trials. Stat Med 10(11):1665-77, 1991

19. Cook DJ, Witt LG, Cook RJ, Guyatt GH: Stress ulcer prophylaxis in the critically ill: a meta-analysis. Am J Med 91(5):519-27, 1991

20. Sankey SS WL, Fine MJ, Kapoor W: An assessment of the use of the continuity correction for sparse data in metaanalysis. Communications in Statistics-Simulation and Computation 25:1031-56, 1996

21. Cox D. The analysis of binary data. London: Methuen & Co Ltd; 1970.

22. Agresti A: An Introduction to Categorical Data Analysis. 1996

23. van Belle G: Statistical Rules of Thumb. 2002

24. Glass GV: Primary, Secondary, and Meta-Analysis of Research. Educational Researcher 5:3-8, 1976

25. Wang MB, BJ. Integrating results through meta-analytic review using SAS software. Cary: SAS Institute Inc.; 1999.

26. Hedges LVO, I.: Statistical Methods for Meta-Analysis. 1985

27. Cohen J: Statistical Power Analysis for the Behavioral Sciences. 1969

28. Hedges LV: Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. Journal of Educational Statistics 6:107-28, 1981

29. Hedges LV: Estimation of Effect Size from a Series of Independent Experiments. Psychological Bulletin 92:490-9, 1982

30. Woolf B: On estimating the relation between blood group and disease. Ann Hum Genet 19(4):251-3, 1955

31. Suissa S: Binary methods for continuous outcomes: a parametric alternative. J Clin Epidemiol 44(3):241-8, 1991

32. DerSimonian R, Laird N: Meta-analysis in clinical trials. Control Clin Trials 7(3):177-88, 1986

33. Palmieri C, Ellis E, 3rd, Throckmorton G: Mandibular motion after closed and open treatment of unilateral mandibular condylar process fractures. J Oral Maxillofac Surg 57(7):764-75; discussion 75-6, 1999

34. Throckmorton GS, Ellis E, 3rd: Recovery of mandibular motion after closed and open treatment of unilateral mandibular condylar process fractures. Int J Oral Maxillofac Surg 29(6):421-7, 2000

35. Worsaae N, Thorn JJ: Surgical versus nonsurgical treatment of unilateral dislocated low subcondylar fractures: a clinical study of 52 cases. J Oral Maxillofac Surg 52(4):353-60; discussion 60-1, 1994

36. Takenoshita Y, Ishibashi H, Oka M: Comparison of functional recovery after nonsurgical and surgical treatment of condylar fractures. J Oral Maxillofac Surg 48(11):1191-5, 1990

37. Joos U, Kleinheinz J: Therapy of condylar neck fractures. Int J Oral Maxillofac Surg 27(4):247-54, 1998

38. Hidding J, Wolf R, Pingel D: Surgical versus non-surgical treatment of fractures of the articular process of the mandible. J Craniomaxillofac Surg 20(8):345-7, 1992

39. Oezmen Y, Mischkowski RA, Lenzen J, Fischbach R: MRI examination of the TMJ and functional results after conservative and surgical treatment of mandibular condyle fractures. Int J Oral Maxillofac Surg 27(1):33-7, 1998

40. Widmark G, Bagenholm T, Kahnberg KE, Lindahl L: Open reduction of subcondylar fractures. A study of functional rehabilitation. Int J Oral Maxillofac Surg 25(2):107-11, 1996

41. Yang WG, Chen CT, Tsay PK, Chen YR: Functional results of unilateral mandibular condylar process fractures after open and closed treatment. J Trauma 52(3):498-503, 2002

42. Santler G, Karcher H, Ruda C, Kole E: Fractures of the condylar process: surgical versus nonsurgical treatment. J Oral Maxillofac Surg 57(4):392-7; discussion 7-8, 1999

43. Konstantinovic VS, Dimitrijevic B: Surgical versus conservative treatment of unilateral condylar process fractures: clinical and radiographic evaluation of 80 patients. J Oral Maxillofac Surg 50(4):349-52; discussion 52-3, 1992

44. Villarreal PM, Monje F, Junquera LM, Mateo J, Morillo AJ, Gonzalez C: Mandibular condyle fractures: determinants of treatment and outcome. J Oral Maxillofac Surg 62(2):155-63, 2004

45. Haug RH, Assael LA: Outcomes of open versus closed treatment of mandibular subcondylar fractures. J Oral Maxillofac Surg 59(4):370-5; discussion 5-6, 2001

46. De Riu G, Gamba U, Anghinoni M, Sesenna E: A comparison of open and closed treatment of condylar fractures: a change in philosophy. Int J Oral Maxillofac Surg 30(5):384-9, 2001

47. Cohen J: Statistical power analysis for the behavioral sciences (2nd ed.). 1988

48. Cooper H HL, eds.: The handbook of research synthesis.399-410, 1994

Vita

Marcy L. Nussbaum was born on April 22, 1982 in Roanoke, Virginia. She went to Cave Spring High School in Roanoke, Virginia. She obtained a B.A. in cognitive science with a minor in psychology from the University of Virginia in Charlottesville, Virginia in May 2004. Marcy will graduate from Virginia Commonwealth University with a M.S. in May 2006. She has accepted a position as a research biostatistician for the R. Stuart Dickson Institute for Health Studies in Charlotte, North Carolina.