UNIVERSITY OF STRATHCLYDE

DEPARTMENT OF
MATHEMATICS AND STATISTICS

Statistical Methods for Meta-Analysis with application to Hyperbaric Oxygen Therapy for Post-concussion Symptoms

by

Sarah Reed

202078239

MSc Applied Statistics in Health Sciences

2022/23

Statement of work in project

The work contained in this project is that of the author and where material from other sources has been incorporated full acknowledgement is made.

Signed	*SReed*
Print Name	SARAH REED
Date	14th May, 2023

Supervised by Professor Chris Robertson

Contents

1 Systematic Review and Meta-analysis

This project outlines the processes followed when doing a systematic review, and the statistical methods and models used in meta-analysis. A short research project using meta-analysis is also completed considering the effects of hyperbaric oxygen therapy for persistent post-concussion symptoms.

1.1 Background

Since the 1940's, scientific research has increased almost exponentially[1]. Evidence from numerous researchers with varying backgrounds and biases accumulates to the point that it becomes difficult to establish which way the strongest evidence points, especially when different results exist showing evidence both for and against an intervention. Until the 1980's, researchers would consider results from multiple studies that addressed the same question and would summarise their findings and come to their own conclusions in the form of a "narrative review". There is no formula for doing a narrative review and different researchers could choose different criteria for the inclusion of studies in their review. The importance of each study was implicitly chosen by the researchers. Therefore, these reviews could be subjective and biased. Indeed, there were cases where researchers studied the same research question and came to different conclusions[2].

In the 1970's researchers started taking a more organised approach to literature research by doing "systematic reviews". For a systematic review, researchers must follow a well-defined and transparent process which is reproducible. Research questions are pre-determined along with inclusion and exclusion criteria for the study and information such as databases searched and keywords used must be disclosed.

Attempts to quantitatively summarise effects from multiple studies began in the

early 1900's by Karl Pearson and Ronald A. Fisher, with limited success[1]. It wasn't until 1976, that "meta-analysis" was truly born. It was Gene Glass who managed to develop statistical methods to combine and summarise quantitative results from a systematic review and provide us with a single overall effect size[3]. Meta-analysis can be applied to any studies that are sufficiently similar to one another that answer the same research question and provide quantitative results. At a similar time, Hunter and Schmidt were also developing their own meta-analysis methods. In the mid 1980's Rebecca DerSimonian and Nan Laird developed a way of calculating random-effects meta-analyses, which accounts for variation of effect sizes in different trials. Further advancements have been made in the field since then that have also made meta-analysis more versatile and robust[1].

In the 1970's a tuberculosis specialist called Archie Cochrane, who was very aware of the limited resources in the UK's National Health Service (NHS), pushed for randomised, controlled trials to take place to evaluate whether an intervention was clinically effective before widespread use. A randomised, controlled trial helps reduce bias in an experiment and is done under controlled conditions where patients with similar characteristics are randomly assigned to different groups and these different groups receive either an intervention, no intervention or a placebo (dummy intervention). With the advances in systematic reviews and meta-analyses, Cochrane was keen that relevant randomised, controlled trials within individual medical specialities should be summarised and updated periodically. This led to the development of the Oxford Database of Perinatal Trials. This set the ball rolling for further advancements in systematic review methodology and led to the creation of the now internationally significant Cochrane Collaboration in 1992. The Cochrane Collaboration reviews follow rigorous rules and are considered the gold standard in systematic reviews. It is a network of 37,000 researchers across 130 countries who work together to gather and

assimilate the most up to date information to help healthcare providers and policy makers make well-informed decisions. Whilst the Cochrane Collaboration focuses on healthcare, the formation of further collaborations have followed in other areas, such as the Campbell Collaboration, formed in 2000, which focuses on the effects of social interventions[3].

More recently research data has become more accessible, and this has allowed a new type of meta-analysis to be performed. "Individual Participant Data (IPD) Meta-analysis" uses the original data from each of the studies rather than the summary statistics. This allows the researcher to investigate how different variables may influence the primary and secondary outcomes of the studies. It also allows any missing data to be imputed and for the same statistical methods to be used across all the data[1].

Meta-analysis is also a useful tool to use before starting any new research. It gives context to a proposed research question and can help with the design of a study, including which statistical methods are best to use and which outcomes may be most useful to focus on, in a given research area. Perhaps more importantly, systematic reviews and meta-analyses can determine whether a new study needs to be done at all. It may be that there are sufficient data available from previous studies that can be combined and summarised to answer a research question without potentially putting more patients at risk in a clinical trial[2].

1.2 Methodology

1.2.1 Research Question

The first thing to do in a systematic review study is to define the most appropriate research question. Scoping searches are helpful for this as they give us an idea of

what research is out there and how broad or narrow the research question needs to be. It is also an ideal time to investigate which keywords highlight the most appropriate studies of interest. The scoping search will indicate whether or not the chosen research question is feasible, or if the question has already been answered. If a meta-analysis is planned, enough studies are required which have appropriate data to extract an effect size from.

In 2013, Cummings and colleagues came up with the 'FINER' framework to help choose a research question. It states that the research question should be Feasible, Interesting, Novel, Ethical and Relevant[1].

Once a research question has been chosen, it is time to personalise the question by thinking about what eligibility criteria are required to focus your search. These criteria will help in choosing which studies to include or exclude in the systematic review. A widely used and helpful approach is to break the question down into a few key areas using a PICO(SS) table. PICO stands for Population, Intervention, Comparison and Outcome and the additional (SS) are sometimes added to include Study design and Setting[3].

If the expectation is to do a meta-analysis of biomedical data, then it is common practice to limit the study design choice to include only randomised, controlled trials in the review and analysis. Furthermore, it is highly recommended to follow the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines, which provide best practice recommendations on how the meta-analysis process should be reported. MARS (Meta-Analysis Reporting Standards) provides similar guidelines to follow for doing meta-analyses of psychological and behaviour research[1].

1.2.2 Analysis Plan

After determining the research question and eligibility criteria, it is useful to construct an analysis plan. This should mention which information will be extracted from each study and any calculations that will be required in the analysis. Based on the expected variation between studies, the choice of models to be used should also be stated beforehand. This is because results from a priori analyses are considered to be more credible than post hoc analyses, that can be easily tweaked to suit the researcher's intentions[1].

1.2.3 Search and Selection

When starting the search for studies, it is worth thinking first about whether only published studies should be included in the search or if unpublished material should be included also. Publication bias exists when the decision to publish a paper is biased by its results. Grey material may sometimes be deemed to be less credible, however, including it can reduce publication bias. Unfortunately, a good deal of quality research is not published if it doesn't provide desired results. This is known as the *file drawer effect*, and to make things worse, well designed and conducted research is *less likely* to produce statistically significant results than poorly designed research. Similarly, some researchers only include papers that are in English, causing a language bias to creep into the analysis. If study selection is restricted by language, it must at least be mentioned as a limitation and included in the eligibility criteria. *Funnel plots* can be used to help identify publication bias. These diagrams plot effect sizes against their standard errors, as demonstrated in section 6.5. These graphs appear asymmetrical when studies with non-significant results are not included. Significance tests, such as Egger's test can be conducted to identify funnel

plot asymmetry. However, a lot of studies (more than 10) are required to make these tests worthwhile and meta-analyses often include too few[4].

Key bibliographic databases to search for health-related topics include Pubmed, CENTRAL, EMBASE, ClinicalTrials.gov, MEDLINE, PROSPERO and the Cochrane Library. Other useful multi-disciplinary databases to search include Web of Science, Scopus and Google Scholar[3].

To find the most appropriate studies, it is important to construct a decent search string. A good starting point is to include keywords from the PICO(SS) table and the eligibility criteria. Connecting them with Boolean operators such as AND and OR can help to refine the search. Truncated words or wildcards using asterisks can be used in many databases. These allow for more words with the same meaning to be picked up. For example truncat* would pick up truncate, truncated, truncating etc. Wildcards are useful for including different spellings of words such as reali*e would pick up both realise and realize[1]. Within databases, specific fields can also be searched, finding papers that use different keywords with the same meaning within that field of interest.

Citation chaining can also be used in the search by screening references that have been used in relevant papers and also by searching for articles that have cited a relevant primary study[3].

The search method should always be reported. This should include what, where, when and how you searched and what was found[3]. The Peer Review of Electronic Search Strategies (PRESS) checklist is a recognised set of guidelines for researchers to follow when reporting their search method. It recommends six key search areas to think about including: Translation of the research question; Boolean and proximity operators; subject headings; text word search; spelling, syntax and line numbers; and limits and filters[5]. The exact way these search ideas are combined can greatly

affect how comprehensive the search is. Ideally, there would be at least two people screening the studies to check that they agree with the choice of studies to be included. Transparency of the search method is therefore key to understanding why any discrepancies may have been found.

Choosing which studies to include in the systematic review and meta-analysis is the next step. First, all the duplicate studies need to be removed. Using a reference management system such as EndNote is particularly useful for doing this. The number of references before and after removing duplicates should be noted. Further elimination of studies takes place as the titles and abstracts are scanned through, removing perhaps 90% of the references[1]. The full articles of the remaining references then need to be read and checked against the eligibility criteria to decide if they are suitable studies to be included in the review and meta-analysis. At this stage, reasons for excluding studies from the research must be disclosed. Some articles may not provide enough information to decide if it would be eligible or not, or there may not be enough information to calculate an effect size. In this case, an attempt to contact the authors should be made twice before excluding it from the review. A flow diagram such as the PRISMA flow chart (see Figure 6 in section 4.3) is the best way to show how and why references were removed from the study[1].

To gather all the data required from the studies, it is best to build a data extraction form or table. Details such as the first author and date of publication should be reported. Other data to include are characteristics of the studies such as sample size, population, types of interventions used, outcomes and possible adverse effects. For meta-analyses, any data required to complete the calculations for an effect size and confidence interval must also be included, if this doesn't exist already as a summary statistic. If subgroup analyses and meta-regressions are also planned, the required data should be pulled out for these too.

1.2.4 Meta-analysis

Meta-analysis is a statistical technique to combine all the summary results from similar, compatible studies into a single pooled effect size. The precision of each study effect size is known from calculating its confidence interval, and a weighting is given to each study based on its sampling error, so typically larger studies with smaller sampling errors are given more importance. Studies that are statistically significant can be seen by scanning the confidence intervals. Studies with confidence intervals that cross the null value (1 for ratio effect sizes and 0 for mean differences) are not considered statistically significant. The pooled summary effect from a meta-analysis is simply a weighted mean of the individual effect sizes. Different models can be used to compute the weighting and these are discussed in section 2.2. The outcome of a meta-analysis is often illustrated using a forest plot, as shown in section 2.4.

1.2.5 Risk of Bias

Before reading too much into our results, the reliability of the included primary studies should be assessed. Information from the studies should be extracted to assess the risk of bias. There are different tools available to help, depending on the types of study designs included. These tools help highlight possible systematic errors. For randomised, controlled trials, Cochrane's Risk of Bias tool is helpful[1]. This helps to judge the level of bias in a study due to the randomisation process, deviations from the planned interventions, missing data, outcome measurement, reported results and gives an overall risk of bias[6]. The tool poses signalling questions for each domain, and an algorithm outputs its suggestion of bias level based on these questions. The investigator's judgement can also be inputted. The Cochrane Handbook recommends

two types of figures to illustrate judgements of study validity in these domains: A barplot showing the proportion of studies with a given risk of bias level for each domain (figure 1), and a traffic light plot, which shows the domain level judgement for each study (figure 2). Red signifies a high level of bias in a domain, yellow shows there are some concerns, and green shows a judgement of low level bias.

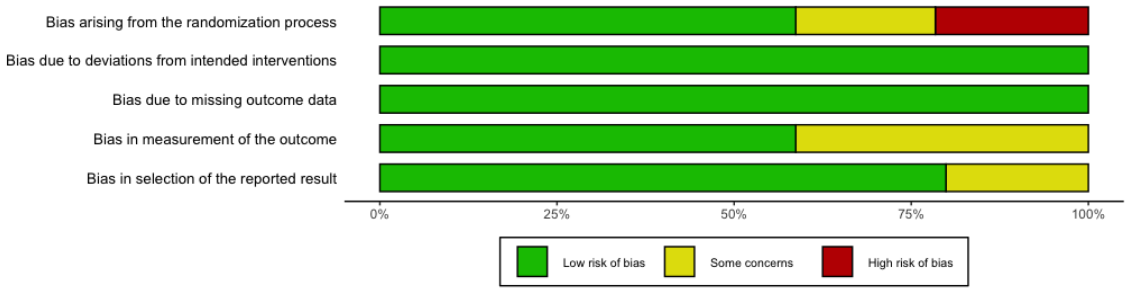

Figure 1: Risk of bias plot showing proportion of studies with a given risk of bias

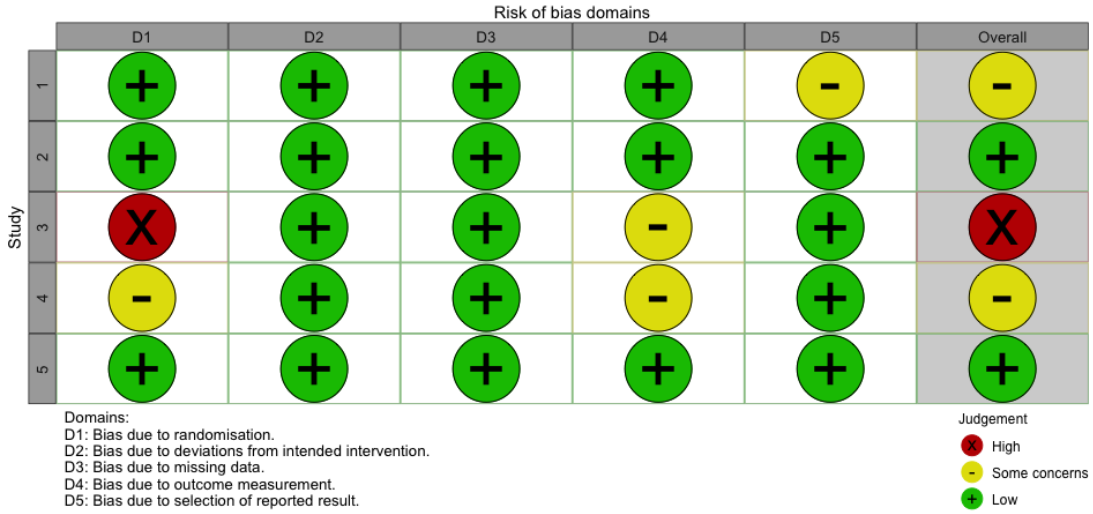

Figure 2: Traffic light plot showing the domain level judgement for each study

Of course there may be other biases that should be considered, such as how a study was funded. Even when a study follows the best practice methods, biases can still exist[3]. If there are highly biased studies, they can be excluded from the meta-analysis. Equally, it may be interesting to complete sensitivity studies to see how much the results differ, with and without the study in question.

In this chapter, systematic reviews and their formal structure have been outlined. Meta-analysis has also been introduced, along with the risk of bias that should be considered. In the next chapter, the mathematical formulae and statistical methods used when undertaking a meta-analysis are described.

2 Meta-analysis

Meta-analysis uses statistical methods to combine effect size summary statistics from similar intervention studies into an overall, pooled effect size. Information such as the sample sizes, observed effect size and its standard deviation are extracted from each study. This information is not always available in an ideal format for meta-analysis. Sometimes, the required results must be calculated from the given study data before implementing a meta-analysis.

This chapter describes the types of effect sizes that are often used in meta-analysis. It shows how to calculate them, along with their estimated standard errors that are required to compute confidence intervals, and how to convert between different types of effect size statistics in order to pool together results in the meta-analysis using a single standardised effect size.

The methods in this chapter have been collated from "Doing Meta-Analysis with R"[1], "Introduction to Meta-Analysis"[2] and the "Cochrane Handbook for Systematic Reviews of Interventions"[7,8].

2.1 Effect sizes for Meta-analysis

2.1.1 Continuous data

For continuous data that are normally distributed the effect size extracted will be either the raw mean difference, D, or the standardised mean difference, d. The raw mean difference (equation 1), can be used when data have a common metric, such as when all weights are recorded in kilograms.

$$D = \bar{x}_A - \bar{x}_B \tag{1}$$

where $\bar{x}_A$ and $\bar{x}_B$ are the means of groups A and B, which could be the treatment

and control groups, for example. The standard deviation, s, of the observed effect size must also be extracted from each group to pool the results. s_{pooled} is the pooled standard deviation of groups A and B and is calculated as follows:

$$s_{pooled} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{(n_A - 1) + (n_B - 1)}} \tag{2}$$

where n_A and n_B are the associated sample sizes of each group within the study and s_A and s_B are the standard deviations of each group. These statistics are also required to calculate the standard error of D:

$$SE_D = s_{pooled}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}} \tag{3}$$

When data are recorded in different metrics, the standardised mean difference, also known as Cohen's d (equation 4), is used instead and provides us with the difference between groups in units of standard deviations, so the results can still be pooled together.

$$d = \frac{\bar{x}_A - \bar{x}_B}{s_{pooled}} \tag{4}$$

$$SE_d = \sqrt{\frac{n_A + n_B}{n_A n_B} + \frac{d^2}{2(n_A + n_B)}} \tag{5}$$

Standardised mean differences can be loosely interpreted using Cohen's convention, however, it is important to consider what a clinically significant result for the study in question would be. As a guide though, Cohen suggests:

$d \approx 0.20$: small effect

$d \approx 0.50$: moderate effect

$d \approx 0.80$: large effect

Standardising, however, has its issues. The standardised results are greatly affected by the amount of variability in the study sample. This means that the standardised mean difference between groups could differ despite the raw mean difference being the same. For small studies (study size, $N \leq 20$), which are common, d is often over-estimated, so a bias correction called J is often used:

$$J = 1 - \frac{3}{4N - 9} \tag{6}$$

This converts Cohen's d to the Hedges' g effect size (equation 7). The standard error of g is expressed in equation 8.

$$g = J \times d \tag{7}$$

$$SE_g = J \times SE_d \tag{8}$$

It should be noted that Hedges' g and Cohen's d are sometimes used interchangeably in research papers so it is wise to confirm with the authors which measure has been used before including in the meta-analysis.

2.1.2 Binary data

Effect sizes commonly used for binary data include the risk ratio, odds ratio and risk difference (equations 9, 10 and 11 respectively). A contingency table (table 1), is useful for calculating these effect sizes for binary data and represents the number of patients in the intervention group who had an event (a), those who didn't have an event (b), and the number of patients in the comparison group who had an event (c) and those who didn't have an event (d). n_A is the total sample size in the intervention group and n_B is the total number in the comparison group.

Table 1: 2 x 2 Contingency table

	Event	No event	Total
Intervention Group (A)	a	b	$n_A = a + b$
Comparison/Control Group (B)	c	d	$n_B = c + d$

$$\text{Risk Ratio (RR)} = \frac{\text{risk of event on treatment}}{\text{risk of event on control}} = \frac{(a/n_A)}{(c/n_B)} \tag{9}$$

$$\text{Odds Ratio (OR)} = \frac{\text{odds of event on treatment}}{\text{odds of event on control}} = \frac{(a/b)}{(c/d)} = \frac{ad}{bc} \tag{10}$$

$$\text{Risk Difference (RD)} = \text{risk on treatment} - \text{risk on control} = \frac{a}{n_A} - \frac{c}{n_B} \tag{11}$$

Risk Ratio

When both groups have the same treatment effect, RR=1. If RR>1, this increases the risk of an event. If RR<1, this decreases the risk of an event. Risk ratios do not follow a normal distribution: If the risks are halved, RR=0.5 but if they are doubled, RR=2. These two values are not equidistant from 1, which causes problems for meta-analysis. The logarithmic transformation of the risk ratio should therefore be taken before pooling the effects (equation 12).

$$logRR = ln(RR) \tag{12}$$

$$SE_{logRR} = \sqrt{\frac{1}{a} - \frac{1}{n_A} + \frac{1}{c} - \frac{1}{n_B}} \tag{13}$$

$$RR^* = e^{logRR} \tag{14}$$

All computations in the meta-analysis are carried out on a logarithmic scale, including calculation of the standard error, (equation 13), and confidence interval. At the end, the pooled logRR is returned to its original metric by taking the exponential, to give the overall risk ratio, RR*, (equation 14).

Calculations of risk ratios can be problematic when there are cells with a value of zero. These are often overcome by adding or subtracting 0.5 to each cell with a value of zero. Similarly, when sample sizes in the two groups are very uneven, a "treatment arm continuity correction" can be applied. Unsurprisingly, using these corrections can lead to biased results too.

Odds Ratio

Computations of the odds ratio in meta-analysis and associated confidence intervals are also carried out on the logarithmic scale using the logarithmic odds ratio and its standard error, as seen in equations 15 and 16. Odds ratios also have problems if there are zero value cells, and continuity corrections can again be used.

$$logOR = ln(OR) \tag{15}$$

$$SE_{logOR} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \tag{16}$$

$$OR^* = e^{logOR} \tag{17}$$

Once again, these logarithmic values are returned to their original metric at the end by taking the exponential to give the pooled odds ratio, OR*, (equation 17).

Risk Difference

The risk difference measures the *difference* between two risks and therefore calculations do not need to be done on the logarithmic scale. Expressions for risk difference and its standard error are shown in equations 18 and 19.

$$RD = \frac{a}{n_A} - \frac{c}{n_B} \tag{18}$$

$$SE_{RD} = \sqrt{\frac{ab}{n_A^3} + \frac{cd}{n_B^3}} \tag{19}$$

2.1.3 Correlations

The correlation coefficient between two continuous variables can also be used as an effect size in meta-analysis studies. This effect is more typically used within psychology studies. The population correlation, ρ, is estimated by the sample correlation, r. The standard error of r is calculated as follows:

$$SE_r = \sqrt{\frac{(1-r^2)^2}{n-1}} \tag{20}$$

Since the variance depends heavily on the correlation, the correlation is usually converted to the Fisher's z scale using equation 21, with all computations done using the z values. The standard error of z is expressed in equation 22. The resulting summary effect and its associated confidence interval would then be converted back to an overall correlation, r* using equation 23.

$$z = 0.5 \times ln\left(\frac{1+r}{1-r}\right) \tag{21}$$

$$SE_z = \sqrt{\frac{1}{n-3}} \tag{22}$$

$$r^* = \frac{e^{2z} - 1}{e^{2z} + 1} \tag{23}$$

2.1.4 Confidence Intervals

For every effect size, a confidence interval (CI) should be calculated as a measure of precision. It is typical to quote a 95% confidence interval (equation 24). This means that on repeated population sampling, 95% of the time, the interval will contain the true, unknown parameter.

$$95\%\ CI = \text{effect size} \pm (1.96 \times SE_{\text{effect size}}) \tag{24}$$

2.1.5 Statistical Conversions

Sometimes, there are studies that all answer the same broad question, but their summary statistics have been measured in different ways. However, provided it still makes sense to combine the results, it is possible to do so by converting these effect sizes to a common metric using some helpful formulae.

Standard errors (SE) from confidence intervals and p-values

The SE of a group's mean difference can be obtained from the upper and lower limits of its 95% CI, CI_{upper} and CI_{lower} as follows:

$$SE = \frac{CI_{upper} - CI_{lower}}{3.92} \tag{25}$$

If the group k sample size is small, < 20, it is better to replace 3.92 with (2×t-statistic). If the standard deviation, s_k, of the group's mean difference is required, it can be calculated from its SE as shown in equation 26 where n is the group size:

$$s_k = \sqrt{n} \times SE \tag{26}$$

When p-values have been given using a Wald test, the z-statistic corresponding to the p-value can be obtained from a table of the standard normal distribution. Then the SE can be found as follows:

$$SE = \left| \frac{\text{observed intervention effect}}{z} \right| \tag{27}$$

When p-values have been given for a group's paired t-test for the mean difference between the before and after treatment values, a table of t-distributions or an inverse t-distribution calculator can be used to find the t-statistic that corresponds to the given p-value, (or half its p-value if a 2-tailed test was used), and the appropriate number of degrees of freedom, $n - 1$. If a p-value is given for the difference of means between two treatment groups, then the degrees of freedom is given by $n_A + n_B - 2$ where n_A and n_B are the sample sizes of the intervention (A) and comparison/control (B) groups. This t-statistic and the difference of means, D, can then be used to obtain the SE as follows:

$$SE = \left| \frac{D}{t} \right| \tag{28}$$

The standard errors can be found in the same way for ratios, except that all calculations are done on the natural logarithmic scale.

Converting from logOR to Cohen's d

It is possible that some of the effect sizes are given as a mean difference, and some are reported as an odds ratio. These effect sizes can still be pooled together: A log odds ratio effect size can be converted to d as shown in equation 29 with its associated standard error in equation 30.

$$d = logOR \times \frac{\sqrt{3}}{\pi} \tag{29}$$

$$SE_d = SE_{logOR} \times \frac{\sqrt{3}}{\pi} \tag{30}$$

In applying this conversion, the underlying data are assumed to be continuous and have a logistic distribution.

Converting from d to logOR

To convert from d to a log odds ratio, and to calculate its standard error, the following formulae can be used:

$$logOR = d\frac{\pi}{\sqrt{3}} \tag{31}$$

$$SE_{logOR} = SE_d\frac{\pi}{\sqrt{3}} \tag{32}$$

Again, in order to use this conversion, the underlying data are assumed to be continuous and have a logistic distribution.

Converting from r to d

Similarly, some studies may report an effect size of the strength of relationship between the control and treatment groups rather than a mean difference. To convert from a correlation, r to a standardised mean difference, d, and its associated standard error, the following formulae can be used:

$$d = \frac{2r}{\sqrt{1-r^2}} \tag{33}$$

$$SE_d = \frac{2SE_r}{\sqrt{(1-r^2)^3}} \tag{34}$$

In applying this conversion the continuous data used to calculate r are assumed to have a bivariate normal distribution.

Converting from d to r

To convert from d to r, the following formulae can be used:

$$r = \frac{d}{\sqrt{d^2 + \alpha}} \tag{35}$$

where α is a correction factor for when $n_A \neq n_B$,

$$\alpha = \frac{(n_A + n_B)^2}{n_A n_B} \tag{36}$$

If n_A and n_B are not known exactly, use $n_A = n_B$, giving $\alpha = 4$.

The standard error of r is then:

$$SE_r = \frac{\alpha SE_d}{\sqrt{(d^2 + \alpha)^3}} \tag{37}$$

To use this conversion, an assumption is made that a continuous variable was dichotomised to form the treatment and comparison groups.

When converting between Fisher's z and d, assumptions are made about the *independent* variable only. Whilst converting between the log odds ratio and d, assumptions are made about the *dependent* variable only. Since these assumptions are completely independent of one another, both assumptions can be safely applied in order to convert a Fisher's z effect through to d, then transformed again to the logOR. This applies in the reverse direction also.

2.2 Basic Meta-analysis models

There are two main statistical models used in most meta-analyses. The Fixed-effect model (also known as the common effect or inverse-variance model) and the Random-effects model.

2.2.1 Fixed-effect Model

The fixed-effect model is the simplest model and assumes that all studies included share the same true population effect size, θ, and the observed effect size of study k, $\hat{\theta}_k$, only differs from θ due to a sampling error ϵ_k as follows:

$$\hat{\theta}_k = \theta + \epsilon_k \tag{38}$$

If each study had an infinite sample size, the sampling error would be zero. As the sample size increases, the sampling error, which is often represented by the standard deviation, s_k, decreases. Since larger studies with a smaller standard deviation (more precise) should be better estimators of the true overall effect than smaller studies, a greater weight can be assigned to them when pooling together the results, thereby giving a weighted average of the studies. The weight of study k, w_k is calculated by taking the inverse of its within-study variance, s_k^2:

$$w_k = \frac{1}{s_k^2} \tag{39}$$

The summary effect, $\hat{\theta}$, which is a pooled estimate of the true effect, is the weighted average and calculated as follows:

$$\hat{\theta} = \frac{\sum_{k=1}^{K} \hat{\theta}_k w_k}{\sum_{k=1}^{K} w_k} \tag{40}$$

The standard error of this pooled effect is estimated to be:

$$SE_{\hat{\theta}} = \sqrt{\frac{1}{\sum_{k=1}^{K} w_k}} \tag{41}$$

The 95% confidence interval of the summary effect can be calculated as follows:

$$95\%\ CI = \hat{\theta} \pm (1.96 \times SE_{\hat{\theta}}) \tag{42}$$

Next a z-value should be calculated to test the null hypothesis, which is that the true effect θ is zero for a difference or one for a ratio:

$$z = \frac{\hat{\theta}}{SE_{\hat{\theta}}} \tag{43}$$

The p-value for a two-tailed test is then given by

$$p = 2[1 - (\Phi(\mid z \mid))] \tag{44}$$

where $\Phi(z)$ is the standard normal cumulative distribution. If the p-value ≤ 0.05, the effect is statistically significant.

For binary data, alternative methods exist such as the Mantel-Haenszel and Peto methods, which both attempt to improve the model.

2.2.2 Random-effects Model

The issue with the fixed-effect model is that often the studies being pooled together are not identical in their design and samples are from slightly different population groups, so there are bound to be some differences in the results. The random-effects model accounts for this additional source of variance, ζ, due to between-study heterogeneity (differences found between the studies). Instead of assuming a fixed value for the true effect, the random-effects model assumes there is a normal distribution of true effect sizes and it is the mean of this distribution that is required for our pooled effect size. The observed effect, $\hat{\theta}_k$, is considered an estimate of the *study's true effect*, θ_k, plus a sampling error, ϵ_k:

$$\hat{\theta}_k = \theta_k + \epsilon_k \tag{45}$$

However, the true effect size of study k, θ_k, is part of a distribution of effect sizes with mean, μ, plus the between-study error, ζ_k:

$$\theta_k = \mu + \zeta_k \tag{46}$$

Substituting equation 46 into equation 45, the observed study effect size, $\hat{\theta}_k$, is found to be:

$$\hat{\theta}_k = \mu + \zeta_k + \epsilon_k \tag{47}$$

In other words, the observed effect size, $\hat{\theta}_k$, deviates from the pooled mean effect, μ, by two error terms, the between-study error, ζ_k, and the sampling error, ϵ_k. Figure 3 illustrates the parameters in a random-effects model.

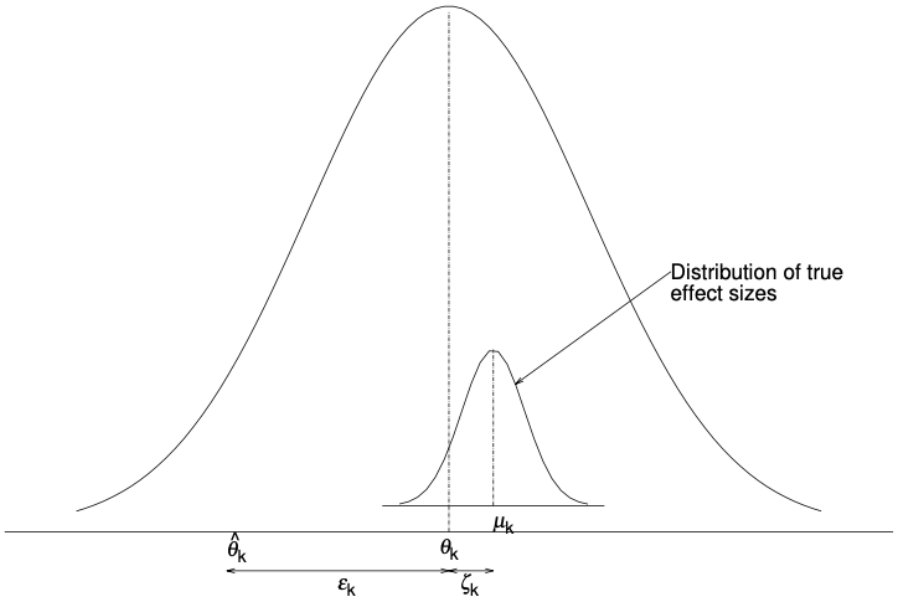

Figure 3: Illustration of the parameters used in a random-effects model[1]

To account for ζ_k, the variance of the normal distribution of true effect sizes, τ^2, must be estimated. There are a few methods to estimate τ^2 but the most common one used is DerSimonian-Laird's method of moments (see equation 55).

This estimator can be biased when heterogeneity is high and the number of studies is small, but for easy replication of results this is the method of choice. The weight calculation can then be adjusted to account for the between-study heterogeneity,

giving an adjusted random-effects weight, w_k^*, as follows:

$$w_k^* = \frac{1}{s_k^2 + \tau^2} \tag{48}$$

The pooled effect estimate, $\hat{\theta}$, can then be calculated as follows:

$$\hat{\theta} = \frac{\sum_{k=1}^{K} \hat{\theta}_k w_k^*}{\sum_{k=1}^{K} w_k^*} \tag{49}$$

The standard error of this summary effect is estimated to be:

$$SE_{\hat{\theta}} = \sqrt{\frac{1}{\sum_{k=1}^{K} w_k^*}} \tag{50}$$

The confidence interval and z-value are calculated in the same way as for the fixed-effect model and the null hypothesis is that the *mean* of the distribution of true effects is zero for a difference, or one for a ratio.

A Knapp-Hartung adjustment can also be applied, which tries to control for the uncertainty in the estimate of τ^2 by changing the calculation of $SE_{\hat{\theta}}$ and basing the significance test on a t-distribution rather than a normal distribution. It is considered sensible to apply this adjustment and usually makes the confidence intervals for the pooled effect slightly larger. However, it can produce anti-conservative results when effect sizes are similar in all studies.

An important assumption in order to use the random-effects model is that ζ_k is independent of study k. This is known as the exchangeability assumption, as all true effect sizes are assumed to be exchangeable. Since some degree of study heterogeneity is expected to exist, it is common practice to use a random-effects model. However, some argue that a fixed-effect model can be preferable as the random-effects model pays comparatively more attention to small studies when calculating the pooled effect size. Since small studies can have more biases attached to them, pooled results from the random-effects model could be more biased.

2.2.3 Mantel-Haenszel method for binary data

The fixed-effect and random-effects models work fine for binary data, however when the number of events or the total sample size of a study is small, the Mantel-Haenszel method is often preferred, and uses different formulae for the weights of studies.
Risk Ratio:

$$w_k = \frac{(a_k + b_k)c_k}{N_k} \tag{51}$$

Odds Ratio:

$$w_k = \frac{b_k c_k}{N_k} \tag{52}$$

where, as in section 2.1.2, a_k is the number of events in the treatment group, b_k is the number of non-events in the treatment group, c_k is the number of events in the control group and N_k is the total sample size.

2.3 Heterogeneity

Often, it is useful to know how much of the variation in the pooled effect size is due to sampling error and how much is due to between-study heterogeneity. There are a few measures used to quantify this. One measure is Cochran's Q, which is a weighted sum of squares on a standardised scale, computed by:

$$Q = \sum_{k=1}^{K} \frac{(\hat{\theta}_k - \hat{\theta})^2}{s_k^2} \tag{53}$$

where s_k^2 is the study's variance and $\hat{\theta}_k$ and $\hat{\theta}$ are the observed effect size and the estimated pooled effect. Q can be used to determine if there is more variation than expected from sampling error alone. If there is excess variation then it can be assumed to be from between-study heterogeneity. Q is assumed to roughly follow a χ^2 distribution with K-1 degrees of freedom when only sampling errors exist. So,

Q can be used to test if the variation in a meta-analysis exceeds the amount that we expect under the null hypothesis of no heterogeneity. However, whether Q is significant depends heavily on the size of the meta-analysis and many argue that Q doesn't really follow a χ^2 distribution and therefore Q and related statistics based on Q may be biased.

Higgins and Thompson's I^2 statistic is based on Q and is defined as the percentage of variability in the effect sizes that is not due to sampling error, and is calculated as follows:

$$I^2 = \left(\frac{Q - (K-1)}{Q}\right) \times 100\% \tag{54}$$

where K is the number of studies. If $I^2 = 0$, there is no between-study heterogeneity. If I^2 is negative, the value is rounded up to zero. As a rough guide, I^2 values of 25%, 50% and 75% are associated with a low, medium or high level of heterogeneity.

The variance of the true effect, τ^2, as mentioned previously is useful as it is expressed on the same scale (squared) as the effects so it is easier to interpret statistically, however, it is harder to determine how meaningful τ^2 is in practical terms. It can be estimated as follows:

$$\hat{\tau}^2 = \frac{Q - (K-1)}{C} \tag{55}$$

where

$$C = \sum w_k - \frac{\sum w_k^2}{\sum w_k} \tag{56}$$

A useful and more meaningful way of quantifying heterogeneity is to show prediction intervals. It is common to give a confidence interval for the pooled effect size but it would be nice to know how the *true effects* are distributed around this summary effect. Using a t-distribution with K-1 degrees of freedom, a 95% prediction interval (PI), can be calculated as follows:

$$PI = \hat{\mu} \pm t_{K-1,0.975}\sqrt{SE_{\hat{\mu}}^2 + \hat{\tau}^2} \tag{57}$$

where $\hat{\mu}$ is the estimated pooled effect, $SE_{\hat{\mu}}$ is the standard error of the pooled effect and $\hat{\tau}^2$ is the estimated between-study heterogeneity variance. Thus while the 95% CI indicates where the mean effect size will lie within, 95% of the time, the 95% PI indicates where the *true effect* in a new study will lie 95% of the time. In other words, the precision of the pooled summary effect is given by the confidence interval and the distribution of *true effect* sizes is given by the prediction interval.

In meta-analysis, issues can sometimes arise due to outlying or influential study effects. Whilst outliers are defined by the size of their effect, they don't always impact the results much. Influential results from a particularly large study, however, by definition have a significant impact on the meta-analysis results, regardless of the effect size. Influential studies can be detected using a "leave-one-out" method, where the meta-analysis results are re-calculated K times, leaving a different study out each time. Influence diagnostics can then use these results to provide information about which studies influence the pooled summary effect estimate the most.

2.4 Forest Plots

The observed effect sizes of each study and the overall, weighted effect size can be displayed in a forest plot (figure 4). The centre of the diamond represents the pooled summary effect size, with the width of it representing its confidence interval. If prediction intervals are given, these are shown as horizontal lines extending either side of the diamond. The centres of the boxes give the observed effect sizes of the individual studies with the size of the box representing the size of the study and its associated weighting. For example, in figure 4 we can see that study 7 was a small study, whilst study 1 was large. The horizontal lines either side of the boxes give the confidence intervals for each study and indicate how dispersed the results

are. Smaller confidence intervals represent more precise results. Typically, forest plots also include a vertical reference line showing the x-axis value of no effect (null hypothesis). In this example, the confidence intervals for studies 1-3 all cross the null value and therefore their results are not statistically significant. The pooled effect size for this meta-analysis is 0.42 with a confidence interval of (0.28, 0.56).

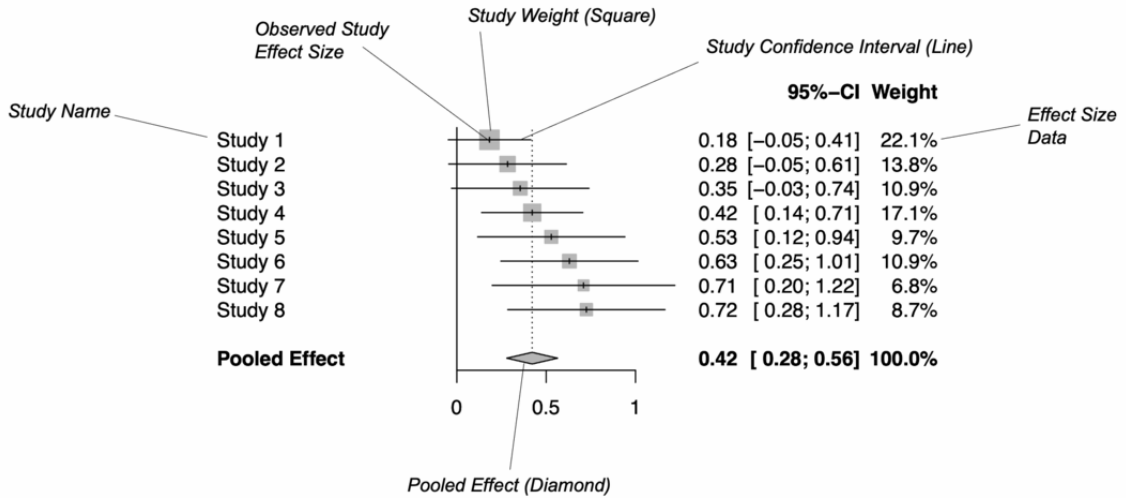

Figure 4: Example of a forest plot of effect sizes[1]

2.5 Power Analysis

As mentioned previously, a major advantage of doing a meta-analysis is the increase in power from including a larger total sample size. This is particularly helpful when the true effect size is small. Power is the probability of rejecting the null hypothesis of *no effect* when there *is* an effect. A study power of at least 80% is desirable. A study is considered to be under-powered if its power is less than 80% and indicates that a larger sample is required. A power calculation should be computed for the minimum size of effect that we are interested in finding, and should ideally be done *a priori*

to ensure that enough studies are included in order to verify that a small effect does exist. There are various functions that have been developed to compute the power of a study. See chapter 14 of "Doing Meta-Analysis in R"[1] for more information.

2.6 Advanced Meta-analysis models

2.6.1 Subgroup analyses

In a meta-analysis a weighted mean effect size is computed for an outcome of interest. However, data from the studies may fall into two or more natural groupings. For example, the effectiveness of a new drug may be of interest, but the studies may use different doses of the drug. Subgroup analyses allow the effect size for each dose size to be computed, to help determine which dose is the most effective. It is also a useful tool to investigate variation shown in the forest plot. Ideally, subgroup tests should be defined a priori. There are a few different statistical models that can be used for subgroup analyses but only a fixed-effect**s** model is outlined here.

Firstly, a pooled effect size, $\hat{\mu}_g$, is calculated for each subgroup, g, typically using a random-effects model. There will now be different estimates of heterogeneity for each subgroup, $\hat{\tau}_g^2$, but it is common to instead use a value of τ^2 that has been pooled across the subgroups, as small subgroups may give imprecise estimates of heterogeneity.

If each subgroup is now treated as though it is a single study in a meta-analysis, heterogeneity between the groups can be assessed in a similar way to a normal meta-analysis. Cochrane's Q statistic can be used to determine if the differences between the groups are larger than what they would be if due to sampling error alone. Q is then compared to a χ^2 distribution with G-1 degrees of freedom. It tests the null hypothesis that all subgroups have the same underlying effect.

There are a couple of caveats to be aware of with subgroup analyses. Firstly, the reason for doing a meta-analysis is to combine studies to *increase* statistical power for a given question. If the data is then divided up again into subgroups, statistical power is lost, so it usually makes no sense to conduct subgroup analyses on a small number of studies. Also, if no statistically significant effect difference is found, it does *not* follow that the group effects are equivalent. Subgroup analyses are also observational so the possibility of confounding variables must never be discounted. Finally, care should be taken to avoid ecological bias caused by grouping by mean values, such as using mean ages to separate groups, as this allows for groups with overlapping ages. To avoid this, all individuals in a study would need to fall into a single group, such as 0-18 years and over 18 years.

Since the fixed-effects model includes both fixed effects and random effects, it is often referred to as a mixed-effects model. Sometimes, it may be that the subgroup levels are not really fixed, but are randomly selected. In this case it is better to let the model estimate variability as a random effect, as in a multi-level model.

2.6.2 Meta-regression

Another use of mixed-effects models is in meta-regression. In primary studies, multiple regression is used to assess relationships between covariates and a dependent variable. In meta-regression, a similar process is followed, only the covariates are considered at the study level rather than at subject level and the dependent variable is the effect size. The usual techniques in standard multiple regression can be performed to assess the impact of covariates to predict the effect size in studies with particular characteristics. Typically it is recommended to have at least ten studies for each covariate for the analysis to be meaningful.

2.6.3 Network meta-analysis

Most meta-analyses compare an intervention group with one other group, such as a control group. However, often there are multiple different interventions that can be used for a single condition. There may be studies comparing each intervention with a control, but no studies comparing the interventions directly. Which intervention is most effective though? Network meta-analysis can be used to compare interventions *indirectly* through the control group that the studies all share, and form a hierarchy of intervention effectiveness. Empirical studies have shown network meta-analysis to provide more precise estimates of effects when compared with a single direct or indirect estimate from an intervention-control study[7]. Figure 5 illustrates how a network of direct and indirect connections may be built up in a network meta-analysis:

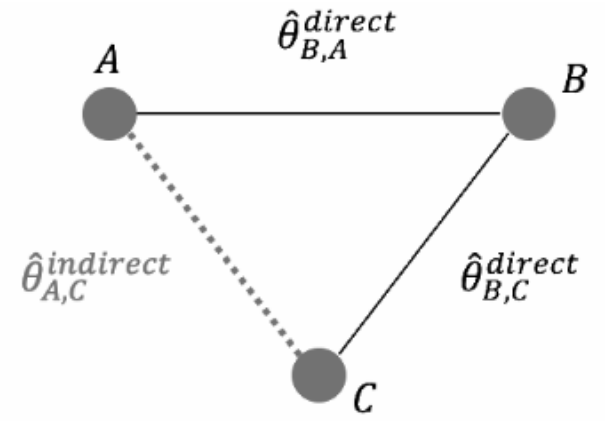

Figure 5: Structure of a network with two direct and one indirect connection. The circles are *nodes* and represent the treatment type. The solid lines, called *edges* represent how the linking treatments relate to each other. The dotted line represents the indirect link between A and C.[1]

In study i, A represents treatment A whilst B represents the control treatment, B, for example, a placebo which all studies use. The line between A and B represents the relationship between their effect sizes, $\hat{\theta}_{i,A,B}$. If there are data from another study, j, which compared treatment C to the placebo treatment B, this can be added to form a network as shown in figure 5, with $\hat{\theta}_{j,C,B}$ being the relative effect sizes between B and C. Since both these effect sizes were *observed* in primary studies, they provide *direct* evidence, so these effect sizes are defined as $\hat{\theta}_{B,A}^{direct}$ and $\hat{\theta}_{B,C}^{direct}$, where B comes first in the notation as it is the reference treatment. However, there is also *indirect* evidence between A and C since B serves as a bridge between the two treatments. This indirect effect estimate, $\hat{\theta}_{A,C}^{indirect}$ is derived as:

$$\hat{\theta}_{A,C}^{indirect} = \hat{\theta}_{B,A}^{direct} - \hat{\theta}_{B,C}^{direct} \tag{58}$$

Network meta-analysis combines these direct and indirect relative effects into one model. To compute the variance of an indirect comparison, the variances of the direct comparisons that form the bridge should be added up, so effect sizes estimated from indirect connections have greater variation and therefore lower precision than those from direct comparisons.

For a network meta-analysis to be valid, however, *transitivity* should be assumed. This means that the studies being compared must be similar in all important factors (such as potential modifiers) other than the intervention used. Transitivity requires all the interventions to be jointly randomisable, such that in theory, all interventions could be compared at the same time in a single multi-arm randomised trial. The statistical manifestation of transitivity is called *coherence* or *consistency*, and implies that all the direct and indirect evidence agrees with each other.

Figure 5 shows how three treatments may be compared. However, often comparison of many interventions is required, requiring a much larger, more complicated network.

If there are K studies, the number of pairwise comparisons is defined as M. The observed effect sizes, $\hat{\theta}_m$, are collected for each comparison m into a vector $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_M)$. The network meta-analysis model is expressed by the formula given in equation 59, where $\mathbf{X}$ is an $m \times n$ matrix with m rows of treatment comparisons and n columns of different treatments.

$$\hat{\theta} = \mathbf{X}\theta_{treat} + \epsilon \tag{59}$$

Vector θ_{treat} contains the *true* effects of the n treatments in the network. Finally, ϵ is a vector containing the sampling errors, ϵ_m of all the comparisons, and are assumed to be a random draw from a normal distribution with a mean of zero and variance of σ_m^2. However, there is a problem of the model being over-parameterised due to the matrix not having full rank, as its columns are not all independent. This means the matrix $\mathbf{X}$ is not invertible, so θ_{treat} cannot be estimated directly using a weighted least squares approach. Instead, it is better to use the *graph theoretical* approach which uses a Moore-Penrose pseudo-inverse matrix, which does allow for the calculation of fitted values using a weighted least squares method.
Cochrane's Q can be calculated to measure the total heterogeneity in the network, and from this an I^2 can be computed to measure the inconsistency in the network.

2.7 Summary

In this chapter, some common techniques used in meta-analysis have been described. Calculations of effect sizes and conversions to different effect sizes for pooling in a meta-analysis have been reported. The fixed-effect and random-effects models have been described and forest plots for presenting results have been shown. Also, power analysis and the meanings and calculations of some measures of heterogeneity

have been discussed. Finally, the more advanced meta-analysis models for subgroup analysis, meta-regression and network meta-analysis have been outlined. In the next chapter, research on post-concussion symptoms (PCS) is introduced.

3 Concussion and Post-concussion Symptoms

There are groups of people in society who suffer quietly every day with symptoms that are very *real* but that doctors fail to treat, as they can't find anything physically wrong with them. People with prolonged symptoms after having suffered a concussion fall into this category, but hyperbaric oxygen therapy is a treatment that may help reduce their symptoms.

3.1 What is a concussion?

A concussion is a form of mild traumatic brain injury (mTBI) caused by a blow to the head or whiplash effect. The required force is surprisingly small and only around 10% of reported concussions involve a loss of consciousness[9]. They often result from car accidents, falls or sports impacts, such as in rugby. After a concussion, patients can suffer from post-concussion symptoms (PCS). These can include headaches, dizziness, memory loss, increased light and sound sensitivity and fatigue. Usually, symptoms only last a few weeks but in about 30% of individuals they can continue for longer and cause prolonged changes in cognitive function. Those whose symptoms have not subsided after three months are sometimes referred to as having persistent post-concussion symptoms (PPCS). This subset of patients continue to report somatic, cognitive, emotional and behavioural difficulties which can significantly interfere with their working and family life[10]. Repeated concussions can cause progressive cognitive decline and a second concussion before the brain has recovered can cause life threatening brain swelling.

Concussion is difficult to manage as the symptoms are often invisible to others and typically nothing untoward shows up on standard computed tomography (CT) or magnetic resonance imaging (MRI) brain scans. Instead, diagnosis normally still re-

lies on self-reporting of symptoms, which may not be immediately recognised. There are imaging techniques, such as single-photon emission computerised tomography (SPECT) that do reveal differences in brain activity in PCS patients by detecting altered blood flow, but these scans are not typically performed.

3.2 Changes in the brain

Within the skull, the brain is surrounded by cerebrospinal fluid and protective membranes called meninges, which are designed to cushion the brain. When a concussion occurs, the brain can be bruised as it knocks against the skull, and different parts of the brain can move at different speeds causing shearing forces that can stretch and tear nerve tissue. PCS are a result of dysfunctional neuro-vascular coupling (NVC), which involves the relationship between neurons and the blood vessels that supply them. Concussion alters the balance of ions and chemicals in the brain, impairing neuron function and their ability to send signals to other parts of the brain. Secondary processes, activated due to the injury include the production of harmful chemicals called free radicals, inflammation, poor movement of molecules within the neurons, and imbalances of important ions required for nerve function. Consequently, cells don't receive the correct amount of oxygen to signal the brain as normal. Instead, other neural pathways that are less efficient are used in an attempt to complete the task. Thus, the hypo-active areas of the brain don't do their normal share of the work, whilst other areas of the brain take on the extra work to complete the task, which is tiring for the brain and leads to PCS. Repairing any injury to the brain is energy-intensive and unfortunately the damage makes it even harder for the neurons to generate the required energy in order to heal[11,12].

Research has shown that even one "silent" sub-concussive blow to the head triggers

changes in how neurons function. These changes don't always appear to have an effect in the short term but seem to have consequences years later. A study published in 2014, following 800,000 conscripts over a 30 year period, found a statistically significant increase in the risk for developing young-onset, non-Alzheimer's dementia after suffering one mTBI.

Also, in a survey of more than 2000 retired professional American footballers, players who had had multiple concussions were three times more likely to have been diagnosed with clinical depression. A study of death certificates also found the death rate from neurodegenerative diseases to be three times higher than in the general population. Chronic traumatic encephalopathy (CTE), is a progressive neurodegenerative disease thought to be linked to repeated head injuries. Its symptoms include memory loss, depression, anxiety and confusion, some of which are also PCS. CTE was first discovered in the early 2000's by Bennet Omalu during a post-mortem of a former player, Mike Webster. Under microscope, he discovered aggregates of tau protein, which is one of two proteins that accumulate in Alzheimer's disease. CTE has since been found in 76 of 79 former American football players who donated their brains for research. However, repeated head injuries do not always lead to CTE. It remains unclear, but it would appear that an individual's genetics may also affect the outcome[13]. Are genetics and tau protein aggregates also contributing factors in PPCS patients?

The research completed in this project specifically covers PPCS. However, when even low level knocks to the head are shown to have such a long lasting impact on the brain, it demonstrates just how critical it is to increase research into how the brain responds to any trauma.

3.3 Persistent post-concussion symptoms or something else?

Several studies have been done to investigate the prevalence of PPCS in the general population who have and haven't had head injuries. In these studies they found no significant difference between these two groups, suggesting that what we consider to be PPCS may not always originate from having a head injury[14,15]. However, van der Vlegel et al[10] found that whilst PCS are evident in patients with and without head injuries, there are differences in their symptom profiles: Emotional complaints existed in both groups, however, cognitive and most somatic complaints were significantly more prevalent in head injury patients.

Studies have also found similar symptoms in people with chronic pain. Smith-Seemiller et al. compared a mTBI group with a chronic pain group, and whilst the total scores from the Rivermead Post-Concussion Questionnaire were not significantly different, again there were differences between the groups for specific areas such as those with mTBI suffered more cognitive problems and those with chronic pain suffered more emotional problems[16].

In one study, Mclean et al attempted to find predictors of PPCS. They summarised that a patient's baseline PCS predicted PPCS, whilst the presence of a head injury did not. They also believe that post-traumatic stress disorder (PTSD) due to minor injury events may play a significant role[17].

Compared to patients without PPCS, those with PPCS typically have a lower health related quality of life, higher health care use and lower return to work rates. The fact that PPCS can be so debilitating and continue to trouble people for many years, causing a significant socio-economic burden, also highlights how important it is to do further research into this area to find ways to optimise recovery. Progress, however, is being hampered by inconsistent assessment and classification. Research should be

aimed at improving imaging techniques to provide visual evidence of any changes that happen in the brain after a concussion. The role that genetics may play in the outcome after a concussion and the affect that pre-concussion health may have on outcomes also require investigation. Crucially, since assessment of PPCS is heavily based on self-reporting of patients, it is imperative that standardised assessment strategies should be agreed.

3.4 Treatment

Seeking relief for PPCS patients often requires a multi-faceted approach due to the complexity of the problem. Difficulties can stem from different areas of the body. A concussion usually causes a whiplash effect, so headaches can be caused by referred pain from the neck, requiring physiotherapy. Patients can also have visual and vestibular dysfunction; issues that equally need to be addressed. Moreover, patients often suffer from psychological difficulties such as anxiety and depression after a concussion, so cognitive behavioural therapy is one of the most common forms of treatment given. Unfortunately, this typically only provides limited relief.

Historically, patients have been advised to limit exercise immediately after a concussion, but research is now suggesting that doing so is counter-productive. Aerobic exercise is considered to be an important part of recovery as it improves blood flow, neuronal function, autonomic regulation and brain healing modulators (BDNF). However, concussion patients frequently have difficulty exercising as it worsens their symptoms[18].

Hyperbaric Oxygen Therapy

Typically those suffering PPCS are given treatment to help reduce their symptoms but there is little known about how to cure the root cause of their issues. Hyperbaric oxygen therapy (HBOT) is the only treatment aimed at curing PPCS by attempting to repair defects in the brain caused by the body's response to a concussion. HBOT is widely known for treating decompression sickness: Behnke and Shaw were the first to successfully treat it in 1937. In 1955, Churchill-Davidson used HBOT to enhance the effects of radiation therapy in cancer patients[19]. Since then, as the mechanisms of HBOT have been further understood, it has been used to treat a variety of medical conditions, including our most recent medical challenge, Long Covid. Other conditions treated by HBOT include multiple sclerosis, fibromyalgia, crush injuries, cerebral palsy, migraines, ulcers and even Lyme disease[20]. However, there is a great deal of controversy regarding the efficacy of its use for certain conditions, one of which is PPCS. Its use for PPCS is therefore considered to be an ideal topic for performing a meta-analysis.

As mentioned previously, after a concussion there is swelling, which causes a reduction in blood flow to the site of injury, and consequently reduces the availability of oxygen required for recovery. Many illnesses and injuries occur and linger at a cellular level since the body's healing mechanism is unable to function properly. During HBOT, rather than breathing the usual 21% oxygen at normal air pressure, a patient breathes in close to 100% oxygen at typically 1.5 or 2.0 ATA (equivalent to 16.5ft or 33ft under water). These conditions allow the body to dissolve up to 10 times the normal concentration of oxygen in the blood plasma, cells, tissues and fluids, finally providing the energy the body needs to repair. By forcing more oxygen into the tissues, new blood vessels are encouraged to grow and the diameter of blood

vessels increases, which in turn provides better access for oxygen to reach the cells that need repair. It also stimulates cells to produce substances such as vascular endothelial growth factor, which attract and stimulate endothelial cells for healing. Similarly, it enhances the ability of white bloods cells to resist and fight infection. In essence, HBOT aims to break the cycle of swelling, oxygen starvation and tissue death and encourages healing[21].

In the remainder of this project, meta-analysis techniques are used with the aim of determining whether or not the statistics are in favour of hyperbaric oxygen therapy being a useful treatment for PPCS.

4 Literature search and selection

In this chapter, research is started with a literature search, using the recommended guidelines mentioned in section 1.2.

4.1 Research Question

The research question has been chosen to be "Does hyperbaric oxygen therapy reduce symptoms of PPCS". During the scoping search, it became evident that there may be a lack of research on the use of HBOT for PPCS. The search for studies was kept as broad and simple as possible due to time constraints. The University of Strathclyde's SUPrimo search tool was used to find appropriate research papers, as these papers must reach a certain standard to be included in the database, according to library services. Only papers written in English were searched to avoid translation issues (so a language bias may exist) and no citation chaining was used.

4.2 Eligibility Criteria

Any case-control studies that used HBOT as a treatment for PPCS were considered to be of interest, provided treatment and control effects were given or could be calculated. The only restriction made on the population to be included was that they had been suffering from PPCS for at least 3 months after concussion. HBOT at 100% oxygen and 1.5 ATA, or an equivalent combination, was chosen for the eligibility criteria as this seemed to be the most commonly used therapy studied, and studies done over an 8 to 12 week treatment period, with at least 5 treatments per week were included, but only 1 treatment per day was allowed. End of study tests had to have been completed within 2 months after treatment. Outcomes required were any appropriate tests for PPCS.

4.3 Search technique

The search string used was *("hyperbaric oxygen" OR HBO) AND ((post AND concuss*) OR postconcuss*)* and the search was made on 21st December, 2022. This provided 183 results. Choosing "Full text online" reduced this to 82 results, of which there were 59 in "peer-reviewed journals". As shown in figure 6, there were 5 duplicates leaving 54 to be screened. 33 papers were irrelevant, and for 1, only an abstract was found. This left 20 papers to be assessed properly for suitability.

For the meta-analysis, case-control studies were desirable; ideally randomised controlled trials. 15 were excluded for unsuitable study designs, including 11 case studies, 1 retrospective study and 3 other studies with issues such as inappropriate outcomes or interventions. The 5 studies that were kept for the meta-analyses included 3 randomly controlled studies; Hadanny[22], Cifu[23] and Miller[24] and 2 cross-over studies; Boussi-Gross[25] and Harch[26]. In the cross-over studies, effectively only the first half of the experiment was used: The treatment group baseline to post-HBOT effect was compared with the baseline to pre-HBOT effect of the cross-over group so that no dependency issues arose by including a sample twice in the study. Whilst Miller's study states it was double blinded, the chamber operators knew the chamber treatment allocation.

There were 11 case studies found in the search results. It is suggested that the increased bias from the typically poorer study designs, and potential for voluntary inclusion make it unusual to include case series studies in a meta-analysis as it lowers the level of evidence. Since there is no *effect difference* in case series, a different meta-analysis model is required. Therefore case studies were not included in this study.

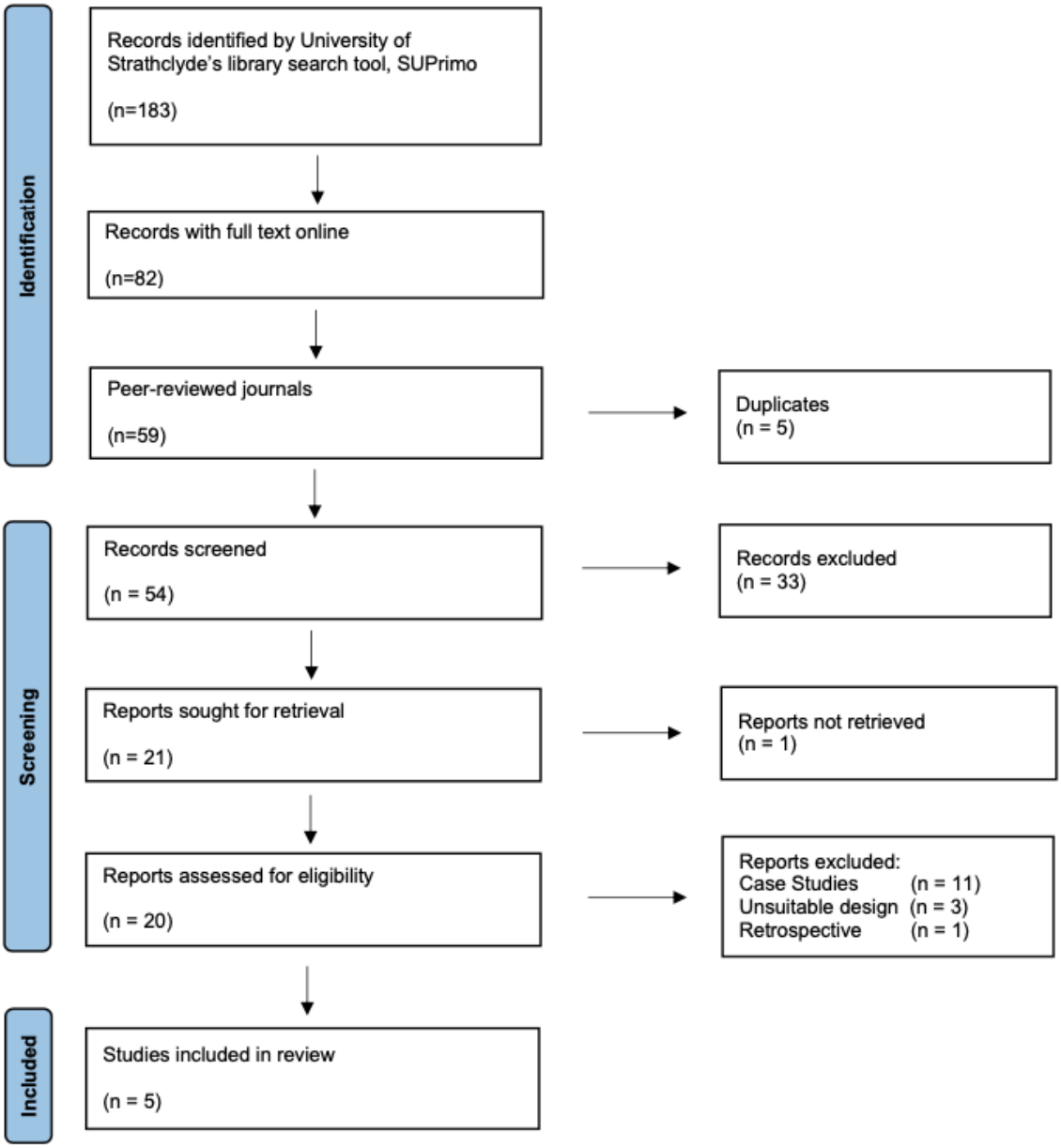

Figure 6: PRISMA flow chart for meta-analysis

5 Data extraction

In the previous chapter, study selection was described. This chapter details the extraction of information from the papers and the calculations that are made to achieve the appropriate effect sizes to input into the meta-analysis model.

5.1 Data extraction tables

Once the choice of studies for inclusion in the meta-analysis has been established, the relevant data and information can be extracted into data extraction tables (tables 2 to 5). General study information can be found in Appendix A and includes study design, population type, primary outcomes, information on the intervention and sham/control treatments, the types of tests used and date of publication.
There were three papers from which effect sizes could be obtained for tests of Memory Index and Executive Function, both of which are commonly impaired in PPCS patients. These papers were authored by Boussi-Gross[25], Harch[26] and Hadanny[22].
All five studies included a measure of PPCS, which were pooled together, despite being drawn from different tests.

5.2 Effect sizes

Study results are frequently presented in different formats: Some may provide a standardised effect size, whilst others provide a raw effect size. Some studies report a confidence interval whilst others may provide a standard error or a p-value, for example. Equally, studies may provide no treatment effect size at all, but have enough information for one to be calculated. Tables 2 and 3 show data for Memory Index and Executive Function treatment effects. Table 4 shows data for PPCS.

Table 2: Memory Index data table

First Author	n_A	n_B	D	s_{pool}	p	CI	d	SE_d
Boussi-Gross	32	24	-	-	< 0.0005	-	0.73	**0.209**
Harch	23	27	6.92	8.6	0.0067	2.01 to 11.83	**0.8**	**0.295**
Hadanny	15	10	-	-	0.017	-	0.48	**0.201**

Table 3: Executive Function data table

First Author	n_A	n_B	D	s_{pool}	p	CI	d	SE_d
Boussi-Gross	32	24	-	-	< 0.0005	-	0.66	**0.189**
Harch	23	27	1.97	5.8	0.2384	-1.36 to 5.28	**0.34**	**0.286**
Hadanny	15	10	-	-	0.830	-	0.13	**0.567**

Table 4: PPCS data table. The sign of Miller's value of Cohen's d contradicts his description of the treatment reducing symptoms.

First Author	n_A	n_B	D	s_{pool}	p	CI	d	SE_d	improve
Boussi-Gross	32	24	-	-	-	-	**1.07**	**0.288**	yes
Boussi-Gross*	32	24	**1.41**	**2.15**	-	-	**0.66**	**0.277**	yes
Harch	23	27	-23.9	9.22	0.0001	-29.2 to -18.6	**-2.59**	**0.384**	yes
Hadanny	15	10	-	-	0.02	-	-0.17	**0.073**	yes
Cifu	18	21	-	-	-	-	**0.15**	**0.322**	no
Miller	24	23	-	-	-	-	**0.30***	**0.293**	yes

In the tables, n_A and n_B are the sample sizes of the treatment and control groups. D is the raw mean effect, s_{pool} is the pooled standard deviation. CI is the confidence interval; p is the p-value; d is the standardised effect, Cohen's d; and SE_d is the standard error of this effect.

Treatment effect sizes were not given in three papers, so the data used to calculate the treatment effect is shown in table 5. In all tables, bold type indicates data that have been calculated, rather than provided in the studies. In table 4, Cohen's d for Miller's study is starred as the sign of the group effects given by the data were contrary (for NSI scores) to their description of the treatment providing improvement: A reduction in NSI score is normally considered an improvement, as shown in Harch's study. Also, the starred Boussi-Gross data were calculated a different way, as explained later in this chapter, due to a potential discrepancy in the given results.

Table 5 shows the mean group scores of group A (treatment) and group B (control) before and after the intervention; the group mean effect size; its confidence interval; its associated p-value from a paired t-test and the t-statistic.

Table 5: PPCS Data to calculate group effect sizes

First Author	$\bar{x}_{A_{bef}}$	$\bar{x}_{A_{aft}}$	Mean effect, $\bar{x}_A$	CI	p_{pair}	t
Boussi-Gross	5.03	6.62	**1.59**	-	< 0.0001	**4.46**
Cifu	24.29	25.38	**1.09**	-	0.21	**1.30**
Miller	-	-	3.7	-3.7 to 11.2	-	**2.07**

First Author	$\bar{x}_{B_{bef}}$	$\bar{x}_{B_{aft}}$	Mean effect, $\bar{x}_B$	CI	p_{pair}	t
Boussi-Gross	5.26	5.21	**-0.05**	-	0.373	**0.909**
Cifu	27.57	27.76	**0.19**	-	0.91	**0.11**
Miller	-	-	-1.1	-7.3 to 5.2	-	**2.07**

5.2.1 Memory Index and Executive Function

For the first two meta-analyses of Memory Index and Executive Function, Boussi-Gross' and Hadanny's papers provide a standardised mean difference (Cohen's d), whilst Harch's paper provides results of the raw mean difference, D. Harch's results include values of $D = 6.92$, $s_{pooled} = 8.6$ and sample sizes $n_A = 23$ and $n_B = 27$. Equation 3 can therefore be used to find SE_D to be 2.44. d and SE_d can then easily be found using equations 4 and 5.

Also, Boussi-Gross' and Hadanny's papers provide a p-value rather than a standard error associated with its value. The theory for these conversions has already been mentioned in section 2.1.5. These conversions were calculated in R using the *se.from.p* function in the *dmetar* package. The function requires the treatment effect size, its p-value and the total study sample size.

5.2.2 PPCS

For the meta-analysis of PPCS, results from four different tests were pooled together as they were all measuring similar symptoms. They included EuroQol Visual Analogue Scale (EQ-VAS), Neurobehavioral Symptom Inventory (NSI), Health and Behaviour Inventory (HBI) and Rivermead Post-Concussion Symptom Questionnaire (RPQ) as shown in table 6. These results must therefore be viewed with a bit of caution.

For Boussi-Gross' study, a treatment group mean effect, $\bar{x}_A$, was calculated for EQ-VAS by taking away the group's baseline score from its post-treatment score, whilst the control group's mean effect, $\bar{x}_B$, was calculated by taking away the control group's baseline score from the group's *before treatment* score (as this study used a cross-over design). The study provided p-values for the paired comparisons of the second

Table 6: PPCS test type

First author of study	PPCS test type
Boussi-Gross	EQ-VAS
Harch	NSI
Hadanny	HBI
Cifu	RPQ-13
Miller	NSI

measurement to baseline in each of these cases. The t-statistic, t_A (for group A) can be found from the p-value given for the paired t-test from statistical tables, an inverse t-distribution calculator, or using R (see code in Appendix C). Then the standard deviation of the treatment group effect, s_A, can be calculated using equation 60. The control group standard deviation, s_B can be found in a similar way. Equations 1 to 5 can then be used to find the standardised treatment effect and its standard error, for pooling in the meta-analysis.

$$s_A = \frac{\bar{x}_A \times \sqrt{n_A}}{t_A} \tag{60}$$

For Cifu's study, treatment and control group mean effect sizes were found in a similar manner, and again, using the same equations, the standardised treatment effect can be found.

Miller's paper provides data for the group mean effects with a 95% confidence interval. In this case, the standard deviation of the group mean effect can be found using its given confidence interval using equations 25 and 26. Using equations 2 to 5, the standardised treatment effect and its standard error can be calculated to input into the meta-analysis model. The standard treatment (no chamber) group was used as

the control group in this study instead of the sham treatment group, since chamber pressure of 1.2ATA is argued to be a treatment in itself.

There is however, a potential issue with the results given in Boussi-Gross' study. Figure 7 shows his results for EQ-VAS. If the p-value of 0.373 is used to find the difference of means between baseline and pre-HBOT measurements for the control group, by the method mentioned above, the standard deviation of this difference is 0.270, which is very small compared to the standard deviations of each of the baseline and pre-HBOT scores, 1.70 and 1.66, which is concerning.

	Treated group (n = 32)				Crossover group (n = 24)					
	Baseline	HBOT	P1	P2	Baseline	Control-Pre HBOT	Post HBOT	P2	P3	P4
EQ-5D	7.87±1.36	6.48±1.07	0.615	<0.0001	7.70±1.11	8.06±1.05	6.75±1.06	<0.01	<0.0001	0.362
EQ- VAS	5.03±2.31	6.62±2.45	0.696	<0.0001	5.26±1.70	5.21±1.66	6.39±1.80	0.373	<0.0001	0.696

Abbreviations:
Values are presented as mean ± STD. P1 stands for the p values for baseline comparison of treated and crossover group; P2 stands for the p values for comparison of the second measurement to baseline in the same group; P3 stands for the p values for comparison of pre- and post-HBOT in the crossover group; P4 stands for the p values for endpoint scores comparison following treatment in both groups. EQ-5D as well as the EQ-VAS scores significantly improved following HBOT, both in the treated group and in the crossover group following treatment, while there was no significant improvement following the control period.
doi:10.1371/journal.pone.0079995.t003

Figure 7: Results from Boussi-Gross' paper

Using equation 61, where s_1 represents the standard deviation of the control group scores at baseline and s_2 is the standard deviation at pre-HBOT, for $s_{diff} = 0.270$, the correlation, r, between the two sets of mean scores would need to be r=0.987. EQ-VAS results give scores between 0 and 100, therefore this seems highly unlikely that the scores would be quite so similar, even for a retest. It is possible that there is a discrepancy in the data for this control group.

$$s_{diff}^2 = s_1^2 + s_2^2 - 2rs_1s_2 \tag{61}$$

Since this is a *randomised* cross-over study, and the comparison of the baseline

scores from each group show that they are comparable with a p-value of 0.696, it is reasonable to calculate an effect size for this study an alternative way, by taking the mean difference from the post-HBOT scores of the treatment group and the pre-HBOT scores from the cross-over group. This gives a raw effect size of $D = 6.62 - 5.21 = 1.41$, and again using equations 2 to 5, the standardised effect and its standard error can be computed to use in the meta-analysis.

In this chapter, it has been demonstrated how study data can be extracted and converted into appropriate effect sizes ready for inclusion in the meta-analysis model. It has also been discovered that sometimes study data can have discrepancies that make calculating the effect size tricky. The decisions a researcher has to make regarding data used in order to include a study, or whether a study should be included at all, can in itself allow bias to enter the results. In the next chapter, the results of the meta-analysis models for HBOT treatment are described.

6 Results

In this chapter, the outputs of the meta-analysis models are analysed and forest plots produced to display the results. Then sensitivity studies and influence analyses are performed to detect studies that may be overly influencing the summary effect. Finally, publication bias is considered.

6.1 Model

Once the standardised effect sizes had been calculated and formatted in a data table, the meta-analysis model was run. The R code used for the meta-analysis can be found in Appendix C. It uses a function called *metagen* from the dmetar package[27]. A random-effects model was used as some heterogeneity was expected since the studies used different population groups and have been performed slightly differently. Three studies provided test scores for Memory Index and Executive Function whilst five studies included tests for PPCS. The outputs from the meta-analysis models are shown in table 7 and Appendix B.

6.2 Memory Index

The results for Memory Index follow. Studying the model output in table 7, it can be seen that the pooled effect size estimate for Memory Index from the random-effects model was d=0.64, ie. the control and treatment group means were 0.64 standard deviations apart. Its 95% confidence interval was between 0.22 and 1.06 standard deviations. This confidence interval may be generous as the Hartung-Knapp adjustment was used to try to control for uncertainty in the estimate of heterogeneity. Using Cohen's convention, 0.64 suggests a moderate to large treatment effect.

Table 7: Meta-analysis output - Memory Index

```
                      SMD      95%-CI      %W(common) %W(random)
Rahav Boussi-Gross 0.7300 [0.3204; 1.1396]       38.7       38.7
Paul G Harch       0.8000 [0.2218; 1.3782]       19.4       19.4
Amir Hadanny       0.4800 [0.0860; 0.8740]       41.9       41.9

Number of studies combined: k = 3

                              SMD       95%-CI       z|t  p-value
Common effect model       0.6390 [ 0.3841; 0.8938] 4.91 < 0.0001
Random effects model (HK) 0.6390 [ 0.2215; 1.0564] 6.59   0.0223
Prediction interval              [-1.0133; 2.2913]

Quantifying heterogeneity:
 tau^2 = 0 [0.0000; 1.0635]; tau = 0 [0.0000; 1.0312]
 I^2 = 0.0% [0.0%; 89.6%]; H = 1.00 [1.00; 3.10]

Test of heterogeneity:
 Q      d.f.   p-value
 1.11    2     0.5732

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp (HK) adjustment for random effects model (df = 2)
- Prediction interval based on t-distribution (df = 1)
```

This effect was considered statistically significant as p=0.022, using a 5% significance level. The number needed to treat, NNT, which is commonly used to signify how many patients must receive treatment to prevent one additional negative event, was approximated by Kraemer and Kupfer's method using only a standardised effect size. Here $NNT = 3$. In other words, 3 patients must be treated to achieve one patient who successfully has a reduction in symptoms.

The restricted maximum-likelihood method was used to estimate between-study heterogeneity as this method is known to work well with continuous data. The variance of true effects, $\tau^2 = 0$, (95% CI: 0.00 to 1.06). This suggests that there was effectively no between-study heterogeneity in the data. The standard deviation of the true effect sizes, $\tau = 0$ (95%CI: 0.00 to 1.03). This was very surprising as heterogeneity in the results was expected. Cochrane's Q test is used to test if the variation in a meta-analysis significantly exceeds the amount we would expect under the null hypothesis of no heterogeneity. Here Q=1.11 and the p-value of 0.573 suggests that there was no statistically significant heterogeneity. The percentage of variability in the effect sizes that wasn't caused simply by sampling error, $I^2 = 0\%$, however it had a huge confidence interval of 0 to 89.6%! There may well have been limited heterogeneity in the studies, however given the large confidence interval, it is likely that the studies were too small to have enough power to estimate heterogeneity well.

The t-test statistic for the pooled effect size was higher for the random-effects model than for the fixed-effect model (which the model output also gives us) and the confidence interval for the fixed-effect model was different, despite the effect estimate being the same, which is unexpected. This must be due to the Knapp-Hartung adjustment, which is based on a t-distribution, that was included in the model to control for uncertainty in the estimate of heterogeneity. When this adjustment was excluded from the model for Memory Index, the significance test of the pooled ef-

fect followed a normal distribution instead, and the random-effects results became identical to the fixed-effect results, with a z-statistic of 4.91 and p-value of < 0.001. As shown in figure 8, using the model without the Knapp-Hartung adjustment, all the effect sizes favoured treatment and the effect sizes were fairly similar. None of the study effect confidence intervals crossed the null value (zero for continuous data), showing that all study effect sizes were statistically significant. The confidence intervals all overlapped, showing the lack of heterogeneity. However the standard errors were all fairly large and the prediction interval was d= -1.01 to 2.29 standard deviations, suggesting that future studies may find both positive and negative treatment effects based on the current evidence.

The study weightings are represented by the size of the squares and the precise weights are displayed in the figure also. The highest study weighting was given to Hadanny's study with 41.9%, which had the smallest sample size of only 25 (but unusually had the smallest sampling error), closely followed by Boussi-Gross's study with 38.7%. The lowest weighting was given to Harch's study with 19.4%.

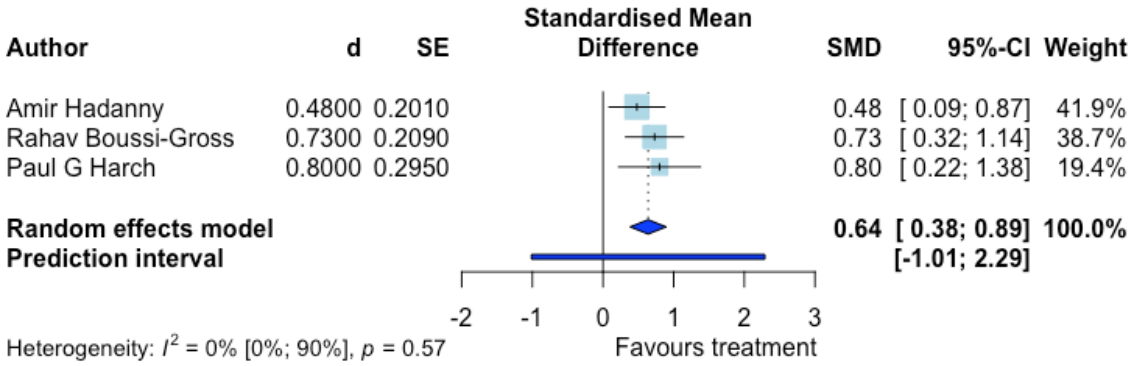

Figure 8: Forest plot for Memory Index

6.3 Executive Function

The model output for Executive Function can be found in table 10 in Appendix B. The pooled estimate of Executive Function, given by the random-effects model was d=0.53 (95% CI: -0.02 to 1.08) with a close to statistically significant p-value of 0.053. The fixed-effect model estimate was also d=0.53 but with a different confidence interval (95% CI: 0.23 to 0.83). The fixed-effect p-value of < 0.001 is statistically significant. For the Executive Function effect, when excluding the Knapp-Hartung adjustment the random-effects results also became identical to the fixed-effect results, with a z-statistic of 3.50 and p-value of < 0.001. Given the results showed zero heterogeneity, the random-effects results were expected to be the same as the fixed-effect result, so it made more sense to exclude the adjustment when no heterogeneity was found. Using Cohen's convention, an effect size of 0.53 suggests a moderate treatment effect. For Executive Function, NNT=4, so 4 patients must be treated to see symptom improvement in 1 patient.

For Executive Function, Q=1.41 with a p-value of 0.494 suggesting that there was no more heterogeneity than would be expected under the null hypothesis of no heterogeneity. The between-study heterogeneity variance was estimated at $\tau^2 = 0$ (95%CI: 0.00 to 2.66), with an I^2 value of 0.0% (95% CI: 0.0 to 89.6%). The prediction interval ranged from d= -1.40 to 2.46, indicating again that negative intervention effects are possible in future studies.

Figure 9 shows that both Hadanny's and Harch's studies had a statistically non-significant effect size, as their effect size confidence intervals spanned the null value, despite their effect estimates both being positive. Boussi-Gross' study, however, had a statistically significant positive effect size and since it had the highest study weighting with 64.6% it contributed the most to the pooled effect size. The lowest

weighting was given to Hadanny's study with just 7.2% and this study showed the smallest effect size. It is worth noting here that Hadanny's study had the highest weighting for the Memory Index effect size. So clearly, in Haddany's study there was comparatively low within-study variance (s_k^2) for the memory test but comparatively high within-study variance for the Executive Functioning test.

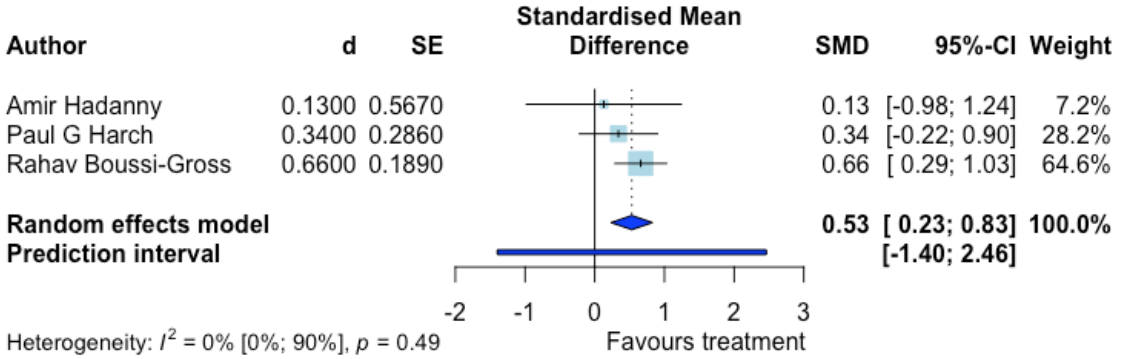

Figure 9: Forest plot for Executive Function

6.4 PPCS

The model output for PPCS can be found in table 11 in Appendix B. The pooled effect estimate using the random-effects model was d=0.77 (95% CI: -0.57 to 2.11) with a statistically non-significant p-value of 0.186. This can also be seen in figure 10, as the width of the diamond, representing the pooled effect and its confidence interval, crosses the null value. However, using Cohen's convention, 0.77 suggests there was quite a high treatment effect. This is a case where we have a strong treatment effect but it isn't statistically significant. It would therefore be prudent to do further research with a larger sample size (or use studies that inlcude the same tests for PPCS!) for there to be enough power to show a statistically significant

result, if there is one, as the potential treatment effect looks promising.

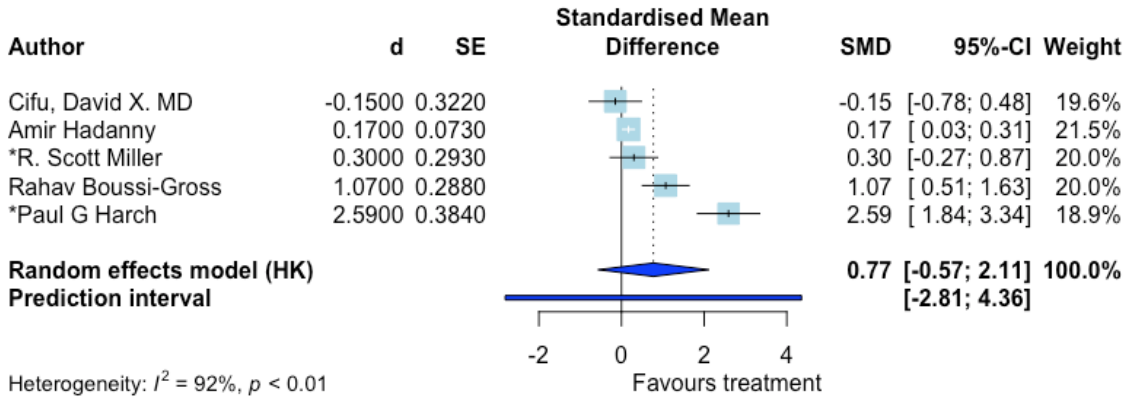

Figure 10: Forest plot for PPCS. The starred studies both used the same type of test.

In this meta-analysis the Hartung-Knapp adjustment was used as the results were notably heterogenous. Q=47.77 with a p-value of < 0.0001 suggesting that there was very high heterogeneity. The between-study heterogeneity variance was estimated at τ^2=1.04 (95%CI: 0.31 to 9.84), with an I^2 value of 91.6% (95% CI: 83.4 to 95.8%). The prediction interval ranged from d= -2.81 to 4.36, indicating that both positive and negative intervention effects are possible in future studies.

In figure 10 it can be seen from the 95% confidence intervals that three studies showed statistically significant results in favour of hyperbaric oxygen treatment, but Cifu's and Miller's studies did not. Also, the confidence intervals for Hadanny's study were very small compared to the others, showing small within-study variance, despite a small sample size. The two starred studies, Miller's and Harch's, both used the same test type, NSI. These two studies have quite different results.

The study weightings were all fairly balanced, with Hadanny's study having the most

at 21.5% and Harch's study having the lowest at 18.9%. Haddany's study had the smallest sample size with just 25, whereas the other studies had between 39 and 56 participants included. It is perhaps worth noting here that the random-effects model pays comparatively more attention to small studies when calculating the pooled effect size. Given that small studies can often have more biases attached to them, the pooled results from a random-effects model can potentially be more biased.

Figures 11 and 12 show the results when the alternative calculation for Boussi-Gross' effect size was used, and when excluding Boussi-Gross' study entirely. The new effect size for Boussi-Gross' study was 0.66 (95% CI: 0.12 to 1.20), which is statistically significant. The study weightings also changed slightly. Including this alternative effect calculation for Boussi-Gross' study in the meta-analysis reduced the overall pooled effect to 0.69 (95% CI: -0.63 to 2.01), with a statistically non-significant p-value of 0.222. The lower prediction limit didn't change much with the new calculation, however, the upper prediction limit was reduced comparatively more (new PI: -2.83 to 4.21). Since Boussi-Gross' study effect was very similar to the pooled effect size, it can be seen in figure 12 that removing the study entirely did not affect the size of the pooled effect much. However, it markedly increased the size of the 95% confidence interval from (-0.63 to 2.01) to (-1.27 to 2.67). Similarly the prediction interval increased in size to (-5.03 to 6.44).

Since Harch's effect size was so high compared to the other studies, there was concern that there was an inconsistency in the data somewhere. Harch's and Miller's studies both used the NSI test, so their results should be fairly similar, but they aren't. These studies were completed over 5 years apart so it is possible that the NSI test has evolved within this time frame, although no evidence of this was found. Otherwise, there are three versions of the NSI test according to three timeframes: one is for symptoms since time of injury, one for symptoms in the past month and one for

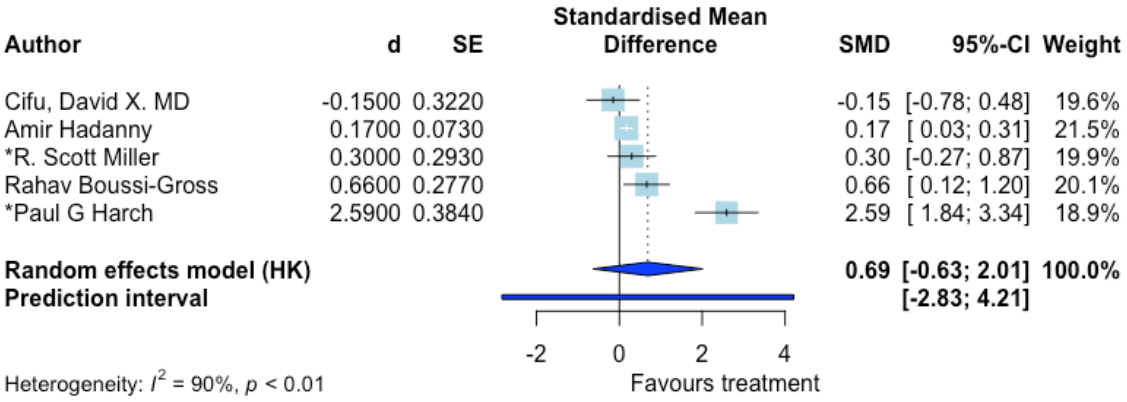

Figure 11: Forest plot for PPCS - alternative calculation for Boussi-Gross

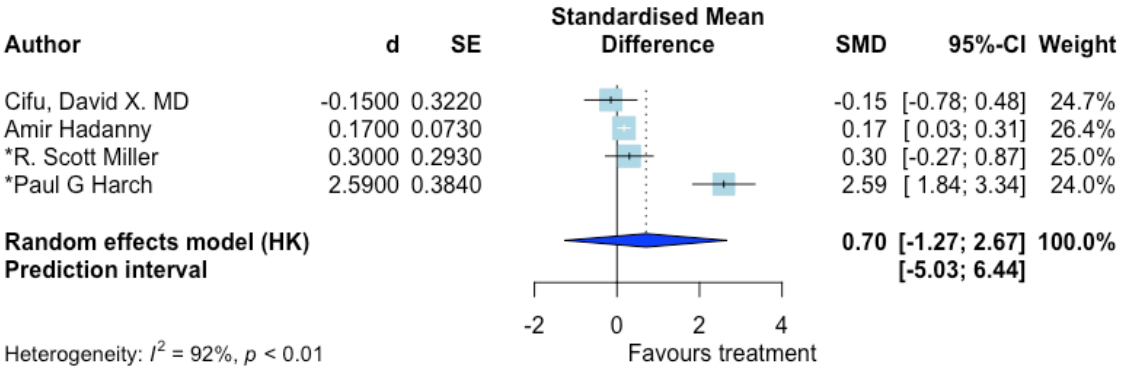

Figure 12: Forest plot for PPCS - Excl. Boussi-Gross

symptoms in the last two weeks. The precise test used doesn't appear to be confirmed in the studies. It was interesting to see how the results changed when Harch's study was excluded (figure 13). The new calculation of Boussi-Gross' effect size has been kept in this meta-analysis and those that follow.

Excluding Harch's study, the model gave a value of Q=4.21 with a p-value of 0.240 suggesting that there is no statistically significant heterogeneity in this model. This can be seen visually as the confidence intervals overlap. The between-study heterogeneity variance was estimated at $\tau^2 = 0.01$ (95%CI: 0.00 to 1.47), with an I^2 value of 28.7% (95% CI: 0.0 to 73.7%). The pooled effect size became 0.20 (95% CI: -0.12 to 0.53), which Cohen would classify as a small positive treatment effect. This effect is not statistically significant as its confidence interval spanned zero. However, with a larger total sample size providing more power, this smaller effect may become statistically significant. The prediction interval was -0.28 to 0.69 so both positive and negative effects would be expected in further studies. The most alarming part of this result was that Hadanny's study was attributed a 75% weighting! This didn't happen when Boussi-Gross' study was excluded, where the weightings remained fairly balanced. The standard error of the treatment effect for Hadanny's study was suspiciously low given how small the study was and its population was different too, being a study of children, so it was worth seeing how the results changed if we excluded Hadanny's study from the model (figure 14).

The meta-analysis of the remaining three studies provided a more balanced weighting, with 30%, 34% and 36% since their sample sizes and standard errors were more comparable. The pooled effect size was a statistically non-significant 0.29 (95% CI: -0.71 to 1.29) and the prediction interval was about -4.2 to 4.8. The effect was 0.09 higher than when Hadanny was included. Q=3.64 with a p-value of 0.16, showing there is no significant heterogeneity, and the percentage of variability in the effect

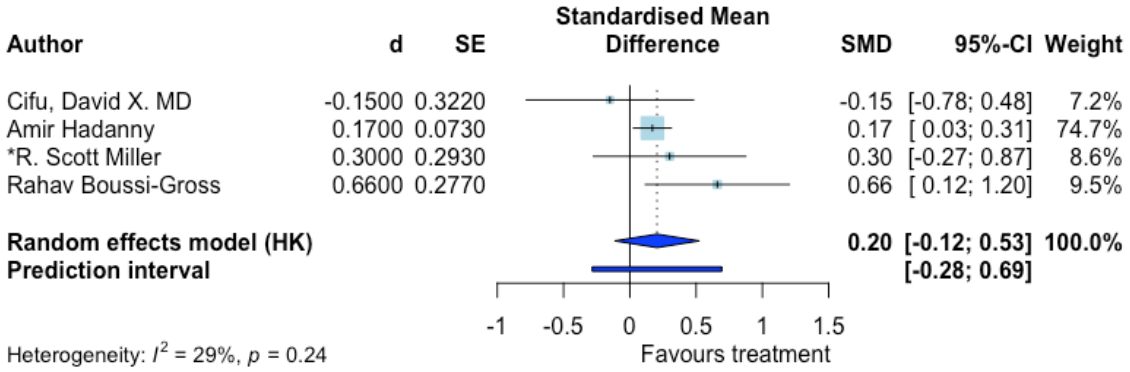

Figure 13: Forest plot for PPCS - Excl. Harch

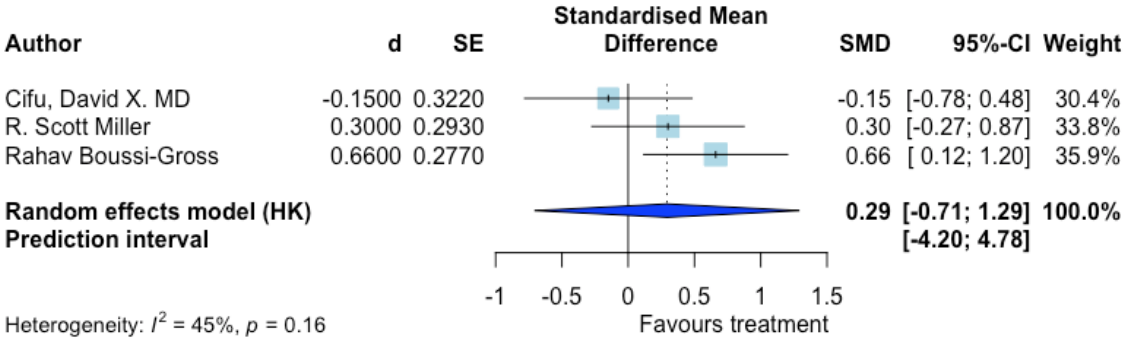

Figure 14: Forest plot for PPCS - Excl. Harch and Hadanny

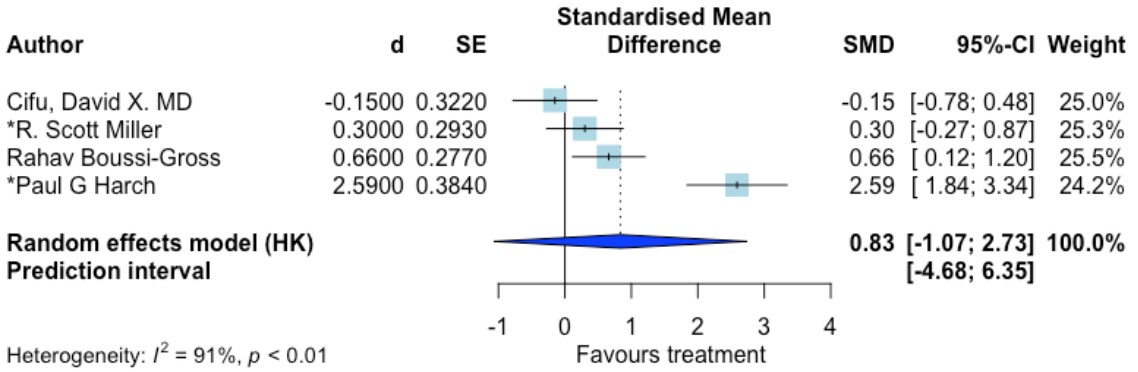

Figure 15: Forest plot for PPCS - Excl. Hadanny

sizes that isn't caused by sampling error, $I^2 = 45\%$.

Figure 15 shows the results when Hadanny's study was excluded from the model. Once again heterogeneity was very high with $I^2 = 91\%$ and the effect size was the largest found, at 0.83 (95% CI: -1.07 to 2.73). Again, this result is not statistically significant as the confidence interval crossed the null value of zero effect, so no treatment effect was found.

6.4.1 Sensitivity studies

Table 8 compiles the PPCS results from the meta-analyses produced above. Over all the PPCS models, the pooled effect ranged from 0.20 to 0.83, depending on which studies were included. The smallest prediction interval of -0.28 to 0.69 was found when excluding Harch's study, whilst the largest was in the model excluding Boussi-Gross' at -5.03 to 6.44. The heterogeneity ranged from $I^2 = 29\%$ to 92%. As stated earlier, these results should be viewed with caution as the models pool scores from different tests.

Table 8: PPCS results from sensitivity studies for the different models

Model	Effect	p-value	Heterogeneity, $I^2(\%)$	p-value
Original	0.77	0.19	92	< 0.01
Alt. Boussi-Gross	0.69	0.22	90	< 0.01
Excl. Boussi-Gross	0.70	0.34	92	< 0.01
Excl. Harch	0.20	0.13	29	0.24
Excl. Harch & Hadanny	0.29	0.33	45	0.16
Excl. Hadanny	0.83	0.26	91	< 0.01

The best model was perhaps the one excluding Harch's study. Individually removing each of the other studies did not change the effect size substantially, however, when removing Harch's study, the effect size reduced significantly from 0.77 to 0.20.

6.4.2 Influence analysis

The results for the PPCS outcome were very heterogenous so there was lower confidence in the pooled summary effect size. The summary effect can be overly influenced by outlying and influential studies. A simple way to spot *outliers* is to look for studies whose confidence intervals lie outside of the pooled effect confidence intervals. There didn't appear to be any outliers in these results. *Influential* studies can still exist though, and can be found using a *leave one out* method. This re-calculates the results of the meta-analysis K times, leaving out a different study, k, each time. Influence diagnostics can then display which studies change the pooled effect the most. There is a function called *InfluenceAnalysis* in the dmetar package in R that calculates influence diagnostic graphs from the meta-analysis model results. The function creates four plots. The first one is a Baujat plot (figure 16), designed to detect studies which overly contribute to *heterogeneity* within the meta-analysis. Studies towards

the right hand side of the graph are regarded as influential as they contribute the most heterogeneity. Studies in the top right of the graph are particularly relevant as they also highly influence the effect size. The study at the far right of figure 16 is Harch's, and it contributed significantly more heterogeneity to the model than any other study.

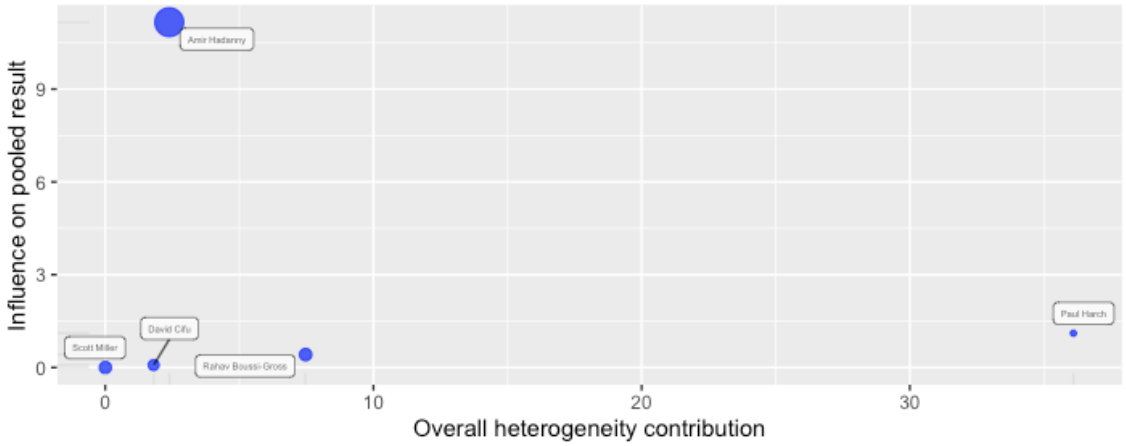

Figure 16: The Baujat plot shows which studies have a large influence on the pooled result and which contribute the most to heterogeneity. Far right study is Harch's.

The next plot it provides shows the externally standardised residuals (figure 17). These show the deviations of the observed study effect sizes from the pooled effect size. An *external* estimate of the pooled effect is made excluding the study being left out (hence a *leave one out method*). The resulting residual is then standardised by three values: the variance of the external pooled effect, the τ^2 estimate of this external effect and the variance of k, the study being left out. DFFITS indicates how much the pooled effect changes when the k^{th} study is removed, measured in standard deviations. Higher values show a more influential study. Cook's distance is

another measure used for influence, similar to DFFITS. The covariance ratio is found by dividing the variance of the external pooled effect by the variance of the pooled effect including all the studies in the meta-analysis. A study value below 1 indicates that the pooled effect size would be more *precise* if this study was removed from the meta-analysis. Studies shown in red are considered to be influential studies. Harch's was the only study with a covariance ratio below 1, indicating removal of the study would be preferred. τ^2 and Q show the heterogeneity values when each study was removed. As one would expect, lower values are preferred. Again, it can be seen in the bottom two graphs of figure 17 that it was Harch's study that had the most influential weighting of the studies and should be removed from the meta-analysis. The hat value is equivalent to the weighting so these graphs look much the same.

Figures 18 and 19 show the effect sizes and I^2 values when omitting each study. Please note that the values differ from the sensitivity studies done in section 6.4.1, as these ones used the original data whilst the sensitivity studies in section 6.4.1 used the alternative calculation for Boussi-Gross' study. The lowest effect was seen when Harch's study was omitted and the largest when Cifu's was left out. Removing Harch's gave the lowest heterogeneity. In these figures, the dotted green line shows the original pooled effect size and the shaded green area it's 95% CI. All influence analyses indicated that Harch's study should be removed.

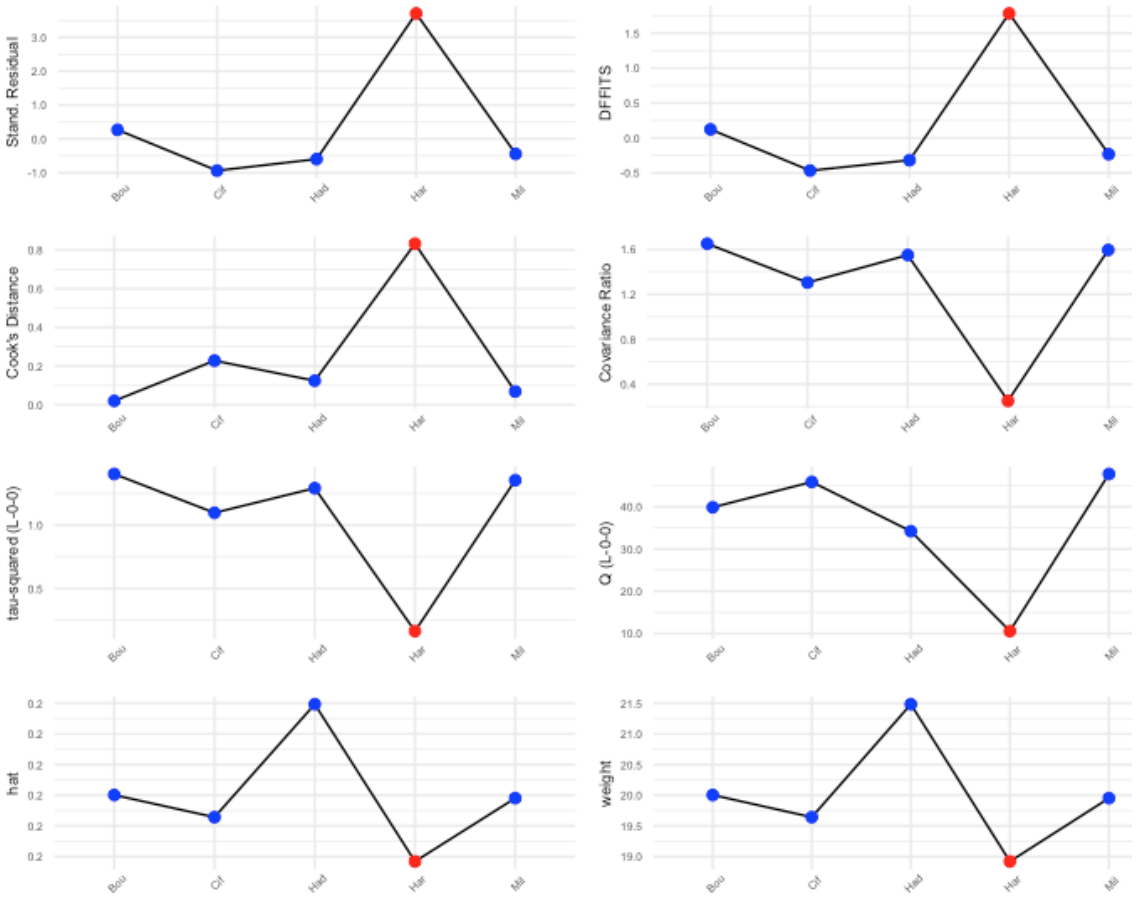

Figure 17: Standardised residuals plot. The names at the bottom, for each study are: Boussi-Gross, Cifu, Hadanny, Harch and Miller. These show the deviations of the observed study effect sizes from the pooled effect size where an *external* estimate of the pooled effect is made excluding the study being left out. Studies shown in red (Harch's) are considered to be influential studies.

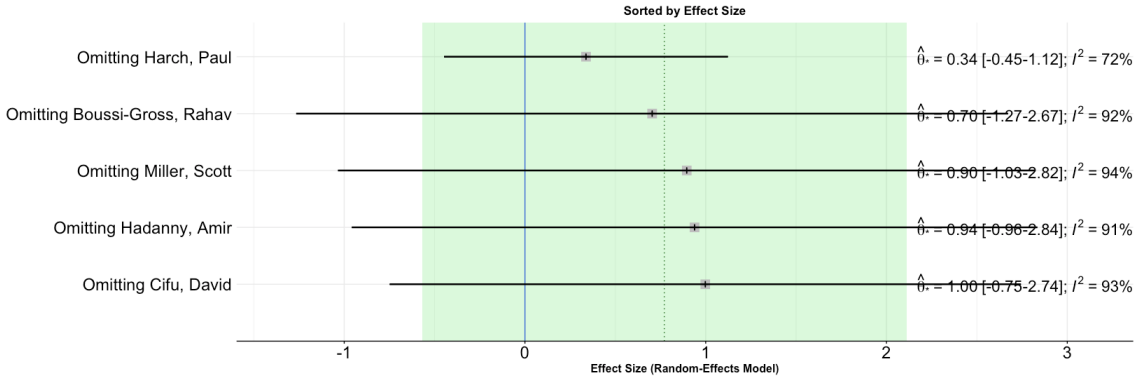

Figure 18: Forest plot of effect sizes: Leave-one-out meta-analysis results

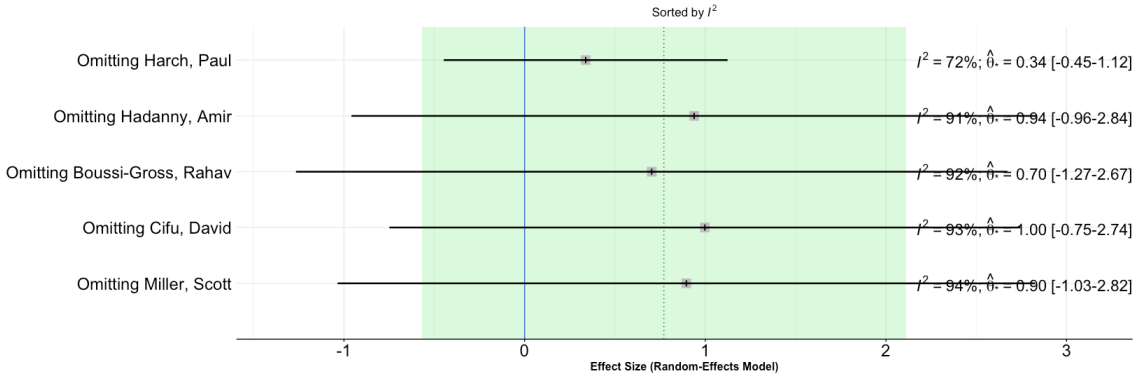

Figure 19: Forest plot of I^2: Leave-one-out meta-analysis results

6.5 Publication Bias

To consider how biased the included studies were, judgements were made with the help of Cochrane's Risk of Bias tool[28]. Figures 20 and 21 show the resulting carefully considered bias. In answering the signalling questions, two of the studies (Cifu and Miller) were considered to have a low risk of bias whilst the other three were assigned a medium risk, with some concerns, all due to questionable errors in the results. It is possible that there are other biases within the studies that exist but are not highlighted by these signalling questions. For instance studies where participants put themselves forward to take part may provide more biased results than participants who have been invited to participate.

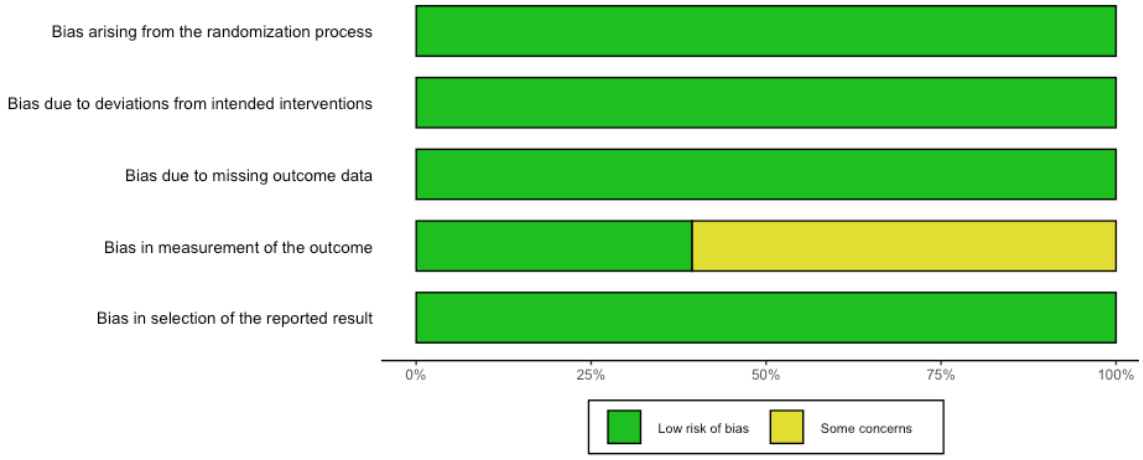

Figure 20: Risk of bias plot showing proportion of studies with a given risk of bias. 3 of the 5 studies show some concerns for bias in the outcome measurement.

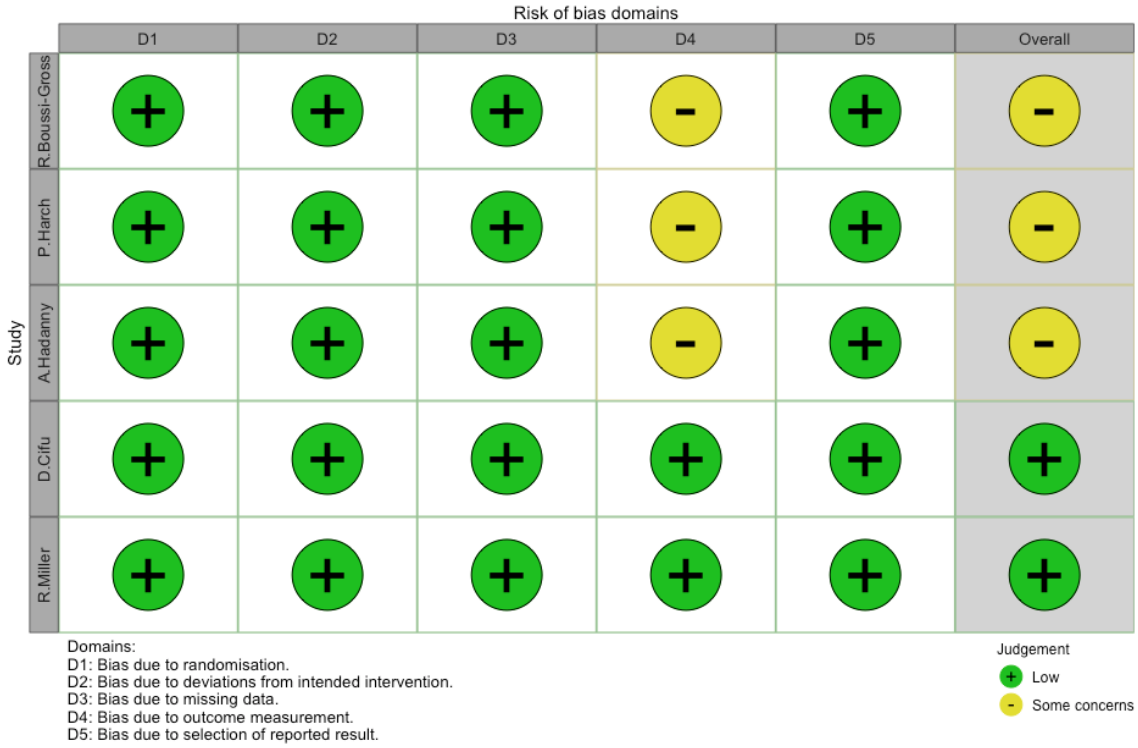

Figure 21: Traffic light plot showing the domain level judgement for each study. This plot shows Boussi-Gross', Harch's and Hadanny's studies cause some concern regarding bias due to the outcome measurement and this affects the study's overall judgement of bias.

In order to discern the likelihood of publication bias affecting the results, funnel plots were produced to look for asymmetry, using the original results (figures 22 to 24). A funnel plot is a scatterplot of the study's observed effect sizes (x-axis) against the standard error of their effect (y-axis). The vertical line shows the pooled effect size. If the plots look fairly symmetrical, with a balance of points on both sides of the funnel, this suggests that there is likely to be limited publication bias in the results. If however, the points are weighted more on one side, this indicates that there may be publication bias affecting the results, due to issues such as the file drawer effect or from only including papers published in a particular language.

There were only three results for Memory Index, so the funnel plot was expected to look slightly unbalanced, but as shown in figure 22 it looks as balanced as we might expect. The colours where the points sit represent the associated p-values of the studies. For example, the two studies to the right of the funnel have a value of $p < 0.01$ and the study to the left of the funnel here is in the orange region of $p < 0.05$, showing that the outcome measures from all the included studies are statistically significant at a 5% significance level.

Again the funnel plot in figure 23 for the Executive Function outcome, also looks fairly balanced given there are only three studies. This time only one study, that of Boussi-Gross, was statistically significant, with a value of $p < 0.01$. Studies in the white area (or red area where $p < 0.1$) are not statistically significant at the 5% level. Frequently, if a study's result falls into the white area, it is common for it to remain unpublished as its result is not deemed to be desirable. So the fact that there are two studies within this region suggests that this meta-analysis is unlikely to be affected by publication bias.

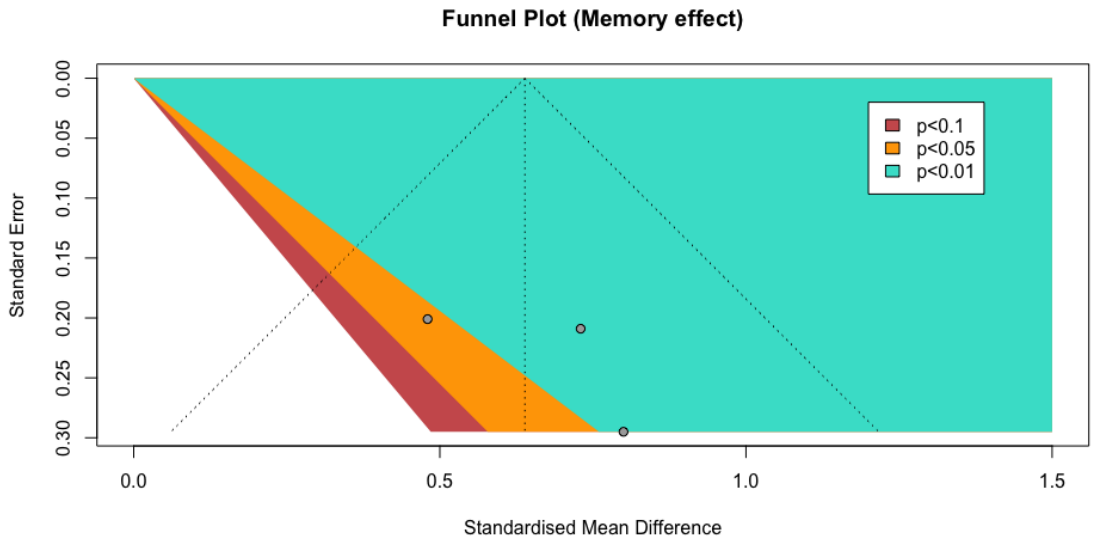

Figure 22: Funnel plot for Memory Index outcome

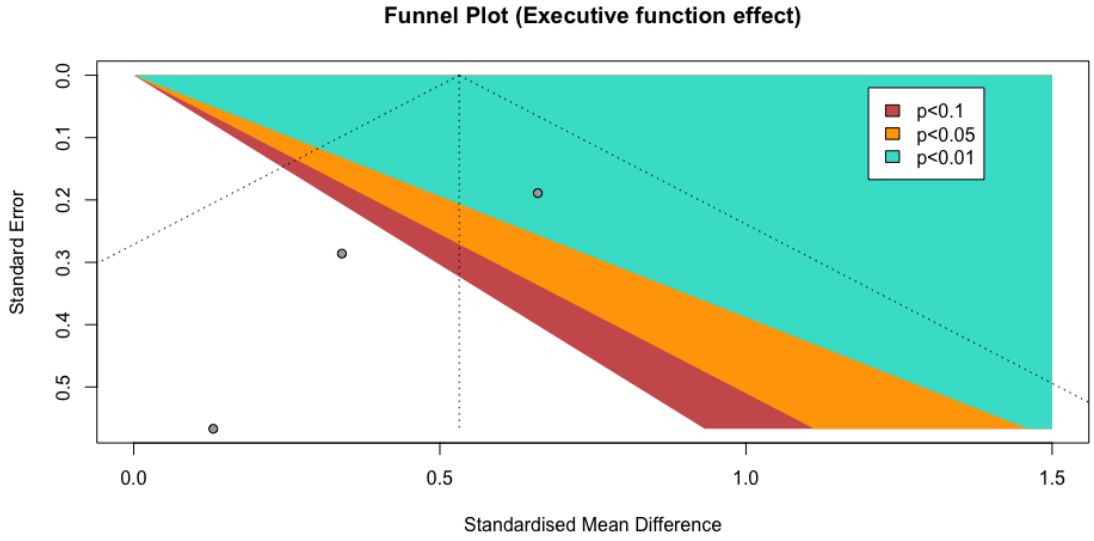

Figure 23: Funnel plot for Executive Function outcome

Figure 24 shows the funnel plot for PPCS. This plot looks more interesting. The funnel shown by the dotted line represents the expected distribution of study effects under the alternative hypothesis that there is a treatment effect, and the vertical dotted line shows the pooled effect size. The funnel formed by the distribution of the p-values, by contrast, is centred on the null hypothesis value of zero effect. It shows that two studies have a highly statistically significant effect with $p < 0.01$ and one study with $p < 0.05$ which is also statistically significant. Haddany's study, in the orange area looks like a bit of an anomaly as its standard error is so small, suggesting fairly high precision in its result. It is worth investigating why this standard error may be so different from the others. The other four study points are all quite dispersed and have relatively high standard errors. This dispersion of results is expected to happen when there is lower precision. These remaining studies are fairly balanced and roughly centred on the funnel. Two of the results lie in the white area, where the treatment effect is not considered to be statistically significant, so from this and the funnel plot being fairly symmetric, it is believed that there was a low publication bias likely in the resulting pooled effect.

When looking to understand Hadanny's study better, with the low standard error, there are perhaps plausible reasons why this study's results differed from the other studies. Firstly, this is the only study which recruited children rather than adults: The average age of participant was 11.6 ± 2.32 years, so the treatment may well work differently for children, and their responses may be more similar to one another. An alarming piece of information in this study is that 88% of the participants suffered from adverse effects during the study, although this is believed to be expected when treating children. Given that the PPCS tests question the participants about their health and wellbeing in the recent past, many of the test results may have been skewed due to participants' adverse effects, as this would affect their overall feeling

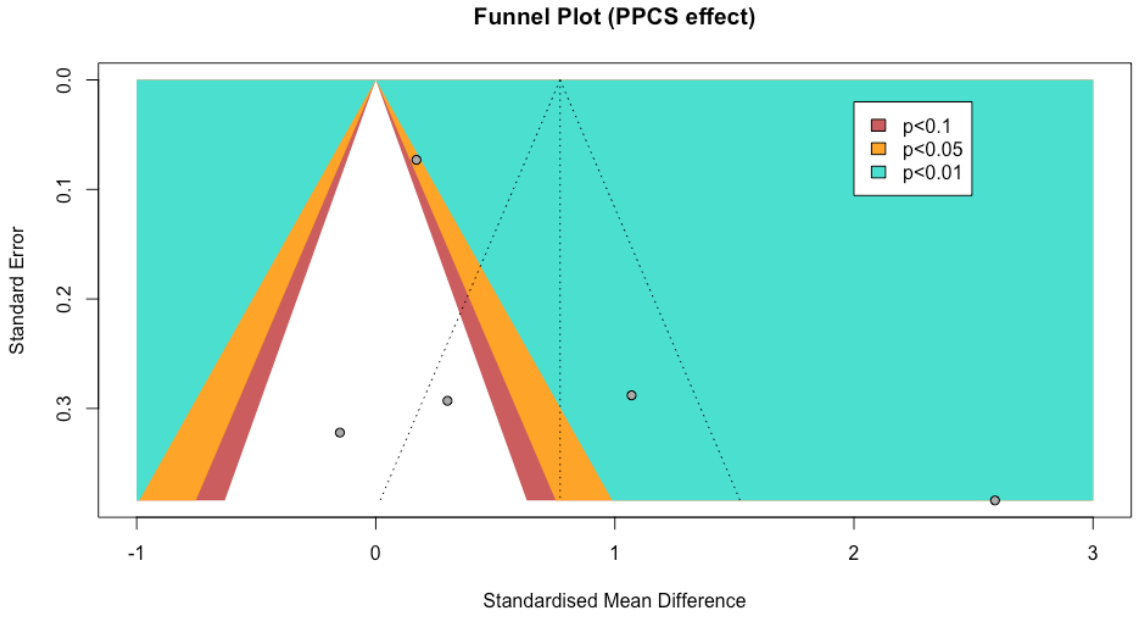

Figure 24: Funnel plot for PPCS outcome. Hadanny's study is in the orange area.

of wellbeing, and may lead to participants giving comparatively similar responses. However, with the standard error being so low, there was concern over whether there was an error in the results somewhere that hasn't been picked up.

In this chapter, the meta-analysis outputs for the Memory Index, Executive Function and PPCS outcomes were interpreted. Forest plots were compared and influence analysis was carried out to determine if there were any studies that were overly influencing the results. Risk of bias, traffic light and funnel plots were then presented to assess study bias. In the final chapter, the results are summarised and final thoughts are made on the meta-analyses.

7 Conclusion and Discussion

The processes involved in performing a systematic review have been presented, and the most common models used for meta-analysis have been described. The area of research relating to the use of hyperbaric oxygen therapy (HBOT) to reduce persistent post-concussion symptoms (PPCS) was introduced and meta-analyses of up to five studies was completed to summarise the treatment effect for three relevant outcomes. Ideally, a formal analysis plan would have been completed before proceeding, however, difficulty finding appropriate studies was expected, and it was considered best to avoid overly restricting the study. During the study, common difficulties were encountered such as issues when calculating effect sizes and handling study data that may appear to be inconsistent or lacking information.

The first two outcomes studied were measures of Memory Index and Executive Function. Heterogeneity in the meta-analysis models for these outcomes was almost non-existent, although the confidence intervals were large, which was likely due to the limited quantity of data. It was therefore appropriate to exclude the Knapp-Hartung adjustment from these models.

Results for Memory Index showed a moderate to large treatment effect size of d=0.64 (95% CI: 0.38 to 0.89) standard deviations, which was statistically highly significant ($p< 0.001$), and an NNT of 3. Every study contributed a statistically significant effect size. The prediction interval spanned zero so future studies may find positive and negative treatment effects. However, it is believed that with increased study power this prediction interval would gravitate towards being entirely positive. This meta-analysis indicates that HBOT treatment improves Memory Index scores.

Results for Executive Function provided a statistically highly significant ($p< 0.001$) moderate treatment effect of d=0.53 (95% CI: 0.23 to 0.83) standard deviations

and an NNT of 4. Therefore statistically speaking, HBOT treatment appears to be effective for improving Executive Function. However, only Boussi-Gross' study contributed a statistically significant study effect to the model, so confidence in this result is lower than for Memory Index. Again, increasing power of the study by including more data would be beneficial. Whilst less compelling, it certainly appears that HBOT treatment improves Executive Function scores also.

Results for the more general PPCS outcome, which pooled together results from five studies using four different tests were significantly heterogenous and required the random-effects model with the Knapp-Hartung adjustment. The original model for this meta-analysis gave a statistically non-significant effect size of 0.77 (95% CI: -0.57 to 2.11). However, after sensitivity studies and influence analysis, it was apparent that Harch's study was overly influencing the results and should be excluded from the analysis. Interestingly, this was the only study included that was not blinded for the patients. Removing Harch's study produced a non-significant small effect size of 0.20 (95% CI: -0.12 to 0.53) with no significant heterogeneity. Whilst this perhaps appeared to be the best model, retaining the most information, the results were a little concerning, as the model gave a weighting of 75% to Hadanny's study, with the smallest sample size. Removing Hadanny's study as well as Harch's from the model was perhaps preferable in this case, and increased the effect size from 0.20 to 0.29, and in turn the study weighting became balanced across the studies. The difference of 0.09 standard deviations in these effect sizes is appreciable, but the model likely to be most accurate cannot be assumed. This highlights a perfect example of where meta-bias can creep into research, as the meta-analysis results that the researcher concludes with depend on the *choices* they make. Results for the PPCS outcome were not statistically significant so there was no evidence here to indicate that the treatment is useful. However, this is at least in part due to combining different

tests that are perhaps not comparable enough. It would be better to pool together results that used a single test. Unfortunately, these results are not robust enough to conclude whether HBOT is a useful treatment for the PPCS outcome or not.

In this study standardised effect sizes were used to pool together the results. However standardised results are heavily affected by the amount of variability in the study sample. Therefore raw mean difference results and standardised mean difference results are not comparable. In small studies d is often over-estimated so the bias correction, J is commonly used (equation 6). This correction wasn't included in the analyses, however it may have improved the results.

It should be noted that the summary treatment effects are only relevant to the same population groups as found in the study. For example, included studies were heavily male dominated. Only Boussi-Gross' study was balanced well for gender. So, the results may not be consistent with results found for females. Similarly, the study participants were typically fairly young (mean ages in the studies ranged from 12 to 44 years), so the treatment may affect older patients differently.

The studies for Memory Index and Executive Function included mostly civilians, although Harch's study also included military personnel. The civilian dominated studies tended to show results in favour of HBOT. The extra two studies that were also included for the PPCS outcome both concerned military personnel. As a generalisation, these military studies showed a less favourable outcome for the use of HBOT for PPCS. This could be due to the different study designs used, or there may be a *real* difference between these population groups: The military personnel are more likely to have received multiple concussive traumas and the type of concussion was more likely to be from an explosion than a direct blow to the head. Therefore, the dynamics of how the concussion was received may make a difference. Equally, military personnel are more likely to be also dealing with post-traumatic

stress disorder (PTSD), which can cause similar psychological difficulties that may overlap or complicate their PPCS.

It was the two military studies of Cifu's and Miller's that were found to have the lowest risk of bias, so perhaps these two studies simply show smaller effect sizes as a result of being drawn from more robust studies, which commonly happens. However, there is concern regarding Cifu's study. Cifu fails to report some interesting study information which is included in another paper that he co-authored: Walker et al.[29] reported results for the same trial. However, Walker's paper describes how 10 out of the 60 included participants were excluded from his analyses for failing a test of memory malingering (ToMM). Cifu retained these participants in his analysis, which will likely have biased his results. This demonstrates how vigilance is required when looking for potential areas of bias, other than those drawn from Cochrane's risk of bias tool. Cochrane's tool provided a small concern over bias in three out of the five studies. However, none of the funnel plots showed any cause for concern regarding publication bias due to the file drawer effect. In general, bias wasn't thought to be a significant issue in these meta-analyses, although it is unknown how different Cifu's results would be if he had removed those 10 participants from his study.

There were two cross-over studies included in the analysis. These study designs are often used to study chronic diseases as symptoms are usually fairly stable over time, so a change can typically be attributed to a treatment effect. However, one could argue that a cross-over design perhaps isn't such a good choice for PPCS as symptoms can change over time, often getting worse[12]. Whilst this shouldn't be a major issue here, given only the first part of the study is used, it may perhaps question whether this study design was the best choice for PPCS research.

Adverse effects should also be considered for HBOT treatment. It was mentioned previously that Hadanny's study, which included only children, had a very high ad-

verse effect rate of 88%, which is common when treating children. However, Miller's study also reported almost 1 in 5 patients had adverse effects caused by the hyperbaric conditions. There were no serious adverse events though and any adverse events were temporary. These difficulties are likely to be minimised by reducing the rate at which the chamber pressure is changed during treatment.

Unfortunately, there were a limited number of studies that could be pooled together in this project, which limited the type of analysis that could be done. With more studies, it would be interesting to perform subgroup analyses or network analyses to help determine which treatment dose is most effective. Different air pressures and oxygen levels have been tested in different studies with differing lengths of treatment course, but it is still unknown which treatment protocols are most effective and least hazardous for the patients. Similarly, with more data it would be interesting to perform a meta-regression to determine which covariates determine the effectiveness of the treatment.

There is debate over whether HBOT simply provides a placebo treatment effect, or whether treatment effects are not found to be clinically significant because the sham treatment that HBOT is often compared to, works as a treatment itself, due to hyperbaric pressure. It may be coincidental, but for all three of the outcomes, results showed that the studies compared with 'no treatment' (Harch and Boussi-Gross) had the largest treatment effect and Cifu's study, whose control treatment was hyperbaric at the highest used pressure level of 2 ATA gave the smallest effect size. This appears to back the argument that sham treatment is itself a treatment.

Research into the mechanism of HBOT needs further study. Studies often focus on oxygen availability as being the primary mechanism for HBOT's treatment effect, however, the effect of different hyperbaric conditions is under-studied and crucial in further research as this also provides a potential causality of the treatment effect.

Measures of PPCS are often subjective and test answers can be unconsciously altered by the patient. This subjectivity can also be influenced by nocebo effects, where patients get sub-optimal pre-treatment scores because they believe that their symptoms are worse than they are, which in turn can produce an exaggerated treatment effect. Alternatively, their results can be influenced in the opposite way by having a low expectation of treatment success. Similarly, post treatment scores can be inflated by a placebo effect where the patient strongly believes that the treatment will work. The placebo effect could explain why Harch's study, where the patients were not blinded to their treatment, may have found such a high treatment effect. Participant expectations therefore need to be managed carefully or at least assessed and categorised as covariates in the analysis. Another improvement in these studies would be to use more objective measures of treatment effect such as using imaging techniques to assess physical changes in the brain, as they are less likely to be disputed.

To resolve the dispute over whether or not HBOT does improve symptoms for PPCS patients, high quality primary research must be completed with improved study designs before doing further meta-analyses. In future research, agreed outcomes and agreed descriptions of what constitutes PPCS should be used. Patient expectations also need to be managed better and suitable ways of masking the pressurised environment need to be addressed. After all, meta-analyses can only be as good as the data they include.

Recently HBOT has been used for an increasing number of medical conditions with optimistic results. Its use for PTSD and depression highlight its potential for helping mental health conditions too, which are notoriously difficult to manage. Improving study designs for trials using HBOT will therefore help not only studies for PPCS but other medical issues too.

References

[1] Mathias Harrer et al. *Doing Meta-Analysis With R: A Hands-On Guide*. 1st. Boca Raton, FL and London: Chapman & Hall/CRC Press, 2021. ISBN: 9780367610074.

[2] Michael Borenstein et al. *Introduction to Meta-Analysis*. eng. 1. Aufl. Somerset: Wiley, 2011, pp. xxix–xxix. ISBN: 9780470057247.

[3] Angela Boland, Gemma Cherry, and Rumona Dickson. *Doing a systematic review: A student's guide*. Sage, 2017.

[4] Martin Bland. *An Introduction to Medical Statistics*. Oxford University Press, 2015.

[5] Robin Featherstone and Anne Littlewood. *Improving Search Strategies through Peer Review: Background, Methods & Pilot Evaluation*. 2019. URL: https://community.cochrane.org/sites/default/files/uploads/CIS-Webinar_Peer-Review.pdf.

[6] *Risk of Bias 2 Cochrane Review Group Starter Pack*. 2022. URL: https://methods.cochrane.org/file/rob-2-starter-pack-cochrane-reviewspdf.

[7] A Chaimani et al. *Chapter 11: Undertaking network meta-analyses in Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. 2022.

[8] JPT Higgins, T Li, and JJ Deeks (editors). *Chapter 6: Choosing effect measures and computing estimates of effect in Cochrane Handbook for Systematic Reviews of Interventions version 6.3. 2022*.

[9] URL: https://www.headway.org.uk/about-brain-injury/individuals/types-of-brain-injury/mild-head-injury-and-concussion/.

[10] Marjolein van der Vlegel et al. “Prevalence of Post-Concussion-Like Symptoms in the General Injury Population and the Association with Health-Related Quality of Life, Health Care Use, and Return to Work”. *J Clin Med* 10.4 (2021), p. 806.

[11] 2015. URL: https://www.brainfacts.org/ask-an-expert/what-happens-in-the-brain-during-and-after-a-concussion.

[12] 2023. URL: https://www.cognitivefxusa.com/blog/post-concussion-syndrome-and-post-concussion-symptoms-pcs.

[13] URL: https://qbi.uq.edu.au/blog/2018/05/what-does-concussion-do-brain.

[14] Philip J. A. Dean, Darragh O’Neill, and Annette Sterr. “Post-concussion syndrome: Prevalence after mild traumatic brain injury in comparison with a sample without head injury”. *Brain Injury* 26.1 (2012), pp. 14–26. DOI: 10.3109/02699052.2011.635354.

[15] Daphne C. Voormolen et al. “Prevalence of post-concussion-like symptoms in the general population in Italy, The Netherlands and the United Kingdom”. *Brain Injury* 33.8 (2019), pp. 1078–1086. DOI: 10.1080/02699052.2019.1607557.

[16] Laura Smith-Seemiller et al. “Presence of post-concussion syndrome symptoms in patients with chronic pain vs mild traumatic brain injury”. *Brain Injury* 17.3 (2003), pp. 199–206. DOI: 10.1080/0269905021000030823.

[17] Samuel A. McLean et al. “Health status, not head injury, predicts concussion symptoms after minor injury”. *The American Journal of Emergency Medicine*

27.2 (2009), pp. 182–190. ISSN: 0735-6757. DOI: 10.1016/j.ajem.2008.01.054.

[18] Cameron Marshall. URL: https://concussiondoc.io/.

[19] Christian Smolle et al. "The History and Development of Hyperbaric Oxygenation (HBO) in Thermal Burn Injury". *Medicina* 57.1 (2021). ISSN: 1648-9144. DOI: 10.3390/medicina57010049.

[20] URL: https://wearecompass.org.uk/wp-content/uploads/2022/02/Oxygen-Therapy-Client-Handbook.pdf.

[21] URL: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/hyperbaric-oxygen-therapy.

[22] Amir Hadanny et al. "Hyperbaric oxygen therapy in children with post-concussion syndrome improves cognitive and behavioral function: a randomized controlled trial". *Scientific Reports* 12.1 (2022), p. 15233. DOI: 10.1038/s41598-022-19395-y.

[23] David X. Cifu et al. "The Effect of Hyperbaric Oxygen on Persistent Postconcussion Symptoms". *The Journal of Head Trauma Rehabilitation* 29.1 (2014).

[24] R. Scott Miller, Lindell K. Weaver, and Nazanin Bahraini et al. "Effects of Hyperbaric Oxygen on Symptoms and Quality of Life Among Service Members With Persistent Postconcussion SymptomsA Randomized Clinical Trial". *JAMA Internal Medicine* (2015).

[25] R Boussi-Gross et al. "Hyperbaric Oxygen Therapy Can Improve Post Concussion Syndrome Years after Mild Traumatic Brain Injury - Randomized Prospective Trial". *PLoS ONE 8(11): e79995* (2013).

[26] Paul G Harch et al. "Hyperbaric oxygen therapy for mild traumatic brain injury persistent postconcussion syndrome: a randomized controlled trial". *Medical Gas Research* (2020).

[27] Mathias Harrer et al. *dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'*. R package version 0.0.9000. 2019. URL: http://dmetar.protectlab.org/.

[28] Sterne JAC et al. "RoB 2: a revised tool for assessing risk of bias in randomised trials". *BMJ 2019; 366: l4898* (2019).

[29] William C. Walker et al. "Randomized, Sham-Controlled, Feasibility Trial of Hyperbaric Oxygen for Service Members With Postconcussion Syndrome: Cognitive and Psychomotor Outcomes 1 Week Postintervention". *Neurorehabilitation and Neural Repair* 28.5 (2014), pp. 420–432. DOI: 10.1177/1545968313516869.

A Study information

Table 9: Study information

First Author	published	Population	Study design
Boussi-Gross	15/11/13	$\geq$ 18 civilian	sgl blind, random cross-over
Harch	13/03/20	18-65 civilian & military	sgl blind, random, stratified cross-over
Hadanny	23/09/22	8-15 civilian	dbl blind, random, sham controlled
Cifu	01/02/14	active duty military	blinded, 3-arm, random, sham controlled
Miller	17/11/14	$\geq$ 18, hospitalised military	*dbl blind, random, sham controlled

First Author	Primary outcome	Intervention	Control
Boussi-Gross	Neurotrax	1hr HBOT 100% O_2, 1.5ATA	Std treatment
Harch	NSI & working memory	1hr HBOT 100% O_2, 1.5ATA	Std treatment
Hadanny	Neurotrax	1hr HBOT 100% O_2, 1.5ATA	21% O_2, 1.03ATA
Cifu	RPQ	1hr HBOT 75% O_2, 2.0ATA	10.5% O_2, 2.0ATA
Miller	RPQ	1hr HBOT 100% O_2, 1.5ATA	Std treatment

First Author	sessions	Weeks	Male:Fem	Mean Age	Retest time	PPCS test
Boussi-Gross	40 daily, 5/wk	8	43:57	44	1-3 wks after end	EQ-VAS
Harch	40 daily, 5/wk	8	-	-	within 2 months	NSI
Hadanny	60 daily, 5/wk	12	80:20	11.6	1-3 wks after end	HBI
Cifu	40 daily in 10wks	10	100:0	23.2	within 1 wk after end	RPQ
Miller	40 daily in 10wks	10	96:4	-	within 1 month after end	NSI

B Meta-analysis outputs

Table 10: Meta-analysis - Executive Function

```
                      SMD       95%-CI      %W(common) %W(random)
Rahav Boussi-Gross 0.6600 [ 0.2896; 1.0304]       64.6       64.6
Paul G Harch       0.3400 [-0.2205; 0.9005]       28.2       28.2
Amir Hadanny       0.1300 [-0.9813; 1.2413]        7.2        7.2

Number of studies combined: k = 3

                              SMD       95%-CI       z|t p-value
Common effect model        0.5317 [ 0.2339; 0.8294] 3.50  0.0005
Random effects model (HK) 0.5317 [-0.0175; 1.0809] 4.17  0.0531
Prediction interval               [-1.3986; 2.4619]

Quantifying heterogeneity:
 tau^2 = 0 [0.0000; 2.6581]; tau = 0 [0.0000; 1.6304]
 I^2 = 0.0% [0.0%; 89.6%]; H = 1.00 [1.00; 3.10]

Test of heterogeneity:
Q      d.f.   p-value
1.41    2     0.4936

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp (HK) adjustment for random effects model (df = 2)
- Prediction interval based on t-distribution (df = 1)
```

Table 11: Meta-analysis - PPCS

```
                        SMD            95%-CI %W(random)
Rahav Boussi-Gross   1.0700 [ 0.5055; 1.6345]       20.0
*Paul G Harch        2.5900 [ 1.8374; 3.3426]       18.9
Amir Hadanny         0.1700 [ 0.0269; 0.3131]       21.5
Cifu, David X. MD   -0.1500 [-0.7811; 0.4811]       19.6
*R. Scott Miller     0.3000 [-0.2743; 0.8743]       20.0

Number of studies combined: k = 5

                              SMD            95%-CI    t   p-value
Random effects model (HK) 0.7710 [-0.5695; 2.1114] 1.60  0.1855
Prediction interval              [-2.8145; 4.3564]

Quantifying heterogeneity:
 tau^2 = 1.0439 [0.3118; 9.8396]; tau = 1.0217 [0.5584; 3.1368]
 I^2 = 91.6% [83.4%; 95.8%]; H = 3.46 [2.46; 4.86]

Test of heterogeneity:
 Q      d.f.   p-value
 47.77   4    < 0.0001

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp (HK) adjustment for random effects model (df = 4)
- Prediction interval based on t-distribution (df = 3)
```

Table 12: Meta-analysis - PPCS (alternative calculation of Boussi-Gross' effect size)

```
Review:     Meta analysis - PPCS

                        SMD            95%-CI %W(random)
Rahav Boussi-Gross   0.6600 [ 0.1171; 1.2029]       20.1
*Paul G Harch        2.5900 [ 1.8374; 3.3426]       18.9
Amir Hadanny         0.1700 [ 0.0269; 0.3131]       21.5
Cifu, David X. MD   -0.1500 [-0.7811; 0.4811]       19.6
*R. Scott Miller     0.3000 [-0.2743; 0.8743]       19.9

Number of studies combined: k = 5

                              SMD            95%-CI    t p-value
Random effects model (HK) 0.6881 [-0.6347; 2.0109] 1.44  0.2222
Prediction interval               [-2.8340; 4.2102]

Quantifying heterogeneity:
 tau^2 = 1.0071 [0.2971; 9.6476]; tau = 1.0035 [0.5451; 3.1061]
 I^2 = 90.5% [80.7%; 95.3%]; H = 3.24 [2.28; 4.62]

Test of heterogeneity:
     Q d.f.  p-value
 42.06    4 < 0.0001

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp (HK) adjustment for random effects model (df = 4)
- Prediction interval based on t-distribution (df = 3)
```

Table 13: Meta-analysis - PPCS (Excl. Boussi-Gross study)

```
Review:     Meta analysis - PPCS

                       SMD            95%-CI %W(random)
*Paul G Harch       2.5900 [ 1.8374; 3.3426]       24.0
Amir Hadanny        0.1700 [ 0.0269; 0.3131]       26.4
Cifu, David X. MD -0.1500 [-0.7811; 0.4811]       24.7
*R. Scott Miller    0.3000 [-0.2743; 0.8743]       25.0

Number of studies combined: k = 4

                               SMD            95%-CI    t p-value
Random effects model (HK) 0.7037 [-1.2656; 2.6729] 1.14  0.3381
Prediction interval              [-5.0303; 6.4377]

Quantifying heterogeneity:
 tau^2 = 1.4041 [0.3784; 21.8018]; tau = 1.1849 [0.6152; 4.6692]
 I^2 = 92.5% [84.0%; 96.5%]; H = 3.65 [2.50; 5.32]

Test of heterogeneity:
     Q d.f.  p-value
 39.88    3 < 0.0001

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp (HK) adjustment for random effects model (df = 3)
- Prediction interval based on t-distribution (df = 2)
```

Table 14: Meta-analysis - PPCS (Excl. Harch's study)

```
Review:     Meta analysis - PPCS

                        SMD            95%-CI %W(random)
Rahav Boussi-Gross   0.6600 [ 0.1171; 1.2029]        9.5
Amir Hadanny         0.1700 [ 0.0269; 0.3131]       74.7
Cifu, David X. MD   -0.1500 [-0.7811; 0.4811]        7.2
*R. Scott Miller     0.3000 [-0.2743; 0.8743]        8.6

Number of studies combined: k = 4

                               SMD            95%-CI    t p-value
Random effects model (HK) 0.2049 [-0.1157; 0.5254] 2.03  0.1348
Prediction interval              [-0.2836; 0.6934]

Quantifying heterogeneity:
 tau^2 = 0.0051 [0.0000; 1.4733]; tau = 0.0714 [0.0000; 1.2138]
 I^2 = 28.7% [0.0%; 73.7%]; H = 1.18 [1.00; 1.95]

Test of heterogeneity:
    Q d.f. p-value
 4.21    3  0.2397

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp (HK) adjustment for random effects model (df = 3)
- Prediction interval based on t-distribution (df = 2)
```

Table 15: Meta-analysis - PPCS (Excl. Harch's and Hadanny's study)

```
Review:     Meta analysis - PPCS

                        SMD            95%-CI %W(random)
Rahav Boussi-Gross  0.6600 [ 0.1171; 1.2029]       35.9
Cifu, David X. MD  -0.1500 [-0.7811; 0.4811]       30.4
R. Scott Miller     0.3000 [-0.2743; 0.8743]       33.8

Number of studies combined: k = 3

                              SMD            95%-CI    t p-value
Random effects model (HK) 0.2925 [-0.7068; 1.2918] 1.26  0.3349
Prediction interval              [-4.1995; 4.7845]

Quantifying heterogeneity:
 tau^2 = 0.0717 [0.0000; 6.4141]; tau = 0.2678 [0.0000; 2.5326]
 I^2 = 45.0% [0.0%; 83.7%]; H = 1.35 [1.00; 2.48]

Test of heterogeneity:
    Q d.f. p-value
 3.64    2  0.1621

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp (HK) adjustment for random effects model (df = 2)
- Prediction interval based on t-distribution (df = 1)
```

Table 16: Meta-analysis - PPCS (Excl. Hadanny's study)

```
Review:     Meta analysis - PPCS

                        SMD            95%-CI %W(random)
Rahav Boussi-Gross   0.6600 [ 0.1171; 1.2029]       25.5
*Paul G Harch        2.5900 [ 1.8374; 3.3426]       24.2
Cifu, David X. MD   -0.1500 [-0.7811; 0.4811]       25.0
*R. Scott Miller     0.3000 [-0.2743; 0.8743]       25.3

Number of studies combined: k = 4

                               SMD            95%-CI    t p-value
Random effects model (HK) 0.8341 [-1.0655; 2.7337] 1.40  0.2567
Prediction interval              [-4.6789; 6.3471]

Quantifying heterogeneity:
 tau^2 = 1.2928 [0.3346; 20.1005]; tau = 1.1370 [0.5784; 4.4834]
 I^2 = 90.9% [79.9%; 95.9%]; H = 3.32 [2.23; 4.95]

Test of heterogeneity:
     Q d.f.  p-value
 33.07    3 < 0.0001

Details on meta-analytical method:
- Inverse variance method
- Restricted maximum-likelihood estimator for tau^2
- Q-Profile method for confidence interval of tau^2 and tau
- Hartung-Knapp (HK) adjustment for random effects model (df = 3)
- Prediction interval based on t-distribution (df = 2)
```

C Code

C.1 Effect size calculations

```
library(dmetar)
library(esc)

#Calculate d and SE_d for effect sizes given in D
D=-23.9
n_A=23
n_B=27
s_pooled=9.22

SE_D=s_pooled*(sqrt((1/n_A)+(1/n_B))); SE_D
d=D/s_pooled; d
SE_d=sqrt(((n_A+n_B)/(n_A*n_B))+(d^2/(2*(n_A+n_B)))); SE_d

##########################################################
#Calculate SE's from p-value
#se.from.p(effect size, p=p-value, N=total study sample)

se.from.p(-0.17, p=0.02, N=24, effect.size.type="difference")

##########################################################
#Calculate mean difference from control effect sizes and treatment effect sizes
x_A = 1.09   #treatment group mean effect
x_B = 0.19  #control group mean effect
D=x_A-x_B  #raw mean difference
n_A=18    #treatment sample
n_B=21     #no treatment sample (control) - choosing this rather than sham treatment
```

```
#t-statistic from p-value
#Need the inverse CDF function of 1-p
p_A=0.21/2 #divide by 2 for 2-tailed test
t_A=qt(1-p_A, n_A-1); t_A #function to give t-statistic

p_B=0.91/2 #divide by 2 for 2-tailed test
t_B=qt(1-p_B, n_B-1); t_B #function to give t-statistic

#standard deviation of differences for group A and B
s_A=abs((x_A*sqrt(n_A))/t_A); s_A
s_B=abs((x_B*sqrt(n_B)))/t_B; s_B

################################################
#Standard deviations from Confidence intervals
x_A = 3.7   #treatment group mean effect
x_B = -1.1  #control group mean effect
D=x_A-x_B  #raw mean difference
n_A=24
n_B=23
t_A=qt(0.975, n_A-1) # For 2-tailed 95% CI
t_B=qt(0.975, n_B-1)

upper_CI_A=11.2  #upper 95% CI of treatment group
upper_CI_B=5.2   #upper 95% CI of control group
lower_CI_A=-3.7
lower_CI_B=-7.3

s_A=sqrt(n_A)*(upper_CI_A - lower_CI_A)/3.92; s_A
s_B=sqrt(n_B)*(upper_CI_B - lower_CI_B)/3.92; s_B
# OR
s_A=sqrt(n_A)*(upper_CI_A - lower_CI_A)/(2*t_A); s_A
```

```
s_B=sqrt(n_B)*(upper_CI_B - lower_CI_B)/(2*t_B); s_B

# OR if given standard deviations
s_A=
s_B=
################################################
#Finding s_pooled, d and SE_d

s_pooled=sqrt(((n_A-1)*s_A^2+(n_B-1)*s_B^2)/((n_A-1)+(n_B-1))); s_pooled
SE_D=s_pooled*(sqrt((1/n_A)+(1/n_B))); SE_D
d=D/s_pooled; d
SE_d=sqrt(((n_A+n_B)/(n_A*n_B))+(d^2/(2*(n_A+n_B)))); SE_d
```

C.2 Meta-analysis

```
######################## meta analysis ######################
library(tidyverse)
library(meta)
library(dmetar)
library(readxl)

meta_data <- read_excel("meta_data.xlsx")
head(meta_data)

#pool memory effect sizes
m.gen.mem <- metagen(TE=memory,
                 seTE=memory_SE,
                 studlab=First_Author,
                 data=meta_data[1:3,],
                 sm="SMD",
                 comb.fixed=FALSE,
                 comb.random=TRUE,
                 method.tau="REML",
                 hakn=F,        #knapp hartung adjustment
                 prediction=TRUE,
                 title="Meta analysis - Memory")
summary(m.gen.mem)

#pool Executive Function effect sizes
m.gen.exec <- metagen(TE=executive_func,
                      seTE=executive_func_SE,
                      studlab=First_Author,
                      data=meta_data[1:3,],
                      sm="SMD",
```

```
                      comb.fixed=F,
                      comb.random=TRUE,
                      method.tau="REML",
                      hakn=F,       #knapp hartung adjustment
                      prediction=TRUE,
                      title="Meta analysis - Executive Function")
summary(m.gen.exec)

#pool PPCS effect sizes
m.gen.ppcs <- metagen(TE=ppcs,
                      seTE=ppcs_SE,
                      studlab=First_Author,
                      data=meta_data,
                      sm="SMD",
                      comb.fixed=F,
                      comb.random=TRUE,
                      method.tau="REML",
                      hakn=TRUE,        #knapp hartung adjustment
                      prediction=TRUE,
                      title="Meta analysis - PPCS")
summary(m.gen.ppcs)

######################## forest plots ######################
forest.meta(m.gen.mem,
            sortvar=TE,
            print.tau2=FALSE,
            print.I2.ci = TRUE,
            comb.fixed=FALSE,
            label.right = "Favours treatment",
            xlim=c(-2,3),
            col.diamond = "blue",
```

```
            col.square = "light blue",
            col.predict = "blue",
            leftlabs = c("Author", "d", "SE"))

forest.meta(m.gen.exec,
            sortvar=TE,
            print.tau2=FALSE,
            print.I2.ci = TRUE,
            comb.fixed=FALSE,
            label.right = "Favours treatment",
            xlim=c(-2,3),
            col.diamond = "blue",
            col.square = "light blue",
            col.predict = "blue",
            leftlabs = c("Author", "d", "SE"))

forest.meta(m.gen.ppcs,
            sortvar=TE,
            print.tau2=FALSE,
            print.I2 = TRUE,
            comb.fixed=FALSE,
            label.right = "Favours treatment",
            xlim=c(-3,5),
            col.diamond = "blue",
            col.square = "light blue",
            col.predict = "blue",
            leftlabs = c("Author", "d", "SE"))

#Forest plots with RevMan5 layout
forest.meta(m.gen.mem, layout="RevMan5")
forest.meta(m.gen.exec, layout="RevMan5")
```

```
forest.meta(m.gen.ppcs, layout="RevMan5")

########################## Funnel Plots ###################################
col.contour=c("indianred", "orange", "turquoise")

funnel.meta(m.gen.mem,
            xlim=c(0,1.5),
            contour=c(0.9, 0.95, 0.99),
            col.contour=col.contour,
            studlab=F)
legend(x=1.2, y=0.02,
       legend= c("p<0.1", "p<0.05", "p<0.01"),
       fill=col.contour)
title("Funnel Plot (Memory effect)")

funnel.meta(m.gen.exec,
            xlim=c(0,1.5),
            contour=c(0.9, 0.95, 0.99),
            col.contour=col.contour,
            studlab=F)
legend(x=1.2, y=0.02,
       legend= c("p<0.1", "p<0.05", "p<0.01"),
       fill=col.contour)
title("Funnel Plot (Executive Function effect)")

funnel.meta(m.gen.ppcs,
            xlim=c(-1,3),
            contour=c(0.9, 0.95, 0.99),
            col.contour=col.contour,
            studlab=F)
legend(x=2, y=0.02,
```

```
       legend= c("p<0.1", "p<0.05", "p<0.01"),
       fill=col.contour)
title("Funnel Plot (PPCS effect)")

######################### Risk of bias plots ###################################
library(robvis)
str(data_rob2)
bias <- read_excel("Risk_of_Bias.xlsx")
rob_summary(bias, "ROB2")
rob_traffic_light(bias, "ROB2")

########## example plots ###########
bias_example <- read_excel("Risk_of_Bias_example.xlsx")
rob_summary(bias_example, "ROB2")
rob_traffic_light(bias_example, "ROB2")

###########     Sensitivity Studies.      ##################################
######### Boussi-Gross measured from treat HBOT and control pre-HBOT ###########
meta_dataV <- read_excel("meta_data_VAS.xlsx")

#pool PPCS effect sizes
m.gen.ppcsV <- metagen(TE=ppcs,
                       seTE=ppcs_SE,
                       studlab=First_Author,
                       data=meta_dataV,
                       sm="SMD",
                       comb.fixed=F,
                       comb.random=TRUE,
                       method.tau="REML",
                       hakn=TRUE,        #knapp hartung adjustment
                       prediction=TRUE,
```

```
                      title="Meta analysis - PPCS")
summary(m.gen.ppcsV)

forest.meta(m.gen.ppcsV,
            sortvar=TE,
            print.tau2=FALSE,
            print.I2 = TRUE,
            comb.fixed=FALSE,
            label.right = "Favours treatment",
            xlim=c(-3,5),
            col.diamond = "blue",
            col.square = "light blue",
            col.predict = "blue",
            leftlabs = c("Author", "d", "SE"))

################ Data with no Boussi.  ####################
meta_dataNB <- read_excel("meta_data_NOBoussi.xlsx")

#pool PPCS effect sizes
m.gen.ppcsNB <- metagen(TE=ppcs,
                        seTE=ppcs_SE,
                        studlab=First_Author,
                        data=meta_dataNB,
                        sm="SMD",
                        comb.fixed=F,
                        comb.random=TRUE,
                        method.tau="REML",
                        hakn=TRUE,        #knapp hartung adjustment
                        prediction=TRUE,
                        title="Meta analysis - PPCS")
summary(m.gen.ppcsNB)
```

```
forest.meta(m.gen.ppcsNB,
            sortvar=TE,
            print.tau2=FALSE,
            print.I2 = TRUE,
            comb.fixed=FALSE,
            label.right = "Favours treatment",
            xlim=c(-3,5),
            col.diamond = "blue",
            col.square = "light blue",
            col.predict = "blue",
            leftlabs = c("Author", "d", "SE"))

######################### New B-G calc and NO Harch ###################################
meta_dataNH <- read_excel("meta_data_VAS_NoHar.xlsx")

#pool PPCS effect sizes
m.gen.ppcsNH <- metagen(TE=ppcs,
                        seTE=ppcs_SE,
                        studlab=First_Author,
                        data=meta_dataNH,
                        sm="SMD",
                        comb.fixed=F,
                        comb.random=TRUE,
                        method.tau="REML",
                        hakn=TRUE,        #knapp hartung adjustment
                        prediction=TRUE,
                        title="Meta analysis - PPCS")
summary(m.gen.ppcsNH)

forest.meta(m.gen.ppcsNH,
```

```
            sortvar=TE,
            print.tau2=FALSE,
            print.I2 = TRUE,
            comb.fixed=FALSE,
            label.right = "Favours treatment",
            xlim=c(-1,1.5),
            col.diamond = "blue",
            col.square = "light blue",
            col.predict = "blue",
            leftlabs = c("Author", "d", "SE"))

################## No Harch or Hadanny ################
meta_dataNHH <- read_excel("meta_data_VAS_NoHarNoHad.xlsx")

#pool PPCS effect sizes
m.gen.ppcsNHH <- metagen(TE=ppcs,
                         seTE=ppcs_SE,
                         studlab=First_Author,
                         data=meta_dataNHH,
                         sm="SMD",
                         comb.fixed=F,
                         comb.random=TRUE,
                         method.tau="REML",
                         hakn=TRUE,        #knapp hartung adjustment
                         prediction=TRUE,
                         title="Meta analysis - PPCS")
summary(m.gen.ppcsNHH)

forest.meta(m.gen.ppcsNHH,
            sortvar=TE,
            print.tau2=FALSE,
```

```
            print.I2 = TRUE,
            comb.fixed=FALSE,
            label.right = "Favours treatment",
            xlim=c(-1,1.5),
            col.diamond = "blue",
            col.square = "light blue",
            col.predict = "blue",
            leftlabs = c("Author", "d", "SE"))

################## No Hadanny ################
meta_dataNHad <- read_excel("meta_data_VAS_NoHad.xlsx")

#pool PPCS effect sizes
m.gen.ppcsNHad <- metagen(TE=ppcs,
                          seTE=ppcs_SE,
                          studlab=First_Author,
                          data=meta_dataNHad,
                          sm="SMD",
                          comb.fixed=F,
                          comb.random=TRUE,
                          method.tau="REML",
                          hakn=TRUE,        #knapp hartung adjustment
                          prediction=TRUE,
                          title="Meta analysis - PPCS")
summary(m.gen.ppcsNHad)

forest.meta(m.gen.ppcsNHad,
            sortvar=TE,
            print.tau2=FALSE,
            print.I2 = TRUE,
            comb.fixed=FALSE,
```

```
            label.right = "Favours treatment",
            xlim=c(-1,4),
            col.diamond = "blue",
            col.square = "light blue",
            col.predict = "blue",
            leftlabs = c("Author", "d", "SE"))

################## NNTs. ##########################
NNT(d=0.53)

################## Influence diagnostics. ######################
#Look for any outliers
find.outliers(m.gen.ppcs)

# Influence diagnostics
m.gen.ppcs.inf <- InfluenceAnalysis(m.gen.ppcs, random=TRUE, text.scale = 2)
plot(m.gen.ppcs.inf)

plot(m.gen.ppcs.inf, "baujat")
plot(m.gen.ppcs.inf, "influence")
plot(m.gen.ppcs.inf, "es")
plot(m.gen.ppcs.inf, "i2")
```