An application of a bivariate random effects meta-analysis in a cost-effectiveness analysis of treatment for sleep apnoea.

Susan Griffin and Neil Hawkins, Centre for Health Economics, University of York.

Aims: To demonstrate the use of a bivariate random effects meta-analysis (BRMA) to synthesise data from a systematic review of continuous positive airways pressure (CPAP) for the treatment of obstructive sleep apnoea syndrome.

Methods: The trials reported a range of relevant outcome data. Two outcomes were of interest for the cost-effectiveness analysis: mean difference in Epworth Sleepiness Scale (ESS) score and mean difference in blood pressure (BP). The available evidence was synthesised using BRMA, incorporating information on within-study correlation and the observed between-study correlation. This paper compares BRMA to separate univariate random effects meta-analyses (URMAs).

Results: ESS was reported in the majority of studies and the pooled treatment effect of CPAP differed little between analyses. BP was infrequently reported and the pooled treatment effect of CPAP varied between the URMA and BRMA according to the ability of the between-study correlation to inform the missing data where trials reported only ESS. An informative prior for the within-study correlation had little impact on the results.

Conclusions: Systematic reviews may focus on multiple outcomes of interest. Multivariate meta-analysis takes advantage of information provided by the between-study correlation. Where not all trials report complete sets of outcome measures, a multivariate meta-analysis can be used effectively to impute missing outcomes on the basis of observed characteristics. Conducting separate URMAs assumes that outcomes are missing completely at random and that alternative outcome measures are uncorrelated. The difference between the two approaches may be important when the results are used to inform a cost-effectiveness analysis, particularly in terms of characterising decision uncertainty.

1.1 Introduction

Systematic reviews to identify the current available clinical evidence for health technologies often form the basis for an economic evaluation. Where multiple trials provide information on the same outcome measure a meta-analysis may be conducted to synthesise the available evidence. In many cases, more than one clinical outcome will be of interest in determining the health outcomes of each technology under consideration. Frequently in published systematic reviews, multiple outcomes are synthesised individually in a series of univariate meta-analyses. This paper explores the use of multivariate meta-analysis in the context of an evaluation of continuous positive airways pressure (CPAP) for the treatment of sleep apnoea.

Sleep apnoea describes a sleep disorder where repeated collapse of the upper airway during sleep causes a reduction (hypopnoea) and in some cases total obstruction (apnoea) of respiratory airflow. This in turn requires increased respiratory effort, which for the sufferer is accompanied by arousal from normal sleep. When sleep apnoea is accompanied by clinical symptoms such as excessive daytime sleepiness it is known as obstructive sleep apnoea-hypopnoea syndrome (OSAHS). There is a strong association between OSAHS and obesity as fat deposits on the upper body and neck cause the airway to narrow. In addition to daytime sleepiness, the symptoms of OSAHS include snoring, nocturnal choking and morning headaches. Other consequences include hypertension, memory disturbance and impaired concentration for tasks that require vigilance, such as driving.

The main treatment goal in OSAHS is to reduce daytime sleepiness. The range of treatments available for OSAHS include the administration of CPAP during sleep, the use of dental devices to reposition the tongue or mandible and lifestyle modifications such as weight loss, improved sleep positioning and avoidance of alcohol. Surgical and pharmacological interventions exist but are rarely used in practice. Due to the variation in provision of treatment for OSAHS within the UK, the National Institute for Health and Clinical Excellence (NICE) requested an evaluation of the clinical and cost effectiveness of CPAP in comparison to best supportive care, placebo and dental devices. This evaluation forms the basis for a comparison of alternative methods to synthesise the available data on clinical outcomes.

1.2 The systematic review and decision model

A systematic review was conducted to identify randomised controlled trials (RCTs) that compared CPAP with either active comparator or placebo, and to identify existing economic evaluations of CPAP. The methods and results of this review have been described in detail elsewhere (*report forthcoming at http://www.nice.org.uk*). In addition, a decision analytic model was developed to compare the costs and consequences of the relevant alternatives for a

lifetime time horizon. This paper focuses on the methods used to synthesise the available clinical evidence for input into the cost-effectiveness model.

The primary outcome measures included in the review were subjective daytime sleepiness as measured by the Epworth Sleepiness Scale and objective sleepiness measured for example by the Maintenance of Wakefulness Test (MWT) or the Multiple Sleep Latency Test (MSLT). The secondary outcomes measures included blood pressure (BP), cardiovascular events (CVEs), road traffic accidents (RTAs) and occupational accidents, quality of life, anxiety and depression, simulated driving performance, neuropsychological functioning, apnoea/hypopneoa index (AHI), oxygen desaturation index and adverse effects of treatments.

The review identified 38 RCTs comparing CPAP to placebo, of which 23 reported information on ESS, 5 on MWT, 7 on MSLT, 10 on BP. The measures of daytime BP included those taken by ambulatory measurement (7 studies) and during office visits (3 studies). One trial reported utility scores as measured by the EuroQol instrument, but it appeared that the QALY gain from baseline was analysed for each arm separately and not compared between arms, the resulting treatment effect for CPAP being akin to that produced by a before and after cohort study. No RCTs reported information about CVEs, occupational accidents or RTAs.

The primary outcome measure for the economic evaluation was the cost per quality-adjusted life-year (QALY) gained. To estimate the total cost and total QALYs associated with each relevant treatment strategy, a Markov cohort model was developed which estimated quality-adjusted survival as a function of survival, daytime sleepiness, CVEs and RTAs. A set of individual patient data were available that allowed ESS scores to be mapped to preference-based measures of health-related quality of life. Thus the baseline ESS score for the cohort was translated directly into a utility value and the improvement in ESS with treatment was translated into an improvement in utility. The Framingham risk equations were used predict risk of CVEs as a function of BP and the utility decrement for CVEs was obtained from previously published studies. Thus from the range of outcomes identified in the systematic review, mean difference in ESS score and mean difference in systolic BP at follow-up were selected for input into the decision-analytic model.

2.1 Methods

Table 1 describes the relevant data extracted on ESS and BP. As there remained uncertainty as to whether treatment effect based on ambulatory blood pressure were comparable to those based on office measurements a scenario incorporating only ambulatory BP was included.

This allows a comparison of the bivariate and univariate approaches when there are fewer data to inform the estimate of the between study correlation.

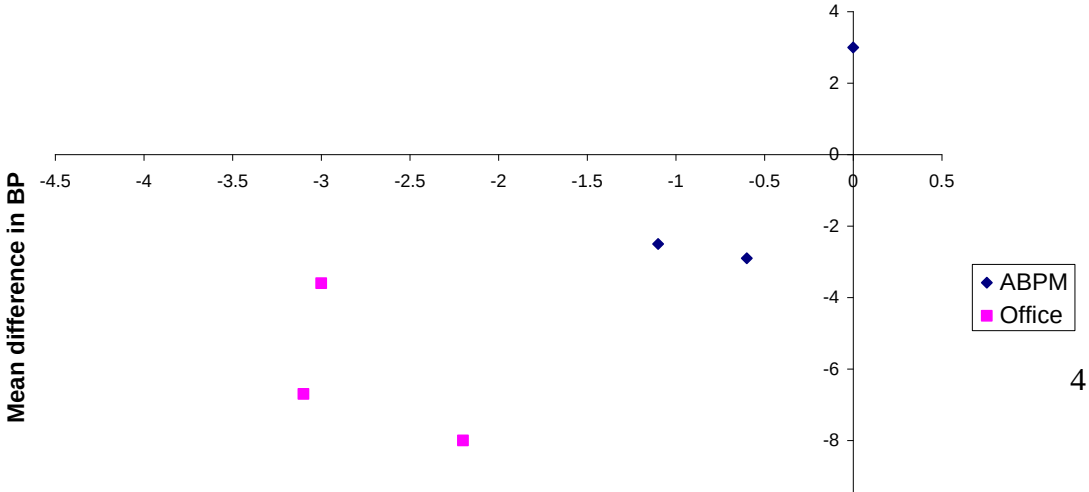
Table 1. Study characteristics and outcomes for inclusion in meta-analysis

Type of study	n	MD ESS	Var(MD ESS)	Type of BP measurement	MD BP	Var(MD BP)
Parallel	54	0	1.37	ABPM	3	13.01
Crossover	35	-1.2	0.17	-	NA	NA
Parallel	105	-5	1.25	-	NA	NA
Crossover	18	0.1	1.40	-	NA	NA
Crossover	23	-6	2.34	-	NA	NA
Parallel	111	-1.09	0.52	-	NA	NA
Crossover	37	-3	1.14	-	NA	NA
Crossover	71	-2.4	0.50	-	NA	NA
Parallel	142	-2.2	0.92	Office	-8	11.69
Crossover	42	-0.6	1.77	ABPM	-2.9	29.16
Parallel	60	-3.8	2.47	ABPM	-10.3	27.88
Crossover	114	-1	0.32	-	NA	NA
Crossover	31	-2.4	0.84	-	NA	NA
Parallel	72	-1	0.56	-	NA	NA
Parallel	56	-1.1	2.00	ABPM	-2.5	8.47
Parallel	101	-3	2.00	Office	-3.6	46.69
Crossover	35	-3.1	0.50	Office	-6.7	3.01
Parallel	42	-4	2.19	-	NA	NA
Parallel	107	-4.8	0.81	-	NA	NA
Parallel	45	-4	4.38	-	NA	NA
Parallel	48	-7.94	1.63	-	NA	NA
Parallel	71	-3	2.47	-	NA	NA
Parallel	118	-4.5	1.03	-	NA	NA
Crossover	13	NA	NA	ABPM	-1	5.62
Crossover	25	NA	NA	ABPM	0	4.45
Parallel	21	NA	NA	ABPM	-1	15.47

MD ESS = Mean difference in Epworth Sleepiness Scale, CPAP vs placebo; MD BP = Mean difference in systolic blood pressure, CPAP vs placebo; Var() = Variance; n = number of trial participants

The treatment effect pairs from the studies that reported both ESS and BP are shown in Figure 1. Only four studies report both outcomes when only ambulatory BP is used, rising to seven when office measures are also incorporated.

Figure 1. Treatment effect pairs from studies reporting both ESS and BP



Individual patient level data from an RCT comparing CPAP to placebo were obtained to inform the within-study correlation between mean difference in ESS and mean difference in BP at follow-up. This dataset was bootstrapped to provide a mean within-study correlation of 0.29 with standard deviation 0.07.

We first set out the framework for a bivariate random effects meta-analysis, which has been reported in Riley *et al.*(Riley et al. 2007a), among others. We then compare the results of bivariate approach to separate univariate analyses, and test the sensitivity of the model to alternative values for the within- and between-study correlation in treatment effects.

2.2 Bivariate random effects meta-analysis

A random effects analysis was performed in both the univariate and bivariate meta-analysis approaches where it was assumed that each study's summary statistics for mean difference in ESS (ess_i) and mean difference in BP (bp_i) represented an estimate of different underlying true values ($essMu_i$, $bpMu_i$), and these underlying true values were assumed to be drawn from a distribution with particular mean (essReMu, bpReMu) and variance ($essReSD^2$, $bpReSD^2$). The framework for the bivariate approach is as follows (ref Riley etc):

$$\begin{pmatrix} ess_i \\ bp_i \end{pmatrix} \sim N\left(\begin{pmatrix} essMu_i \\ bpMu_i \end{pmatrix}, \delta_i \right), \quad \delta_i = \begin{pmatrix} essSD_i^2 & \rho W_i.essSD_i.bpSD_i \\ \rho W_i.essSD_i.bpSD_i & bpSD_i^2 \end{pmatrix}$$

$$\begin{pmatrix} essMu_i \\ bpMu_i \end{pmatrix} \sim N\left(\begin{pmatrix} essReMu \\ bpReMu \end{pmatrix}, \Omega_i \right), \quad \Omega_i = \begin{pmatrix} essReSD^2 & \rho B.essReSD.bpReSD \\ \rho B.essReSD.bpReSD & bpReSD^2 \end{pmatrix}$$

Where $\rho W_i.essSD_i.bpSD_i$ is the within-study covariance and $\rho B.essReSD.bpReSD$ is the between-study covariance. None of the trials reported within-study covariance between mean difference in ESS and mean difference in BP, and so a set of patient-level data containing

both outcomes was obtained from which an informative prior could be specified. In addition, due to the small number of studies reporting both outcomes when incorporating only ambulatory measurements of systolic BP, an informative prior was specified for the variance ($essSD_i^2$, $bpSD_i^2$) to be used where studies did not report one of the outcomes of interest. This prior was specified by multiplying the variance by the sample size, with a crude adjustment for the study design (i.e. doubling the sample size for cross-over studies), so that the prior variance could be informed by the known sample sizes. A product-normal specification was used to characterise the model, and so it was necessary to constrain the estimates for the correlation parameters. The prior for the within-study correlation was specified with a normal distribution truncated between -1 and 1. An uninformative prior for the between-study correlation was specified using a uniform distribution from -1 to 1. The WinBUGS code for the model can be found here (*report forthcoming at http://www.nice.org.uk*).

The univariate meta-analysis models the treatment effects on ESS and BP separately, and is equivalent to assuming that the between-study and within-study correlation between treatment effects on ESS and BP are zero. In both models a burn-in of 50,000 iterations were discarded and 10,000 iterations were used to estimate the pooled treatment effects.

3. Results

Table 2 shows the pooled estimates from the alternative models. The pooled mean difference in ESS is relatively similar across all of the models. The mean difference in BP varies more widely between the univariate and bivariate specifications, and the bivariate approach is associated with a slightly larger standard deviation around the pooled estimate. With ambulatory measurements alone and with both ambulatory and office measurements, the effect of incorporating the between-study correlation is to increase the size of the pooled treatment effect on BP.

Table 2. Pooled estimates of treatment effect on ESS and BP, CPAP vs placebo

	Mean difference in ESS mean (SD)	Mean difference in BP mean (SD)
Ambulatory BP only		
URMA	-2.69 (0.44)	-1.19 (1.57)
BRMA	-2.65 (0.43)	-1.64 (1.72)
Ambulatory + office BP		
URMA	-2.69 (0.43)	-3.03 (1.54)

BRMA	-2.61 (0.43)	-3.70 (1.55)

SD = standard deviation; ; ESS = Epworth Sleepiness Scale; BP = blood pressure; URMA = univariate random effects meta-analysis; BRMA = bivariate random effects meta-analysis

The pooled estimates are shown alongside the estimated underlying true treatment effects ($essMu_i$ and $bpMu_i$) for each study in Figure 2. The estimated underlying treatment effects on BP for each study are dispersed in a flat line around the pooled estimate in the case of the URMA. In contrast, the estimated underlying treatment effects for those studies that did not report BP are informed by the positive between-study correlation in the BRMA. The estimated between-study correlation (ρB) in treatment effects was 0.13 (SD=0.55) for ambulatory BP only, and 0.36 (SD=0.50) when ambulatory and office measurements were combined.

Figure 2. Comparison of the results of the bivariate random effects meta-analysis to separate univariate analysis for ESS and BP

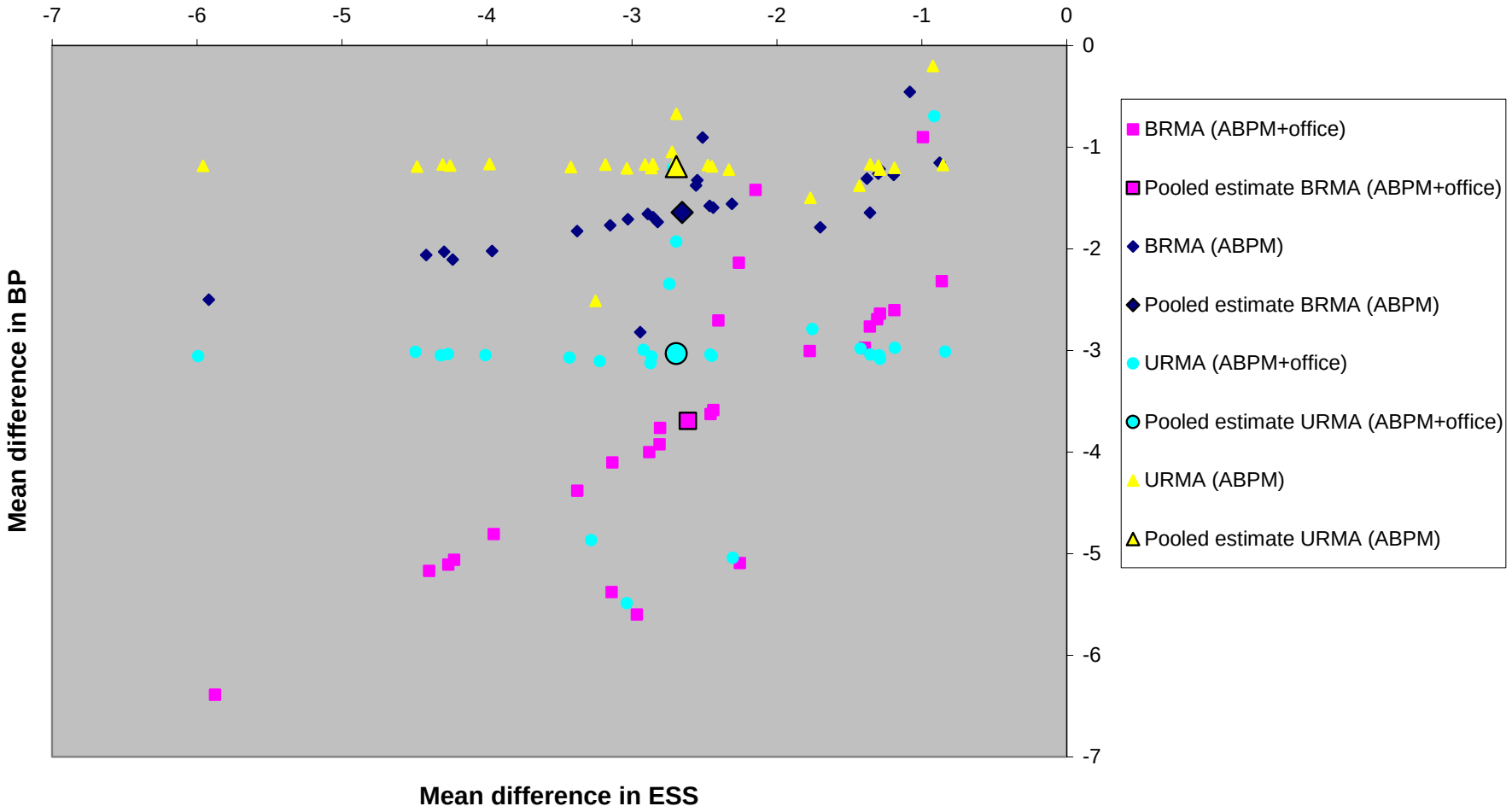


Table 3 shows the results of one-way sensitivity analyses varying the within- and between-study correlations to 0 and 1. Omitting the within- or between-study correlation has little impact on the pooled treatment effect of CPAP on ESS. The pooled treatment effect of CPAP on BP is reduced if the correlation is omitted. The impact of omitting the between-study correlation is increased when both ambulatory and office measurements of BP are included as the estimated between-study correlation is larger. Perfect correlation in treatment effects would reduce the pooled treatment effect of CPAP on ESS, and increase the pooled treatment effect on BP. Again the impact of setting the within- and/or between-study correlation to one is only small for the treatment effect on ESS, and larger for the treatment effect on BP.

Table 3. Sensitivity of BRMA to extreme values for within and between-study correlation

	Mean difference in ESS mean (SD)	Mean difference in BP mean (SD)
Ambulatory BP only		
Full model	*-2.65 (0.43)*	*-1.64 (1.72)*
Omit correlation:		
Between study correlation = 0	-2.68 (0.44)	-1.39 (1.57)
Within study correlation = 0	-2.69 (0.45)	-1.36 (1.75)
Between & within study correlation = 0	-2.69 (0.44)	-1.23 (1.65)
Extreme value correlation:		
Between study correlation = 1	-2.61 (0.43)	-2.24 (1.96)
Within study correlation = 1	-2.55 (0.42)	-2.55 (1.74)
Between & within study correlation = 1	-2.43 (0.42)	-3.89 (1.91)
Ambulatory + office BP		
Full model	*-2.61 (0.43)*	*-3.70 (1.55)*
Omit correlation:		
Between study correlation = 0	-2.65 (0.43)	-3.20 (1.45)
Within study correlation = 0	-2.65 (0.44)	-3.44 (1.54)
Between & within study correlation = 0	-2.67 (0.44)	-3.11 (1.50)
Extreme value correlation:		
Between study correlation = 1	-2.53 (0.41)	-4.45 (1.42)
Within study correlation = 1	-2.48 (0.42)	-4.48 (1.43)
Between & within study correlation = 1	-2.40 (0.44)	-5.11 (1.08)

SD = standard deviation; ; ESS = Epworth Sleepiness Scale; BP = blood pressure; URMA = univariate random effects meta-analysis; BRMA = bivariate random effects meta-analysis

4. Discussion

In the URMA any missing endpoints are assumed to be missing completely at random (MCAR). Data may be considered MCAR if the mechanism for missingness is completely random and does not depend on the missing values themselves or on any other variables in the dataset. The treatment effects in the univariate approach are informed only by the pooled estimate of the observed data, as in a complete case analysis. In contrast, in the BRMA the missing endpoints are assumed to be missing at random (MAR), and to depend on the observed between-study correlation in treatment effects on ESS and BP as well as the pooled estimates for each. Data may be considered MAR if the mechanism for missingness does not depend on the missing values themselves but can be explained by other variables in the dataset. For example, when BP is missing and ESS is observed, the imputed effect for BP is informed in part by the between-study correlation and the observed effect on ESS. If treatment effects on multiple outcomes are thought to be correlated, then any missing data will be MAR; the MCAR assumption is violated and applying a univariate approach can result in biased pooled estimates. The specification of alternative approaches to synthesis had little impact on the treatment effect that was nearly fully observed: 23 out of 26 studies reported ESS and so only three endpoints were missing. In contrast, BP was reported in only seven or ten studies, depending on the measures used, allowing the method of imputation to have a larger impact on the pooled treatment effect.

The bivariate model described in this paper could be extended to incorporate more than two outcomes. A benefit of the multivariate approach is that the same set of studies is used to inform all the pooled estimates of all of the outcomes of interest. If separate URMAs are conducted incorporating only those studies that report each outcome then the pooled estimates for each endpoint could be based on different patient populations, and heterogeneity could well be overlooked when assessing the overall impact of treatment from some weighted combination of outcomes calculated in a decision model. In addition, the approach allows the correlation between parameters to be incorporated in subsequent analyses. Unfortunately, in many cases the number of studies available to inform the estimate of between-study correlation may be small. In this case study the 95% credible interval for the between-study correlation was wide even with seven studies reporting both outcomes (-0.78 to 0.98). A set of individual patient level data were available from which an informative prior for the within-study correlation could be derived. In many cases, no information may be available on the within-study correlation but even then it is still possible to employ a multivariate approach.(Riley et al. 2007b)

The correlation between treatment effects on different endpoints matters after the data have been synthesised when the output from the meta-analysis is used to inform a decision-analytic model. By using the 10,000 iterations from which the pooled estimates are derived (i.e. the WinBUGS output) directly in the probabilistic decision-analytic model any correlation between treatment effects can be maintained. This in turn can affect the estimates of decision uncertainty and value of information analysis. In the evaluation of CPAP, utilising the results from the BRMA marginally improved the incremental cost-effectiveness ratio for CPAP due to the improved pooled estimate for BP. In addition it marginally increased the expected value of perfect information when compared to the analysis based on separate URMA for ESS and BP due to the increased standard deviation around the pooled estimates. In models that analyse individual patients rather than cohorts, the within-study correlation may also impact the expected costs and QALYs.

Meta-analysis may be considered to be observational as it involves the synthesis of data from those RCTs that are published. This issue of publication bias, where positive or statistically significant results may be more likely to be published may have a corollary in that only those secondary outcomes on which the treatment effect is positive or statistically significant may get reported in a published paper. If this is the case then the MAR assumption could be violated, as the missing endpoint data could depend on the missing value itself. However in the case of CPAP the issue of statistical significance does not seem to be of importance in the reporting of treatment effects on BP.

5. Conclusions

In general systematic reviews may focus on more than one outcome of interest. The model described here is generalisable to any number of outcome measures and takes advantage of information provided by the between-study correlation. In contrast conducting separate URMAs for each outcome assumes that any missing outcomes are MCAR and that there is no correlation between alternative outcome measures. A multivariate approach is therefore preferred if treatment effects are correlated as it will minimise missing data bias. The difference between the two approaches may have important consequences when the results are used to inform a cost-effectiveness analysis, particularly in terms of characterising uncertainty and estimating the value of further research.

References (to be completed)

Riley RD, Abrams KR, Lambert PC, et al. An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. Stats in Med 2007a **26**(1): 78-97.

Riley RD, Thompson JR and Abrams KR An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. Biostatistics 2007b **0**(0).