STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:1351–1375 (DOI: 10.1002/sim.1761)

What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data

Michael J. Sweeting[1], Alexander J. Sutton[2,*,†] and Paul C. Lambert[2]

[1]*MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, U.K.*
[2]*Department of Health Sciences, University of Leicester, 22-28 Princess Road West, Leicester LEI 6TP, U.K.*

SUMMARY

Objectives: To compare the performance of different meta-analysis methods for pooling odds ratios when applied to sparse event data with emphasis on the use of continuity corrections.

Background: Meta-analysis of side effects from RCTs or risk factors for rare diseases in epidemiological studies frequently requires the synthesis of data with sparse event rates. Combining such data can be problematic when zero events exist in one or both arms of a study as continuity corrections are often needed, but, these can influence results and conclusions.

Methods: A simulation study was undertaken comparing several meta-analysis methods for combining odds ratios (using various classical and Bayesian methods of estimation) on sparse event data. Where required, the routine use of a constant and two alternative continuity corrections; one based on a function of the reciprocal of the opposite group arm size; and the other an empirical estimate of the pooled effect size from the remaining studies in the meta-analysis, were also compared. A number of meta-analysis scenarios were simulated and replicated 1000 times, varying the ratio of the study arm sizes.

Results: Mantel–Haenszel summary estimates using the alternative continuity correction factors gave the least biased results for all group size imbalances. Logistic regression was virtually unbiased for all scenarios and gave good coverage properties. The Peto method provided unbiased results for balanced treatment groups but bias increased with the ratio of the study arm sizes. The Bayesian fixed effect model provided good coverage for all group size imbalances. The two alternative continuity corrections outperformed the constant correction factor in nearly all situations. The inverse variance method performed consistently badly, irrespective of the continuity correction used.

Conclusions: Many routinely used summary methods provide widely ranging estimates when applied to sparse data with high imbalance between the size of the studies' arms. A sensitivity analysis using several methods and continuity correction factors is advocated for routine practice.

KEY WORDS: meta-analysis; continuity corrections; zero cells; method comparison; sparse data

*Correspondence to: Alex J. Sutton, Department of Health Sciences, University of Leicester, 22-28 Princess Road West, Leicester LE1 6TP, U.K.
†E-mail: ajs22@le.ac.uk

Received August 2002
Accepted October 2003

BEFORE CONTINUITY CORRECTION

Group	Event	No Event	Total
Treatment	a	b	$a+b$
Control	c	d	$c+d$
Total	$a+c$	$b+d$	n

$$\hat{\Omega} = \frac{ad}{bc}$$

$$\text{var}(\ln(\hat{\Omega})) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

AFTER CONTINUITY CORRECTION (k)

Group	Event	No Event	Total
Treatment	$a+k$	$b+k$	$a+b+2k$
Control	$c+k$	$d+k$	$c+d+2k$
Total	$a+c+2k$	$b+d+2k$	$n+4k$

$$\hat{\Omega} = \frac{(a+k)(d+k)}{(b+k)(c+k)}$$

$$\text{var}(\ln(\hat{\Omega})) = \frac{1}{(a+k)} + \frac{1}{(b+k)} + \frac{1}{(c+k)} + \frac{1}{(d+k)}$$

Figure 1. Continuity corrections added to a 2×2 table of observed counts in a study comparing a binary outcome between two groups.

1. INTRODUCTION

The use of meta-analysis for combining results from several studies has become increasingly popular in medical research [1, 2]. When studies with binary outcomes (i.e. event or no event) are being combined, a number of competing methods of estimation, both classical and Bayesian, are available for obtaining a pooled study effect measure [1].

In some instances the outcome of interest may be very rare (or in some circumstances very common). For example, in a meta-analysis of RCTs evaluating a drug intervention, adverse side effects may be rare but serious and hence important [3]. An example where events are rare, leading to sparse data, is the randomized evidence on the risks of serious diseases such as breast cancer and heart disease from the use of hormone replacement therapy [4]. Additionally, events may also be sparse in data from studies using epidemiological designs. For example, a recent meta-analysis investigated the potential association between connective tissue disease and breast implants using very sparse data from epidemiological studies [5, 6]. In this paper trial terminology is often used to describe studies although this is for clarity and convenience and the results are directly generalizable to meta-analysis of other comparative study types, such as those found in epidemiology. Indeed, considerable attention is paid to studies with imbalanced group sizes, which occur commonly in epidemiology.

When zero events occur in either or both arms of a study, the log-odds ratio and the log risk ratio become undefined (as are their variances) and hence this causes problems with several of the potential analysis methods. To overcome this problem, a continuity correction factor, k, is often added to each cell of the 2×2 table for the studies with zero events in either arm (Figure 1).

Using a continuity correction for studies with zero counts allows the log-odds ratio or log risk ratio to be estimated, and hence allows synthesis via standard meta-analysis methods. Figure 1 shows how the use of a continuity correction has an influence on the odds ratio and its variance on the log scale.

Yates first used the term correction for continuity in 1934 [7]. The premise for the use of $\frac{1}{2}$ as a continuity correction is through the approximation of a discrete distribution by a continuous one. In standard 2×2 tables, it has been argued, that the use of $k = \frac{1}{2}$ makes the average approximation of the error equal to zero [8]. For further details of the general use of continuity corrections in 2×2 tables see Plackett [9]. Alternatives to $k = \frac{1}{2}$ have been discussed previously, including smaller values [10] and values dependent on the row and column of the cell entry [11, 12], but, to our knowledge, these or other alternatives have not been considered previously in a meta-analysis context.

When the proportion of studies in a meta-analysis with zero (or all) events in one or both arms is small, the choice of k usually has little influence on the results, but when event data is sparse for the majority of studies, as will be seen, continuity correction factors may exert an undesirable influence on the analysis results.

A previous study [13] has compared the strategy of using a continuity correction factor of $\frac{1}{2}$ where required, versus no continuity correction when meta-analysing odds ratios using the Mantel–Haenszel (M–H) method (see Section 2) and the standard meta-analysis random effects model [14]. In this investigation, studies with zero total events were included in the analysis by adding continuity corrections, however, although seen in practice [15], we consider this to be incorrect as such studies provide no information regarding the odds ratio (see Section 3). The investigation concluded that the uncorrected method performed better only when using the M–H odds ratio with very little heterogeneity present.

A more recent empirical investigation compared numerous methods for combining binary outcome data using meta-analysis when the outcome is rare (study currently only available in abstract form) [16]. Among the methods compared were the M–H, Peto, inverse variance and random effect models for combining on the odds ratio scale (see Section 2). This study came to the finding that the Peto method outperformed the other methods examined in many scenarios, which was rather unexpected since it had previously been reported to perform poorly by others under situations when there is severe imbalance in the numbers in the two groups being compared [17] and when the estimated odds ratio is far from unity [18]. However, this study only considered the single continuity correction factor of $\frac{1}{2}$ equal allocation ratio to both study groups, and no Bayesian meta-analysis methods were explored.

In this paper, a simulation study is described which aims to extend and address limitations in the previous work on sparse data meta-analyses. Here, classical and Bayesian fixed effect methods are explored for meta-analysis of binary outcomes on the odds ratio scale when data is sparse. Additionally, three approaches to setting k, the continuity correction factor, are examined in conjunction with the different methods, where required, and their influence on the pooled odds ratio is discussed. Consideration is given to scenarios where the numbers of subjects in the two groups are unbalanced.

The remainder of this paper is structured as follows. Meta-analysis methods for pooling odds ratios are briefly outlined in Section 2. The routine use of adding a continuity correction of $k = \frac{1}{2}$ is discussed in Section 3 and alternative continuity corrections are proposed and their rationales are discussed. Section 4 presents two motivating examples of sparse event meta-analysis. Section 5 outlines the simulation methods and scenarios used to create and analyse sparse event meta-analysis data. Section 6 presents the results of these simulations. Section 7, summarizes the key findings of this investigation and provides suggestions for further research.

2. META-ANALYSIS METHODS FOR POOLING ODDS RATIOS

This section outlines the different methods used for combining odds ratios in the illustrative examples and the simulation study described in Section 4; fuller accounts of these methods are available elsewhere [1]. It should be realized that all methods are estimating essentially the same parameters, although the method of estimation varies and intrinsic differences between classical and Bayesian inference mean the Bayesian model treats the parameters as random with prior distributions placed on them, while parameters are considered fixed in the classical approaches.

2.1. The inverse variance-weighted fixed effect model

The most common fixed effect method for meta-analysis is the inverse variance-weighted method. A pooled effect size is estimated by taking a weighted average of individual studies' effect estimates, weighting each study by the inverse of its variance. The log-odds ratio scale is often used for combining due its desirable statistical properties [19]. Hence, the inverse variance weighted pooled estimate is calculated by

$$\log(\hat{\Omega}_{\mathrm{IV}}) = \frac{\sum_{i=1}^{s} w_i \log(\hat{\Omega}_i)}{\sum_{i=1}^{s} w_i} \tag{1}$$

where the subscript i denotes the study number ($i = 1, \ldots, s$), $\hat{\Omega}_i$ is the odds ratio estimate for study i and $w_i = 1/\mathrm{Var}(\log(\hat{\Omega}_i))$. The equations used to estimate Ω_i and $\mathrm{Var}(\log(\Omega_i))$ are provided in Figure 1 (with and without continuity correction factor). Assuming normality of the pooled effect size, the variance of (1) can be calculated

$$\mathrm{Var}(\log(\hat{\Omega}_{\mathrm{IV}})) = \frac{1}{\sum_{i=1}^{s} w_i} \tag{2}$$

If any of the cells of the 2×2 table are zero, then a continuity correction factor is required if this method is used.

2.2. The Mantel–Haenszel summary odds ratio

This measure was first used by Mantel and Haenszel [20] in combining odds ratios for stratified case-control studies. In meta-analysis, it has been adapted so that each study forms an individual stratum. The pooled estimate of the common odds ratio is calculated using

$$\hat{\Omega}_{\mathrm{MH}} = \frac{\sum_{i=1}^{s} a_i d_i / n_i}{\sum_{i=1}^{s} b_i c_i / n_i} \tag{3}$$

where a, b, c, d and n are defined for a single study in Figure 1 and the subscript i indicates the ith of s studies to be combined. A commonly used estimate of the variance of the log of the M–H estimate, proposed by Robins *et al.* [21], (which is used throughout this paper) is given by the formula

$$\mathrm{Var}(\log(\hat{\Omega}_{\mathrm{MH}})) = \frac{\sum_{i=1}^{s} P_i R_i}{2(\sum_{i=1}^{s} R_i)^2} + \frac{\sum_{i=1}^{s} (P_i S_i + Q_i R_i)}{2(\sum_{i=1}^{s} R_i)(\sum_{i=1}^{s} S_i)} + \frac{\sum_{i=1}^{s} Q_i S_i}{2(\sum_{i=1}^{s} S_i)^2} \tag{4}$$

where $P_i = (a_i + d_i)/n_i$, $Q_i = (b_i + c_i)/n_i$, $R_i = a_i d_i / n_i$, and $S_i = b_i c_i / n_i$. Unlike the inverse variance-weighted method, the M–H pooled estimate can tolerate some studies with zero

cells, although continuity corrections are still commonly added. It has been suggested that this method outperforms other estimators when the study sample sizes are moderate [22]. However, the estimator is not consistent in sparse event data situations [22].

2.3. Peto's method for combining odds ratios

This method was first described by Yusuf *et al.* [23]. The pooled log-odds ratio estimate is

$$\log(\hat{\Omega}_{\mathrm{PETO}}) = \frac{\sum_{i=1}^{s}(O_i - E_i)}{\sum_{i=1}^{s} V_i} \tag{5}$$

where $O_i = a_i$ (the observed number of events in the treatment group for study i), $E_i = (a_i + b_i)(a_i + c_i)/n_i$ (the expected number of events in the treatment group for study i, *under the null hypothesis of no treatment difference*) and $V_i = (a_i+b_i)(a_i+c_i)(c_i+d_i)(b_i+d_i)/n_i^2(n_i-1)$. An estimate of the variance of equation (5) is given by

$$\mathrm{Var}(\log(\hat{\Omega}_{\mathrm{PETO}})) = \frac{1}{\sum_{i=1}^{s} V_i} \tag{6}$$

It should be noted that O_i, E_i and V_i are all equal to zero for studies with no events in either study arm. These studies therefore do not contribute to either the point estimate or variance of the pooled odds ratio. However, when a study has zero events in one study arm, the pooled odds ratio can still be calculated without the use of continuity corrections. It has been shown that the Peto method performs poorly in unbalanced designs and when the odds ratio differs substantially from one [24]. However, as mentioned in the introduction, Deeks *et al.* have shown the Peto method to provide the least biased and most powerful results when applied to simulated sparse event data with less extreme group imbalances, typically observed from RCT designs [16].

2.4. Logistic regression

Logistic regression can be used to pool binary outcomes from multiple studies. A fixed effect model is specified in equation (7). The numbers of events in the study arms are modelled using binomial distributions. The log odds for the control group for the ith study is denoted by μ_i. The log-odds ratio, denoted by δ, is fixed and the same for each study in the meta-analysis.

$$\begin{aligned} a_i &\sim \mathrm{Binomial}(a_i + b_i, p_{Ti}) \\ c_i &\sim \mathrm{Binomial}(c_i + d_i, p_{Ci}) \\ \mathrm{logit}(p_{Ci}) &= \mu_i \\ \mathrm{logit}(p_{Ti}) &= \mu_i + \delta \end{aligned} \tag{7}$$

Logistic regression is implemented by most statistical packages using iterative weighted least squares. The analogous Bayesian model can be fitted [25] although it is necessary to specify prior distributions for the unknown parameters—the log odds for the control group (μ_i) for each of the s studies and the underlying log-odds ratio (δ). In this paper we assume vague prior distributions, i.e.

$$\mu_i \sim \mathrm{Normal}(0, 10000) \;\; \delta \sim \mathrm{Normal}(0, 10000)$$

From a Bayesian perspective, the most popular approach for obtaining marginal densities for the parameters of interest is to use numerical simulation methods, in particular the Markov

Chain Monte Carlo (MCMC) method. Gibbs sampling is a particular type of MCMC method and is easily implemented using the WinBUGS software package [26]. The use of Gibbs sampling has become increasingly popular since the method requires that we only have to sample from the conditional densities in order to obtain the marginal densities for each of the unknown parameters.

Since this approach assumes events in the treatment groups are binomially distributed, the models can handle studies with zero events. Therefore, there is no need for continuity corrections to be added to any studies for either the classical or the Bayesian implementation.

3. CONTINUITY CORRECTIONS

As discussed in Sections 1 and 2, continuity correction factors are required for some meta-analysis methods if zero events exist in study arms. However, it is currently unclear whether there is an optimum approach to the use of continuity correction factors in meta-analysis.

Clearly, adding different factors to each cell will have different influences on the conclusions drawn. Agresti [10] suggests adding very small constants (such as 10^{-8}) to the cells when there is sparse event data (though not specifically for meta-analysis). A sensitivity analysis was also suggested to measure the effect of using constants of various sizes (e.g. 10^{-8}, 10^{-4}, 0.01, 0.1) [10]. It should not be forgotten that continuity correction factors influence the estimates of the variance, and hence the weighting given to each study in addition to the estimate of magnitude of effect.

It is common practice to remove studies in which there are no events in both treatment arms from a meta-analysis on a relative scale (such as an odds ratio). Such studies are referred to as zero total event studies. Whitehead and Whitehead [27] argue that these studies should be dropped from the meta-analysis since they provide no information on the magnitude of the treatment effect. However, an argument has been made that such studies should be included in a meta-analysis [28] in order to take into account the sample sizes of these studies [13]. A preliminary investigation that preceded the main simulation study described here, by analysing data sets both including and excluding such studies using the Bayesian model that does not require continuity correction factors, confirmed that zero total event studies do not contribute to a fixed effect meta-analysis. (Further details of this investigation are available on request from the authors.) As a result of this, it was decided to exclude total zero studies when using all methods.

The account below describes the rationale for three different approaches to implementing continuity correction factors in a meta-analysis model.

3.1. Adding a constant k to each cell in the 2×2 contingency table

When zero event data occur in a meta-analysis, most often researchers use a constant correction factor. By far the most widespread constant that is chosen is $k = \frac{1}{2}$. The reason for choosing $k = \frac{1}{2}$ may have arisen from analysis by Cox [29] on the odds of a single study group. Cox's analysis states that when using the odds as the effect measure, choosing a correction factor of $\frac{1}{2}$ gives the least biased estimator of the true log odds in a single treatment group situation. However, using the continuity correction of $k = \frac{1}{2}$ may be outperformed in terms of both bias and coverage by other choices of correction factor when studying the odds ratio between two groups. Particularly, the correction factor of $\frac{1}{2}$ may not perform well when the groups are

Table I. An illustration of using a continuity correction of $\frac{1}{2}$ on the effect of the odds ratio.

Group	Event	No event	Total
Treatment	0	100	100
Control	1	400	401
Total	1	500	501

$k = 1/2 \Rightarrow$ *Odds ratio estimate* $(\hat{\Omega}) = \dfrac{0.5 \times 400.5}{100.5 \times 1.5} = 1.33.$

Table II. The effect of continuity corrections on the odds ratio of a study with no events.

1:1 Balanced groups*			1:2 Unbalanced groups†			1:4 Unbalanced groups‡		
Group	Event	No event	Group	Event	No event	Group	Event	No event
Treatment	0	100	Treatment	0	100	Treatment	0	100
Control	0	100	Control	0	200	Control	0	400

*$CC = 1/2 \Rightarrow \hat{\Omega} = 1;\ CC = 1/100(1/TA) \Rightarrow \hat{\Omega} = 1.$
†$CC = 1/2 \Rightarrow \hat{\Omega} \approx 2;\ CC = 1/200\ \&\ 1/100 \Rightarrow \hat{\Omega} = 1.$
‡$CC = 1/2 \Rightarrow \hat{\Omega} \approx 4;\ CC = 1/400\ \&\ 1/100 \Rightarrow \hat{\Omega} = 1.$

severely unbalanced with respect to the number of subjects in each. This problem is illustrated by example in Table I for a situation where the common $k = \frac{1}{2}$ correction factor can give a misleading odds ratio. In the example, results of a theoretical trial, in which the control group has four times as many patients, are presented. There is only one event in the trial which occurs in the control group. Implementing a continuity correction of $\frac{1}{2}$ gives an odds ratio of 1.33 suggesting an increased risk in the treatment group where no events were observed. A further observation is that although the odds ratio scale should be symmetrical, in the sense that the same result will be obtained whether the number of persons experiencing an event or the number of persons not experiencing an event are modelled, this does not hold true when a constant continuity correction factor is used.

3.2. The reciprocal of the opposite treatment arm size

Another possible continuity correction to add is a factor of the reciprocal of the size of the opposite treatment arm to the cells. For example, in Table I the treatment group would be adjusted by a constant $k/401$ while the control group would be adjusted by $k/100$, where k is a proportionality constant of a chosen size. This adjustment may cause less bias when the size of the treatment groups are severely unbalanced. The reason for adding a factor of the reciprocal of the size of the opposite treatment arm becomes clear when considering a study with zero events in each group (Table II). The total zero event studies shown in Table II are for illustration purposes only, we re-iterate we do not advocate including such studies in a meta-analysis but these studies clearly illustrate the effect of different continuity corrections as the odds ratio estimate will be derived solely from the continuity correction that is used. Table II shows that a continuity correction of $\frac{1}{2}$ gives approximate odds ratio estimates of 1 for balanced groups, 2 for an imbalance of 1:2 and 4 for an imbalance of 1:4, etc. However, the

Table III. Derivation of the empirical continuity corrections.

Group	Event	No event
Treatment	$0 + k_T$	$n_T + k_T$
Control	$0 + k_C$	$n_T \times R + k_C$

'treatment arm' continuity correction always leads to an odds ratio of one in this situation. This allows the continuity correction factor not to favour one group over the other in the meta-analysis.

It can also be noted that using the 'treatment arm' continuity correction gives an odds ratio of 0.010 for the example given in Table I ($OR = (0.0025 \times 400.01)/(100.0025 \times 1.01) = 0.010$). This odds ratio estimate, unlike the one obtained using a continuity correction of $\frac{1}{2}$, is (much) less than one.

3.3. An empirical continuity correction

The influence of the constant continuity correction factor depends on the ratio of the group sizes. When the 'treatment arm' continuity correction is used this has the influence of 'pulling' the estimated odds ratio arbitrarily towards no effect (i.e. $OR = 1$). Perhaps, more desirable still, would be for the continuity correction factor to 'pull' the estimate in the direction of the pooled effect size estimate obtained in the analysis. An empirical approach can be adopted, where all the studies in the meta-analysis without a zero event are used to calculate a pooled odds ratio. Using this estimate, a continuity correction can be calculated, which will produce odds ratio estimates close to the pooled odds ratio in the studies with zero events in both arms. Using pseudo Bayesian terminology we can think of this continuity correction factor acting as a 'prior' (empirically derived from the other studies) added to the observed events. For example, suppose that an estimated pooled odds ratio, $\hat{\Omega}$, was obtained using the non-zero studies. Let n_T be the number of persons in the treatment group, R be the group ratio imbalance and hence the number of persons in the control group is $n_T \times R$. Then a total zero event study with continuity corrections k_T and k_C, for the treatment and control groups respectively, is shown in Table III.

Then for k_T to k_C be empirical continuity corrections, we want

$$\frac{k_T(n_T R + k_C)}{k_C(n_T + k_T)} = \hat{\Omega} \tag{8}$$

The left-hand side of equation (8) can be approximated by Rk_T/k_C when the group sample sizes are reasonably large. If a restriction that $k_T + k_C = 1$ is imposed as with a continuity correction of $\frac{1}{2}$, then the following equations are obtained:

$$\frac{R(1 - k_C)}{k_C} \approx \hat{\Omega}$$

$$\Rightarrow k_C \approx \frac{R}{R + \hat{\Omega}} \tag{9}$$

$$\Rightarrow k_T \approx \frac{\hat{\Omega}}{R + \hat{\Omega}} \tag{10}$$

Table IV. Illustrative example of empirical continuity correction.

	Empirical CC('Prior')*		Study data (Observed events)†		Data + empirical CC‡	
	Event	No event	Event	No event	Event	No event
Treatment	0.23	100.23	0	100	0.23	100.23
Control	0.77	200.77	3	200	3.77	200.77

*$\hat{\Omega} \approx 0.60$ (derived from the other studies in meta-analysis).
†$\hat{\Omega}$ = Undefined.
‡$\hat{\Omega} = 0.12$.

Equations (9) and (10) give the appropriate empirical continuity corrections that can be added to the control and treatment groups, respectively. In essence, we are combining our 'prior' information before any events are observed with the data from the specific study to produce an odds ratio estimate. As an example, assume the pooled odds ratio using all the non-zero studies was estimated as $\hat{\Omega} = 0.6$. Suppose a particular study had zero events in the treatment group, three events in the control group and an imbalance of 1:2 as set out in Table IV. The empirical continuity correction can be calculated as 0.23 for the treatment group and 0.77 for the control group. Our 'prior' odds ratio estimate for the study is $\hat{\Omega} = 0.6$. However, when combined with the data our estimate changes to $\hat{\Omega} = 0.12$.

As suggested by Agresti [7], adding smaller continuity corrections to the cells may provide better estimates for the true odds ratio. Therefore, the constraints for the empirical continuity corrections can be changed accordingly. For example, if we wish $k_T + k_C = 0.01$ then equations (9) and (10) become

$$k_C \approx \frac{R}{100(R + \hat{\Omega})} \tag{11}$$

$$k_T \approx \frac{\hat{\Omega}}{100(R + \hat{\Omega})} \tag{12}$$

This method of obtaining a continuity correction appears to be new in meta-analysis.

In summary, three different continuity corrections have been discussed in this section. The constant continuity correction gives a 'prior' odds ratio estimate equal to the ratio of the two treatment arms. The 'treatment arm' continuity correction gives a 'prior' odds ratio of 1 while the empirical continuity correction gives a 'prior' odds ratio of our 'best guess' of the pooled odds ratio (from the non-zero event studies). Table V provides a comparison of the continuity corrections for five studies that differ by the group size ratio.

4. MOTIVATING EXAMPLES

Two meta-analyses data sets in which data are sparse are analysed below using various combinations of pooling method and continuity correction factors. Two of each of the three types of continuity corrections described in Section 3, as outlined in Table VI, were applied to

Table V. Comparison of three continuity corrections and their influence on the estimated odds ratio (assuming empirical CC estimates an odds ratio of 0.5 from previous studies).

				Estimate of odds ratio using continuity correction		
	Event	No Event	Group ratio	1/2	'Treatment arm' where $k_T + k_C = 1$	'Empirical' where $k_T + k_C = 1$
Treatment	0	100	1:1	0.33	0.33	0.20
Control	1	99				
Treatment	0	100	1:2	0.66	0.40	0.22
Control	1	199				
Treatment	0	100	1:4	1.33	0.44	0.23
Control	1	399				
Treatment	0	100	1:8	2.65	0.47	0.24
Control	1	799				
Treatment	0	100	1:16	5.31	0.48	0.25
Control	1	1599				

Table VI. Continuity corrections used in the motivating examples and simulations.

	Type of continuity correction		
Magnitude of Correction	Constant CC	'Treatment Arm' CC	Empirical CC (see equations 9 and 10)
		1:R Imbalance	1:R Imbalance
$TCC + CCC = 1$	$TCC = \frac{1}{2}$	$TCC = 1/(R+1)$	$TCC = \hat{\Omega}/(R+\hat{\Omega})$
	$CCC = \frac{1}{2}$	$CCC = R/(R+1)$	$CCC = R/(R+\hat{\Omega})$
		1:R Imbalance	1:R Imbalance
$TCC + CCC = 0.01$	$TCC = 1/200$	$TCC = 1/(100(R+1))$	$TCC = \hat{\Omega}/(100(R+\hat{\Omega}))$
	$CCC = 1/200$	$CCC = R/(100(R+1))$	$CCC = R/(100(R+\hat{\Omega}))$

Key: CC = Continuity correction; TCC = Treatment group Continuity Correction; CCC = Control group Continuity Correction; R = Ratio of group sizes; $\hat{\Omega}$ = Estimated common odds ratio from the non-zero event studies.

the data sets as required. The constant continuity correction in which $\frac{1}{2}$ is added to both the treatment and control group cells (Table VI) is the continuity correction used routinely. As a comparison to this method, the 'treatment arm' and empirical continuity corrections are also constrained to sum to one. This ensures that the same number of counts and hence 'information' is being added to each study, where required, no matter which continuity correction method is used. The second row of Table VI constrains the continuity corrections for the treatment and control group to sum to 0.01. This will investigate the effect of adding smaller corrections to each cell. As discussed previously, studies with total zero events were removed from all analyses.

Table VII. Raw data from 23 trials of hormone replacement therapy and cardiovascular events and breast cancer.

	Women allocated		Cardiovascular disease	
Study no.	Control	Treatment	Events in control	Events in treatment
1	137	1128	—	—
2	174	701	0	5
3	78	39	—	—
4	42	40	0	0
5	32	46	—	—
6	14	15	1	0
7	51	100	—	—
8	39	36	—	—
9	25	50	0	0
10	19	41	—	—
11	40	116	—	—
12	16	15	0	1
13	19	21	—	—
14	20	20	1	1
15	39	61	—	—
16	54	60	—	—
17	24	76	—	—
18	48	44	—	—
19	26	29	0	1
20	121	56	—	—
21	84	84	3	1
22	30	120	0	0
23	66	68	0	3

— = Data not mentioned in study report assumed to be 0 in the analyses.

4.1. Example 1: Impact of postmenopausal hormone therapy on cardiovascular events: pooled data from clinical trials

Observational studies have suggested that postmenopausal hormone replacement therapy may decrease the incidence of cardiovascular diseases. To investigate these potential associations further, Hemminki and McPherson [4] pooled data from randomized clinical trials that compare hormone therapy with a placebo. Twenty-three trials with a total of 4164 women were identified as suitable. Of these 23 trials, 13 (56.5 per cent) did not mention any cardiovascular events and the remaining 10 trials reported very small incidences in both treatment arms. The data is therefore sparse and hence it is suspected that the results will be sensitive to the method of analysis conducted. Table VII presents the raw data collected from the 23 trials for cardiovascular diseases. Note many trials did not specifically report this outcome (indicated by dashes in the table). It is assumed here, as it was in the original analysis, that where no events are reported there are zero total events. Hence, such trials are excluded from the analysis. The sparse nature of the data can immediately be seen. Many of the trials have severely unbalanced group sizes, since different treatment regimes and doses have been pooled to create a single treatment group.

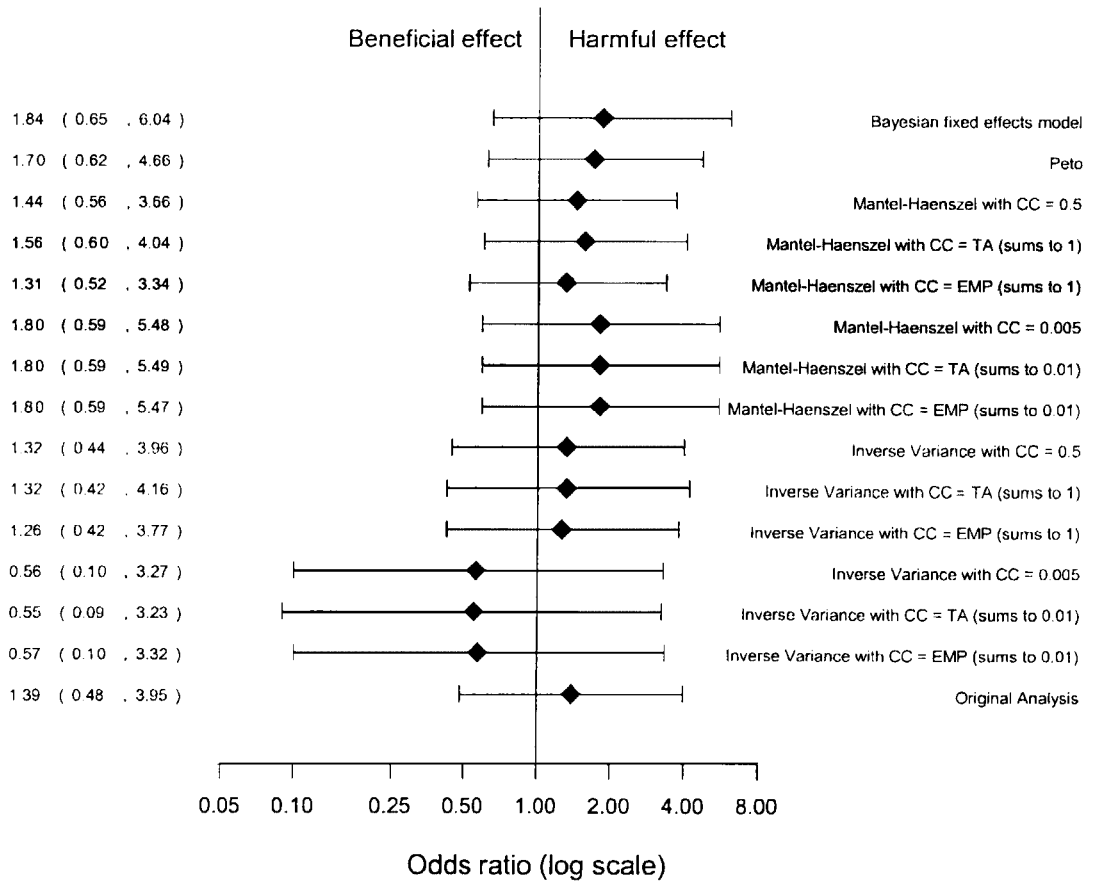

Figure 2. Pooled odds ratio estimates with 95 per cent confidence intervals studying the impact of HRT on cardiovascular diseases (TA = treatment arm continuity correction, EMP = empirical continuity correction).

The authors used a marginal analysis to estimate a pooled odds ratio (i.e. merge all trial results to produce one single two by two table from which the odds ratio is estimated). The problems with this approach are well documented [30, 31] and a potentially misleading result is produced since it makes the assumption that the underlying risk of an event is constant across trials. However, such an approach does allow studies with no events in either arm to be included.

Figure 2 presents the analysis of the impact of HRT on cardiovascular diseases using various meta-analysis pooling methods and continuity corrections. The results suggest that the odds ratio for cardiovascular diseases may be as high as 1.84 (95 per cent CI (0.65, 6.04)) when using a Bayesian analysis. However, a protective effect is suggested when using an inverse variance weighted pooling method with a small continuity correction of 0.005. Although the conclusions drawn are consistent across the methods, with all 95 per cent confidence intervals

Table VIII. Raw data from seven comparative cohort studies examining the effect of electronic fetal heart monitoring (EFM) on perinatal mortality.

	Number of subjects		Number of perinatal deaths		Group size ratio
Study No.	EFM given	EFM not given	EFM given	EFM not given	
1	1162	5427	2	17	1:4.67
2	150	6836	0	15	1:45.57
3	608	6179	1	37	1:10.16
4	4210	2923	1	9	1:0.69
5	554	692	1	3	1:1.25
6	4978	8634	0	2	1:1.73
7	45880	66208	10	45	1:1.44

containing an odds ratio of 1, point estimates do vary qualitatively across methods and there is also considerable variation in the width of the confidence intervals.

4.2. Example 2. Electronic fetal heart rate monitoring and perinatal mortality

The use of electronic fetal heart rate monitoring (EFM) in the 1970s coincided with a fall in the overall perinatal mortality rate. The causation of EFM on the fall in the perinatal mortality rate is plausible since healthy babies sometimes die during labour and timely caesarean section usually prevents this. However, systematic reviews of randomized controlled trials have concluded that EFM has not been shown to reduce perinatal mortality. A recent systematic review addressed this problem by including non-randomized evidence from observational studies [32]. Table VIII shows the raw data from seven comparative cohort studies examining the effect of EFM on perinatal mortality; note the extreme size imbalances between groups in some of these studies.

Pooled estimates using different methods and continuity correction factors are shown in Figure 3. While there is broadly more agreement between methods than for the hormone replacement therapy data set described above, the Peto summary estimate differs slightly from the other methods, possibly due to the imbalanced nature of the data (especially in study 2).

Hence, variation in point estimates can be obtained by using different pooling methods. The patterns in the various estimates for each example are quite different. In the hormone replacement therapy example, there was wide variation between many of the methods, whereas there was broad agreement with the exception of the Peto method in the EFM example. The simulation study that follows investigates the performance of different pooling methods on sparse event data sets with both balanced and imbalanced group sizes using the continuity correction factors outlined in Section 3.

5. SIMULATIONS

5.1. Scenarios

Table IX shows the list of parameters and their assigned values used to simulate the sparse event meta-analysis data sets. It was assumed that the studies within a meta-analysis would

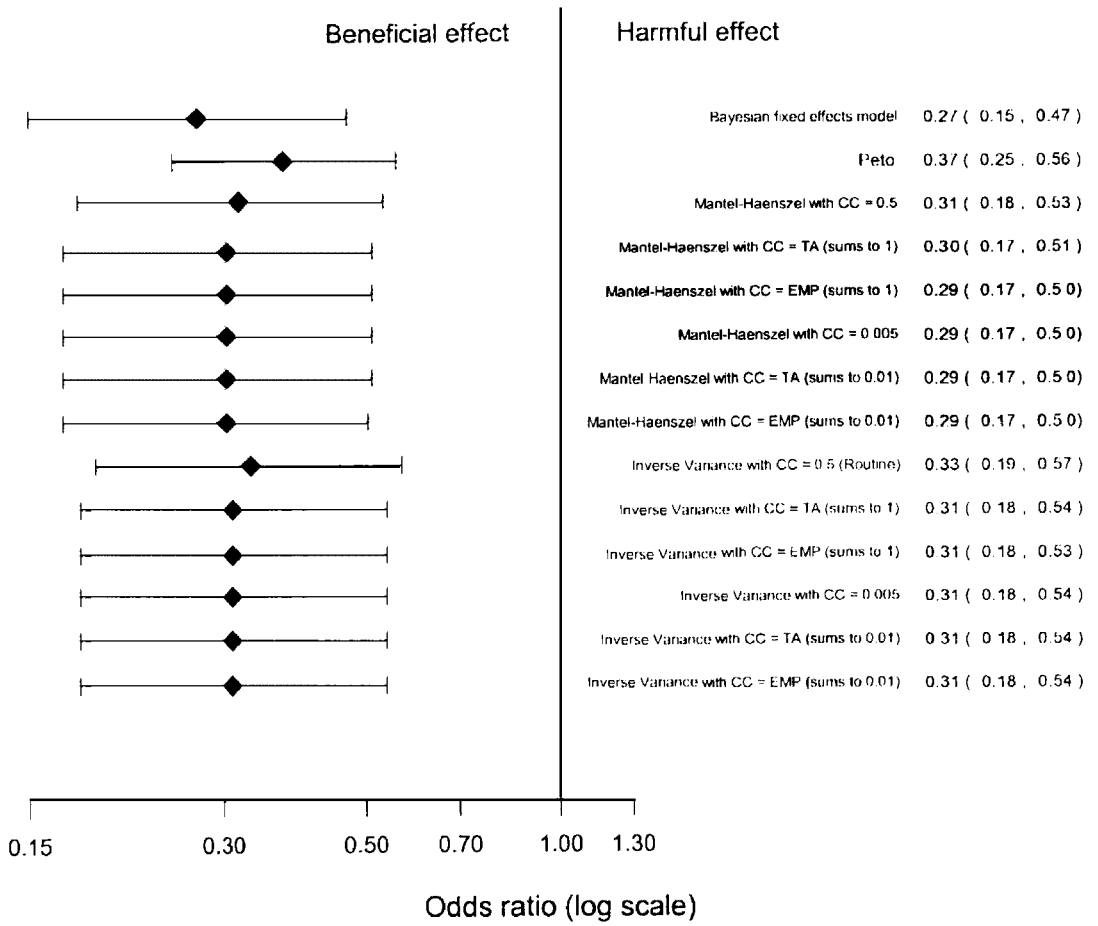

Figure 3. Pooled odds ratio estimates with 95 per cent confidence intervals studying the impact of EFM on perinatal mortality from seven comparative cohort studies (TA = treatment arm continuity correction, EMP = empirical continuity correction).

all have the same group size ratio (R). It was decided to let R vary from balanced groups ($R=1$) to severely unbalanced groups ($R=8$). It was assumed that the treatment effects were homogeneous (i.e. the between study variation in treatment effects was set to 0). We did examine more heterogeneous scenarios. These are not reported in detail here, but are considered in the discussion.

An odds ratio (Ω) of 0.5 simulates data in which patients in treatment group have decreased odds of 50 per cent of experiencing an event compared to the control group. The control group event rate (p_C) was chosen to be 0.01. This simulates sparse event data (an average of 1 in 100 patients in the control group will experience the event of interest). The treatment group event rate (p_T) for study i was then obtained from the control group event rate and the odds ratio (Ω). Each meta-analysis was simulated to contain 10 studies as we believe this represents

Table IX. Parameter values used in the simulation of sparse event data meta-analyses.

Fixed parameters	Assigned value
Odds ratio (Ω)	0.5
Control group event probability (p_C)	0.01
Number of studies in each meta-analysis (s)	10
Number of simulated meta-analysis data sets (n)	1000
Number of patients in treatment group ($n_T = a + b$)	(50,50,100,100,100,200, 200,300,300,400)
Between study variance	0
Varied parameters	
Ratio of group sizes (R)	1:1, 1:2, 1:4, 1:8 (T:C) ($R = 1, 2, 4, 8$)

a typically sized systematic review in medicine. In each of the simulated meta-analyses the composition of the size of the treatment arms in the 10 studies was always the same (i.e. 50, 50, 100, 100, 100, 200, 200, 300, 300, 400). One thousand meta-analyses were simulated for each set of parameter values. The model used to simulate the data sets is defined in equations (13)–(15):

$$p_T = \frac{\left(\frac{p_C}{1 - p_C}\right)\Omega}{1 + \left(\frac{p_C}{1 - p_C}\right)\Omega} \tag{13}$$

$$c_i \sim \text{Binomial}((R \times n_{Ti}), p_C) \tag{14}$$

$$a_i \sim \text{Binomial}(n_{Ti}, p_T) \tag{15}$$

5.2. Applying continuity corrections

The six continuity correction factors outlined in Table VI were applied to all simulated data sets as required. Studies with total zero events were removed from all analyses. A median of 2 (IQR 1–2, range 0–6) studies with total zero events were removed from each balanced group meta-analysis. Note, even with 1000 data sets the bias and coverage are of course prone to some (small) sampling error themselves.

5.3. Software and implementation issues

Simulations for all data set scenarios were performed in S-PLUS [33]. Analysis of the classical pooling methods was also carried out in S-PLUS, while the Bayesian fixed effect analyses were carried out using WinBUGS [26]. In all Bayesian analyses a Normal (0, 10 000) distribution was specified as the prior distribution for the summary log odds ratio and baseline log odds for each study (i.e. δ and the μ_is in equation (7)). Autocorrelation and trace plots of the 1000 estimated pooled log odds ratio values were examined to assess convergence of the MCMC chains. All the autocorrelation plots suggested little autocorrelation existed within MCMC chains and the trace plots provided no evidence that convergence of the sampler had not been

attained. A burn-in of 1000 iterations was followed by a further 3000 iterations. Estimation and inferences were based on all 3000 sample iterations.

6. RESULTS

Summary odds ratios from each scenario were compared for bias on the log odds scale from the true odds ratio of 0.5. The coverage properties of the 95 per cent confidence/credible intervals for the pooled (log) odds ratio were also compared. By definition, the 95 per cent confidence intervals for the pooled odds ratio should include the mean true odds ratio for 95 per cent of simulations. Coverage is estimated over the 1000 meta-analyses by calculating the percentage of times the true odds ratio of 0.5 is actually included within the estimated 95 per cent confidence intervals. Using 1000 simulated data sets, the standard error of the estimated coverage, will be approximately 0.7 per cent. All of the simulation scenarios calculated the standard error of the log odds estimate (and hence bias) to be 0.01, correct to two decimal places. The largest standard error of the log odds ratio was calculated to be 0.014 by the Bayesian method.

6.1. Balanced groups (Ratio 1:1)

The top portion of Table X presents the bias and coverage estimates from the simulations where the group sizes were balanced. These are presented graphically in Figure 4(a). It should be noted that the summary estimate for each Bayesian fixed effect analysis is the median of the posterior for the pooled log-odds ratio. The Peto, logistic regression and Bayesian fixed effect models have one result for each group imbalance, since, they do not require continuity corrections. Additionally, in the case of balanced groups, the constant and reciprocal of the opposite arm size continuity correction factors are identical.

From Figure 4(a) and Table X it can be seen that both the Mantel–Haenszel pooled odds ratio using a small continuity correction and logistic regression provide the least biased estimate for the odds ratio when the treatment groups are balanced. The majority of the classical methods are positively biased and hence underestimate the true treatment effect. The Bayesian fixed effect model slightly overestimates the treatment effect (negative bias of -0.04). The inverse variance method has the highest bias no matter what continuity correction is used. The Peto summary estimate gives almost unbiased results although the confidence intervals are too wide. Using a smaller continuity correction for the Mantel–Haenszel method produces a less biased estimate. However, the small continuity corrections have an opposite effect on the bias and coverage when applied to the inverse variance method. The bias is now considerably higher. The Mantel–Haenszel summary odds ratio was also investigated with no continuity correction. The method performed poorly in the balanced group situation (bias $=-0.195$, coverage $=94.6$).

Nearly all the pooling methods are conservative as shown by coverage greater than 95 per cent. The Bayesian fixed effect analysis gives the best coverage of 95.2 per cent. The confidence intervals calculated from the inverse variance method give very high coverage ranging from 96.9 to 97.9 per cent depending on the continuity correction used.

An empirical continuity correction that sums to one appears to be slightly better than the standard continuity correction of $\frac{1}{2}$ in all the main classical pooling methods.

Table X. Results of simulations with underlying odds ratio of 0.5 for different group imbalances, different continuity corrections and different meta-analysis methods: Mean absolute bias of log-odds ratio (bold), and coverage of the 95 per cent confidence/credible intervals (*italics*).

Group imbalance	Type of continuity correction	Sum of CC in the two arms	Mantel–Haenszel	Inverse variance	Peto (using no CC)	Logistic regression (using no CC)	Bayesian fixed effect analysis (using no CC)
1:1	None	0	**−0.20** *94.6%*	NA	**0.04** *97.1%*	**−0.02** *96.7%*	**−0.04** *95.2%*
	Constant (1/2)*	1	**0.09** *96.7%*	**0.15** *97.9%*			
	Constant (0.05)*	0.01	**−0.01** *96.8%*	**0.25** *96.9%*			
	Empirical	1	**0.03** *96.1%*	**0.13** *97.8%*			
	Empirical	0.01	**−0.01** *96.7%*	**0.26** *96.9%*			
1:2	None	0	**−0.10** *97.1%*	NA	**0.09** *97.5%*	**−0.05** *97.1%*	**−0.04** *95.1%*
	Constant (1/2)	1	**0.14** *95.4%*	**0.26** *93.5%*			
	Constant (0.005)	0.01	**−0.04** *97.1%*	**0.40** *88.5%*			
	α 1/No. in Opposite Group	1	**0.05** *96.9%*	**0.23** *94.1%*			
	α 1/No. in Opposite Group	0.01	**−0.04** *97.1%*	**0.41** *88.3%*			
	Empirical	1	**0.00** *96.6%*	**0.22** *93.8%*			
	Empirical	0.01	**−0.04** *97.1%*	**0.41** *88.3%*			
1:4	None	0	**−0.08** *95.1%*	NA	**0.13** *95.9%*	**−0.07** *95.0%*	**−0.11** *93.2%*
	Constant (1/2)	1	**0.16** *92.5%*	**0.33** *85.3%*			
	Constant (0.005)	0.01	**−0.06** *95.0%*	**0.45** *79.4%*			
	α 1/No. in Opposite Group	1	**0.00** *95.7%*	**0.28** *87.1%*			
	α 1/No. in Opposite Group	0.01	**−0.06** *95.0%*	**0.46** *79.0%*			
	Empirical	1	**−0.03** *94.4%*	**0.29** *87.2%*			
	Empirical	0.01	**−0.06** *95.0%*	**0.46** *79.0%*			
1:8	None	0	**−0.04** *95.9%*	NA	**0.17** *95.1%*	**−0.04** *95.9%*	**−0.08** *94.4%*
	Constant (1/2)	1	**0.20** *92.0%*	**0.39** *78.9%*			
	Constant (0.005)	0.01	**−0.04** *95.9%*	**0.51** *71.6%*			

Table X. *Continued.*

Group imbalance	Type of continuity correction	Sum of CC in the two arms	Mantel–Haenszel	Inverse variance	Peto (using no CC)	Logistic regression (using no CC)	Bayesian fixed effect analysis (using no CC)
1:8	α 1/No. in Opposite Group	1	**−0.00** *96.4%*	**0.38** *79.4%*			
	α 1/No. in Opposite Group	0.01	**−0.04** *95.9%*	**0.52** *70.6%*			
	Empirical	1	**−0.02** *95.8%*	**0.38** *79.8%*			
	Empirical	0.01	**−0.04** *95.9%*	**0.52** *70.6%*			

*Equivalent to 1/No. in Opposite Group.

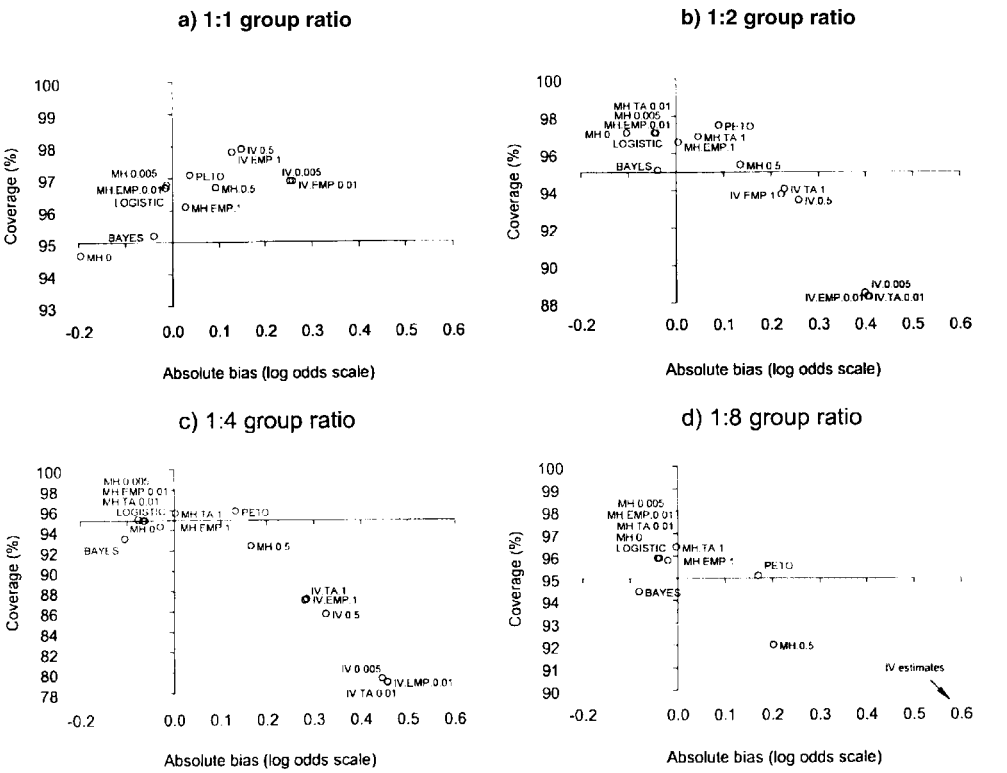

Figure 4. Scatter plot of bias against coverage for meta-analyses with (a) 1:1 group ratio, (b) 1:2 group ratio, (c) 1:4 group ratio and (d) 1:8 group ratio. MH = Mantel Haenszel method, IV = Inverse Variance-weighted method, PETO = Peto pooling method, BAYES = Bayesian fixed effect method, LOGISTIC = Logistic regression, 0.5 = Continuity correction of $\frac{1}{2}$, 0.005 = Continuity correction of 0.005, TA.1 = Ratio of treatment arms that sums to 1, TA.0.01 = Ratio of treatment arms that sums to 0.01, EMP.1 = Empirical CC that sums to 1, EMP.0.01 = Empirical CC that sums to 0.01.

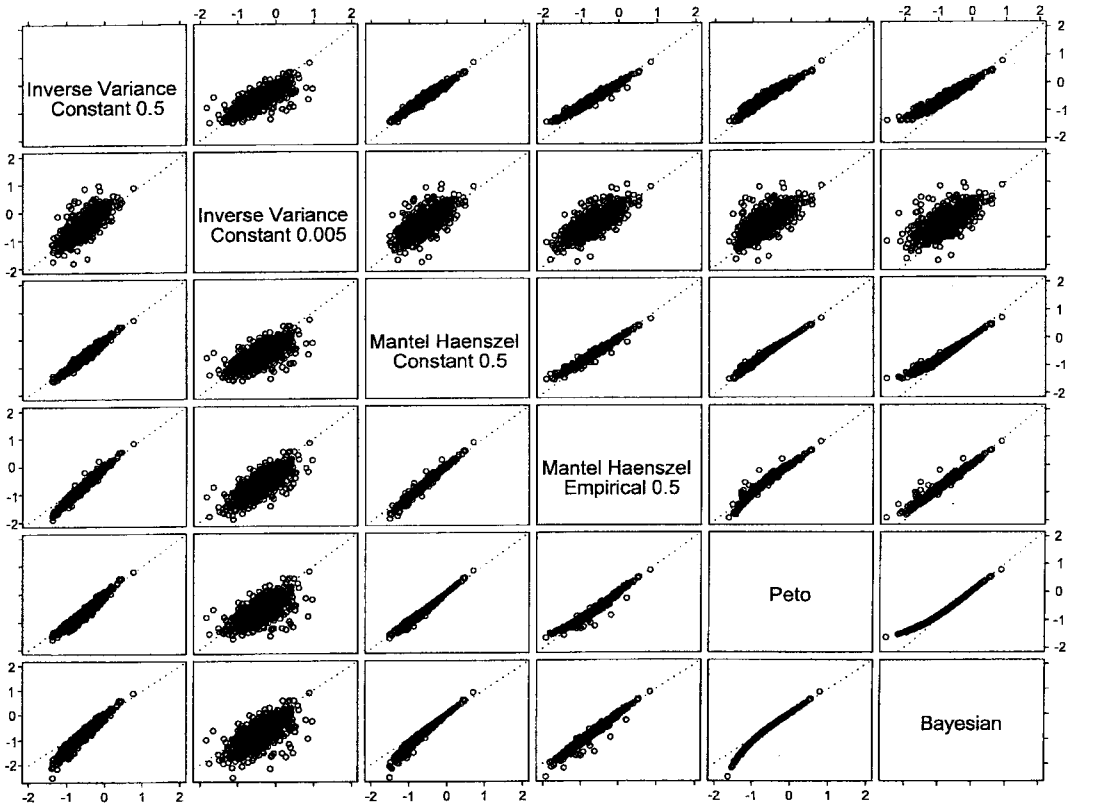

Figure 5. Scatter plot matrix showing relationship between the log-odds ratios of selected pooling methods for each of the 1000 simulated data sets.

Figures 5 and 6 show scatter plots for a selection of the different pooling methods/continuity corrections for balanced groups for the agreement between the log-odds ratios and the variance of the log-odds ratios, respectively. It can be seen that, generally, there is disagreement between the estimates of the log-odds ratio and its associated variance. The best agreement is between the Peto and Mantel–Haenszel method using a continuity correction of $\frac{1}{2}$, for both the log-odds ratio and its variance. Considerable disagreement can be seen between the variances of the log-odds ratios with systematic patterns clear. For example, the variance is always larger using the inverse variance with a constant 0.005 continuity correction compared to a constant $\frac{1}{2}$ continuity correction. This is considered further in the discussion.

6.2. Group size ratio 1:2

Figure 4(b) displays the simulation results for unbalanced group sizes of ratio 1:2. As with the balanced group scenario, the Mantel–Haenszel method with an empirical continuity correction (sum to 1) gives the least biased log-odds ratio estimates (0.004). The routine methods of using the Mantel–Haenszel or inverse variance method with a continuity correction of $\frac{1}{2}$

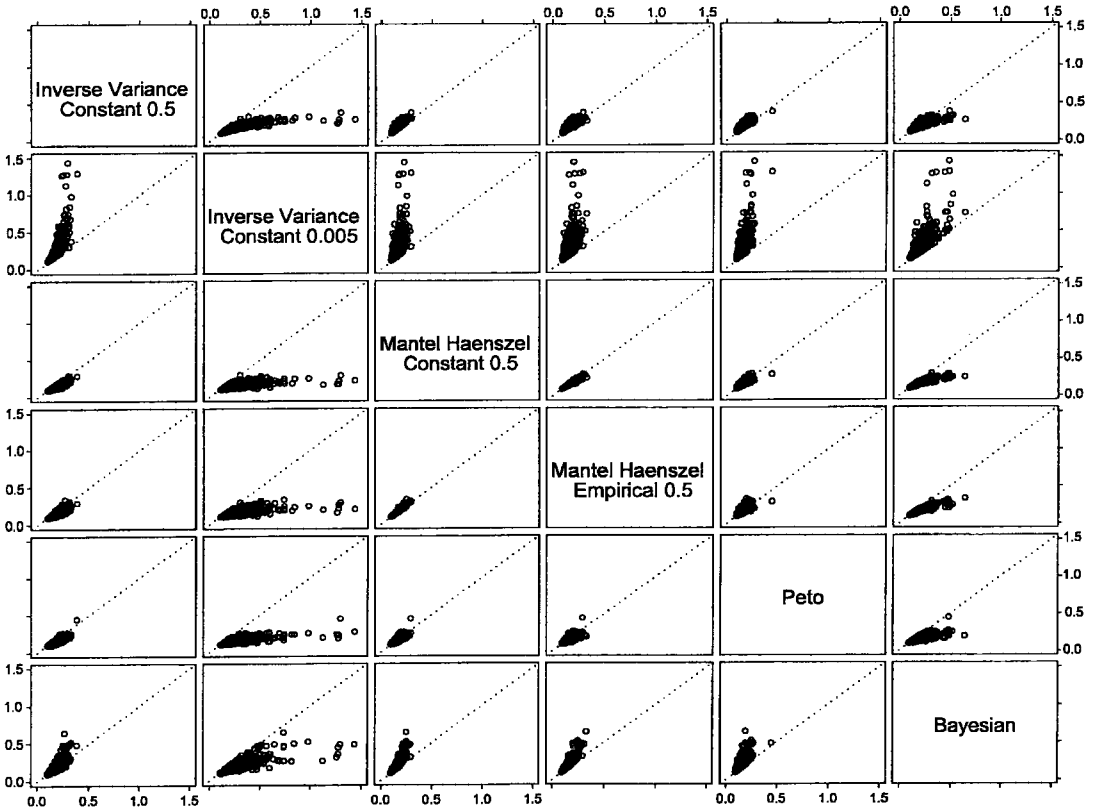

Figure 6. Scatter plot matrix showing relationship between the variance of the log-odds ratios of selected pooling methods for each of the 1000 simulated data sets.

can be seen to give high positive bias, underestimating the treatment effect. The inverse variance method is highly biased and performs the worst. The Peto summary estimate is moderately biased while the Bayesian and logistic regression methods produce similar biases as the Mantel–Haenszel pooled odds ratio.

Once more, when smaller continuity corrections are applied, the Mantel–Haenszel method performs quite well compared with the inverse-variance method, which performs worse. The empirical continuity correction, as in the balanced group scenario, outperforms the constant and 'treatment arm' corrections when using the Mantel–Haenszel method. The coverage statistics are good for the Bayesian model and most of the classical methods.

6.3. Group size ratio 1:4

When the treatment groups are unbalanced in a ratio of 1:4 the performance of the Peto and inverse variance methods worsens (Figure 4(c)). The Mantel–Haenszel method, using either a 'treatment arm' or empirical (sum to one) continuity correction, gives least biased results. The Mantel–Haenszel estimate using smaller continuity corrections, logistic regression and

the Bayesian fixed effect method again generally slightly over estimate the treatment effect but give good coverage. The inverse variance estimates are severely biased and the Mantel–Haenszel method using a continuity correction of $\frac{1}{2}$ also performs badly. Implementing a smaller correction factor improves the bias only with the Mantel–Haenszel constant correction method.

6.4. Group size ratio 1:8

Figure 4(d) shows the impact of severely unbalanced 1:8 group ratios on the classical and Bayesian summary estimates. The differences between the methods are now even more profound than the 1:4 ratio. All the Mantel–Haenszel estimates with the exception of using a continuity correction on $\frac{1}{2}$ plus the logistic regression method give good bias and coverage statistics. The Bayesian estimate gives results consistent with other group imbalances. The Peto summary odds ratio has worsened with the high group imbalance. It is clear that the inverse variance summary estimate and the Mantel–Haenszel with a continuity correction of $\frac{1}{2}$ estimate are ill-equipped to cope with sparse data situations with severely unbalanced groups.

6.5. Comparison across group allocation ratios

It is not possible to compare the group ratio results directly for each pooling method since, for example, the 1:2 imbalance studies contained 1.5 times as many patients as the 1:1 balanced group studies. As the sample size increases, one would expect the accuracy of the estimates to also increase. Despite this, the inverse variance method appears to be increasing in bias as the group imbalance increases. In contrast, the Mantel–Haenszel method gives decreasingly biased estimates as the group imbalance increases although this may be due to the increased sample sizes in the individual studies.

7. DISCUSSION

Researchers performing a meta-analysis involving sparse event dichotomous data face decisions concerning the summary measure and estimate to use. We have shown that the widespread use of the inverse variance-weighted summary odds ratio can give extremely biased results when event data is sparse and the group imbalance is high. Furthermore, when required, alternative corrections, such as the empirical and ‘treatment arm’ continuity corrections, have been shown to generally outperform the more usual constant continuity correction of $\frac{1}{2}$. The key points found in this research are presented below:

- The choice of fixed effect pooling method in a sparse event meta-analysis is important since certain estimators perform poorly with respect to bias and coverage; especially when group imbalances exist.
- The Peto method gives virtually no bias in a balanced group situation but the bias increases with the group imbalance. Confidence intervals for this method were too wide resulting in coverage higher than 95 per cent.
- The Mantel–Haenszel method performs consistently better than the inverse variance weighted method for all group imbalances and all continuity corrections applied. Perhaps

this is unsurprising since the Mantel–Haenszel method does not rely on a single-study statistic as a base.

- The Bayesian fixed effect model using vague priors (which does not require continuity correction factors) performs consistently well irrespective of group size imbalance, although is sometimes outperformed by certain Mantel–Haenszel/continuity correction combinations.
- Logistic regression is shown to perform similarly to the Mantel–Haenszel method with small continuity corrections. The summary odds ratio is generally unbiased with good coverage.
- The empirical and 'treatment arm' continuity corrections perform better than a constant correction factor for both the Mantel–Haenszel and inverse variance methods when the continuity corrections are constrained to sum to one. Therefore, it is recommended that either of the alternative continuity corrections should be used instead of $k = \frac{1}{2}$ despite these corrections not being available automatically in current software.
- The difference in performance of the empirical and 'treatment arm' continuity corrections varies depending on the group imbalance and pooling method used.

The main conclusions drawn in this project support many of the findings from previous literature [13, 16]. As reported by Deeks *et al.* [16], the Peto method is an unbiased method of pooling study results when the groups are balanced. Sankey *et al.* [13] conclude that the continuity correction (of $\frac{1}{2}$) should be employed over no continuity correction in nearly all circumstances. This work goes a step further in suggesting alternative correction factors that outperform adding a constant $\frac{1}{2}$.

Other methods do exist to combine odds ratios that have not been considered in the simulation study. Of particular interest are standard random effect methods, and those which are slightly more sophisticated which take into account the uncertainty inherent in the estimation of the between study variance from classical [34, 35] and Bayesian [36] perspectives. Classical and Bayesian random effects models were considered initially in this investigation, however this was abandoned for the following reasons. First, it was found that when data was estimated from simulation conditions very similar to those described in this paper, but with the between study variance set to be 0.3 (i.e. extreme heterogeneity with the underlying study effects varying between odds ratios of 0.17 and 1.49 when the underlying population mean odds ratio equals 0.5), the between study variance in each meta-analysis model was estimated to be non-zero less than 5 per cent of the time in a classical analysis. Consequently, the results for the random effects model were near identical to the inverse variance model with only fractionally improved performance which was still very poor and did not out perform most of the other fixed effect models, even under conditions of heterogeneity. This is consistent with the study of Sankey *et al.* [13] who report that 'the test for homogeneity … was much too conservative regardless of the method employed'. It should be noted that the application of the empirical continuity correction factor might not be applicable when using random effects models because the underlying treatment effect varies between studies.

Practical problems were encountered when fitting the Bayesian random effect model [36] in the WinBUGS software due to the sparseness of the data. This prevented routine analyses of 1000 data sets to be completed. Additionally it should be realized that intentionally 'vague' priors can be overly influential when data are sparse for variance component parameters as required in a random effect analysis [37], an area which requires further work.

Further, Decks *et al.* [16] considered fixed and random effects Poisson regression although their performance was reported not to challenge the more commonly used methods. Hence, while it would be interesting to examine the performance of random effect estimators under sparse data conditions, having carried out the simulation study described above, we believe that issues regarding heterogeneity are secondary when data are very sparse.

An exact fixed effect method utilizing network algorithms is now available to efficiently calculate exact confidence intervals for pooled odds ratios [38]. The computer package StatXact [39] implements such algorithms and provides exact analysis. We did not consider these here, but exact methods are well known to be conservative and Deeks reported [16] they did not outperform the more common pooling methods in their simulation study. Finally, a Bayesian analysis of the Mantel–Haenszel model has been proposed, however, its performance for sparse data is unknown [40].

Perhaps a more interesting extension to this study would be to look at variations on the methods proposed, for example, Gart and Zweifel [41] propose an alternative formula for the variance of a log-odds ratio which could be substituted in the inverse variance method which may well perform better than the estimate used here. By consideration of the standard variance formula for the variance of a log-odds ratio (Figure 1) it can be seen why methods based on this formula may do poorly when a small continuity correction is applied as inverse of the continuity correction factor is one component of this formula which will greatly inflate the variance. The alternative formula avoids this problem.

In addition to different methods of pooling, there are potentially infinite options for the exact continuity correction used, although we believe using the relatively 'large' and 'small' correction factor for each of three general approaches gives good representation. As previously mentioned, constants other than $\frac{1}{2}$ have been suggested by other authors; for example $\frac{1}{6}$ by Mosteller and Tukey [11] and a range of different values by Agresti [10]. We certainly condone the idea of using a number of continuity corrections as a form of sensitivity analysis. Further, we chose not to add continuity correction factors to studies with zero events in both groups, and only add correction factors where necessary (i.e. one group contained zero events). While we do justify the first of these decisions, we did not examine the performance of adding correction factors to all studies (or to all where the number of events was less than a certain threshold e.g. 5). While this would be interesting, our intuition is that unnecessary correction factors would only artificially inflate the amount of information and hence produce confidence intervals that are too narrow. An alternative formulation of the empirical continuity correction presented in this paper would be to put it into a more formal empirical Bayes framework.

It should be realized that the absolute levels of bias and coverage observed in these simulation studies are not directly applicable to other scenarios. For example, a constant correction of $\frac{1}{2}$ applied to unbalanced groups with ratio 1:4 gives an odds ratio of 4 before any events are observed. When the true underlying odds ratio is 0.5, this may have an effect of biasing estimates towards one (i.e. underestimating the true treatment effect). The same continuity correction applied to data with an underlying odds ratio of 2 may however bias results away from an odds ratio of one (i.e. overestimating the true treatment effect). Therefore, the bias obtained from the simulations in this project will be influenced by the specified underlying odds ratio. We did re-run the simulations changing the underlying odds ratio to 2 and $\frac{1}{4}$ and while absolute levels of bias and coverage varied the findings were qualitatively consistent with those presented in this paper (data available from the authors on request).

The odds ratio has been the only outcome measure examined in this paper. Other outcome measures, such as the risk difference and risk ratio, also exist for pooling dichotomous data in meta-analysis. The issue of deciding which outcome to use in a particular application is non-trivial [42] and it should not be forgotten that the relative weightings given to the individual studies in a meta-analysis may change considerably between outcomes. In a sparse data scenario this is particularly true, and a total zero event study that would be excluded on a relative scale would be included, and even given most weight on a risk difference scale [43]. Hence, investigation of the performance of methods to pool studies using other outcome measures and even a comparison between outcome measures is required.

To conclude, the work presented here suggests that the standard continuity correction factor of a fixed $\frac{1}{2}$ used routinely (and implemented in most software) can be improved upon. While some methods/continuity correction approaches are clearly inferior, there is no one superior method in all situations and therefore a sensitivity analysis across a number of correction factors is suggested. It is recommended that researchers attempting to perform a meta-analysis with sparse dichotomous outcome data should carefully consider and clearly report the choice of outcome measure(s), pooling method(s) and possible continuity correction(s) used since all may have an influence on the conclusions drawn. We also recommend that meta-analysis software permits the user to have control of continuity corrections used.

ACKNOWLEDGEMENTS

The authors wish to thank Professors John Thompson, David Jones and Keith Abrams for valuable discussions that improved this paper and two anonymous referees whose comments improved this paper considerably. Part of this work was carried out when Michael Sweeting was funded by AstraZeneca on the MSc in Medical Statistics at the University of Leicester.

REFERENCES

1. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. *Methods for Meta-Analysis in Medical Research*. Wiley: London, 2000.
2. Egger M, Davey Smith G, Altman DG. *Systematic Reviews in Health Care: Meta-Analysis in Context*. BMJ Books: London, 2000.
3. Sutton AJ, Cooper NJ, Lambert PC, Jones DR, Abrams KR, Sweeting MJ. Meta-analysis of rare and adverse event data. *Expert Review in Pharmacoeconomics Outcomes Research* 2002; **2**:367–379.
4. Hemminki E, McPherson K. Impact of postmenopausal hormone therapy on cardiovascular events and cancer: pooled data from clinical trials. *British Medical Journal* 1997; **315**:149–153.
5. Austin H, Perkins LL, Martin DO. Estimating a relative risk across sparse case-control and follow-up studies: a method for meta-analysis. *Statistics in Medicine* 1997; **16**:1005–1015.
6. Perkins LL, Clark BD, Klein PJ, Cook RR. A meta-analysis of breast implants and connective tissue disease. *Annals of Plastic Surgery* 1995; **35**:561–570.
7. Yates F. Contingency tables involving small numbers and the chi-squared test. *Journal of the Royal Statistical Society* 1934; **1**(Supplement):217–235.
8. Cox DR. The continuity correction. *Biometrika* 1970; **57**:217–219.
9. Plackett RL. The continuity correction in 2×2 tables. *Biometrica* 1964; **51**:327–337.
10. Agresti A. *An Introduction to Categorical Data Analysis*. Wiley: New York, 1996.
11. Mosteller F, Tukey JW. Data analysis and regression. *A Second Course in Statistics*. Addison-Wesley Publishing Company: Philippines, 1977.
12. Fienberg SE. Preliminary graphical analysis and quasi-independence for two-way contingency tables. *Journal of the Royal Statistical Society. Applied Statistics* 1969; **18**:153–168.
13. Sankey SS, Weissfeld LA, Fine MJ, Kapoor W. An assessment of the use of the continuity correction for sparse data in metaanalysis. *Communications in Statistics-Simulation and Computation* 1996; **25**:1031–1056.
14. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986; **7**:177–188.

15. Bow EJ, Laverdiere M, Lussier N, Rotstein C, Cheang MS, loannou S. Antifungal prophylaxis for severely neutropenic chemotherapy recipients. A meta-analysis of randomized-controlled clinical trials. *Cancer* 2002; **94**:3230–3246.
16. Deeks J, Bradburn M, Localio R, Berlin J. Much ado about nothing: statistical methods for meta-analysis with rare events. *Presented at Systematic reviews*: *Beyond the basics*, *January*, *Oxford* (*Abstract available at http://www.ihs.ox.ac.uk/csm/talks.html#p23.*) 1999.
17. Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research* 1993; **2**:173–192.
18. Berlin JA, Laird NM, Sacks HS, Chalmers TC. A comparison of statistical methods for combining event rates from clinical trials. *Statistics in Medicine* 1989; **8**:141–151.
19. Fleiss JL. Measures of effect size for categorical data. In *The Handbook of Research Synthesis*, Cooper H, Hedges LV (eds). Russell Sage Foundation: New York, 1994; 245–260.
20. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute* 1959; **22**:719–748.
21. Robins J, Greenland S, Breslow NE. A general estimator for the variance of the Mantel-Haenszel odds ratio. *American Journal of Epidemiology* 1986; **124**:719–723.
22. Emerson JD. Combining estimates of the odds ratio: the state of the art. *Statistical Methods in Medical Research* 1994; **3**:157–178.
23. Yusuf S, Peto R, Lewis J, Collins R, Sleight P *et al*. Beta blockade during and after myocardial infarction: an overview of the randomised trials. *Progress in Cardiovascular Diseases* 1985; **27**:335–371.
24. Greenland S, Salvan A. Bias in the one-step method for pooling study results. *Statistics in Medicine* 1990; **9**:247–252.
25. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* 2001; **10**:277–303.
26. Spiegelhalter DJ, Thomas A, Best NG. WinBUGS Version 1.2. User Manual. MRC Biostatistics Unit: Cambridge, 2000.
27. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomised clinical trials. *Statistics in Medicine* 1991; **10**:1665–1677.
28. Cook DJ, Witt LG, Cook RJ, Guyatt GH. Stress ulcer prophylaxis in the critically ill: a meta-analysis. *The American Journal of Medicine* 1991; **91**:519–527.
29. Cox DR. *The Analysis of Binary Data*. Methuen & Co Ltd: London, 1970.
30. Simpson EH. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society Series B* 1951; **13**:238–241.
31. Altman DG, Deeks JJ. Meta-analysis, Simpson's paradox, and the number needed to treat. *BMC Medical Research Methodology* 2002; **2**:(available at: http://www.biomedcentral.com/1471-2288/2/3).
32. Hornbuckle J, Vail A, Abrams KR, Thornton JG. Bayesian interpretation of trials: the example of intrapartum electronic fetal heart rate monitoring. *British Journal of Obstetrics and Gynaecology* 2000; **107**:3–10.
33. S-Plus. (4.0). 97. Statistical Sciences, Inc, 1700 Westlake Ave. North, Suite 500, Seattle, WA 98109.
34. Biggerstaff BJ, Tweedie RL. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Statistics in Medicine* 1997; **16**:753–768.
35. Hardy RJ, Thompson SG. A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* 1996; **15**:619–629.
36. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine* 1995; **14**:2685–2699.
37. Lambert P, Sutton AJ, Jones DR, Abrams KR. How vague is vague? *Presented at the 23rd Annual Conference of International Society for Clinical Biostatistics*, Dijon, France, September 2002.
38. Mehta CR, Patel NR, Grey R. Computing an exact confidence interval for the common odds ratio in several 2×2 contingency tables. *Journal of the American Statistical Association* 1997; **80**:969–973.
39. StatXact. Cytel Software Corporation, Cambridge, MA, 2000.
40. Leonard T, Duffy JC. A Bayesian fixed effects analysis of the Mantel-Haenszel model applied to meta-analysis. *Statistics in Medicine* 2002; **21**:2295–2312.
41. Gart JJ, Zweifel JR. On the bias of various estimators of the logit and its variance, with application to quantal bioassay. *Biometrica* 1967; **54**:471–475.
42. Deeks JJ, Altman DG. Effect measures for meta-analysis of trials with binary outcomes. In *Systematic Reviews in Health Care. Meta-Analysis in Context*, Egger M, Davey Smith G, Altman DG (eds). BMJ Publishing Group: London, 2001.
43. Tang JL. Weighting bias in meta-analysis of binary outcomes. *Journal of Clinical Epidemiology* 2000; **53**: 1130–1136.

