© International Epidemiological Association 2002 Printed in Great Britain *International Journal of Epidemiology* 2002;**31**:96–104

Being sceptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction

Julian PT Higgins and David J Spiegelhalter

Background	There has been extensive discussion of the apparent conflict between meta-analyses and a mega-trial investigating the benefits of intravenous magnesium following myocardial infarction, in which the early trial results have been said to be 'too good to be true'.
Methods	We apply Bayesian methods of meta-analysis to the trials available before and after the publication of the ISIS-4 results. We show how scepticism can be formally incorporated into an analysis as a Bayesian prior distribution, and how Bayesian meta-analysis models allow appropriate exploration of hypotheses that the treatment effect depends on the size of the trial or the risk in the control group.
Results	Adoption of a sceptical prior would have led early enthusiasm for magnesium to be suitably tempered, but only if combined with a random effects meta-analysis, rather than the fixed effect analysis that was actually conducted.
Conclusions	We argue that neither a fixed effect nor a random effects analysis is appropriate when the mega-trial is included. The Bayesian framework provides many possibilities for flexible exploration of clinical hypotheses, but there can be considerable sensitivity to apparently innocuous assumptions.
Keywords	Bayes theorem, meta-analysis, magnesium sulphate, myocardial infarction
Accepted	6 September 2001

A common phenomenon in medical research concerns an intervention which, on the basis of early experiments, appears to be of potentially great benefit, but those high hopes are dashed by subsequent investigations. The case of magnesium and myocardial infarction (MI) is a classic example, where early enthusiasm for magnesium based on meta-analysis was tempered by the results of the ISIS-4 mega-trial. One theme in the continuing debate on the methodologies of meta-analysis and mega-trials is the interpretation of unusually strong treatment effects, in which the need for a degree of scepticism has been emphasized. A number of authors[1–7] have suggested that this can be formally accommodated within a Bayesian approach, in which a *prior* distribution expresses the belief that large treatment effects are unlikely. By combining this prior with the evidence from a trial, analysts can formally allow for the common observation that results are 'too good to be true'. Bayesian methods are attracting increasing interest in evaluation of health care interventions and have recently been reviewed.[8]

MRC Biostatistics Unit, Cambridge, UK

Correspondence to: Dr Higgins, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK. E-mail: julian.higgins@mrc-bsu.cam.ac.uk

Magnesium in Myocardial Infarction

Intravenous magnesium has long been believed to play an important role in patients with acute MI. Physiological studies in animals and humans, and epidemiological studies in humans have suggested a protective effect, particularly through preventing serious arrhythmias.[9–11] We briefly review the history of the magnesium trials and meta-analyses in Table 1. There have been many responses to the apparent contradiction between the meta-analysis and the mega-trial, which can be summarized under the following four broad headings.

Essential scepticism about large effects

In response to the ISIS-4 results, Yusuf (the main author of the optimistic *Circulation* editorial) and Flather claimed 'since most treatments produce either no effect or at least moderate effects on major outcomes such as mortality, investigators should be sceptical if the results obtained deviate substantially from this expectation'.[17] Yusuf later added that 'if one assumed that only

Table 1 Brief summary of landmark publications from trials and meta-analyses in magnesium for acute myocardial infarction

Year	Publication	Finding concerning magnesium treatment	Conclusion
1981	First trial report[12]	Trend toward smaller infarct size (43, later 76, patients)	'Magnesium … is promising and deserves further study'[13]
1991	Meta-analysis by Teo *et al.*[10]	55% reduction in odds of mortality (8 trials)	'further large scale trials to confirm (or refute) the findings are desirable'
1992	LIMIT-2 large trial[14]	24% reduction in mortality	'a simple, safe and widely applicable treatment'
1993	*Circulation* editorial[15]	'An effective, safe, simple and inexpensive treatment'	Recommends further trials to obtain 'a more precise estimate of the mortality benefit'
1995	ISIS-4 mega-trial[16]	Non-significant adverse mortality (58 050 patients)	'Overall, there does not now seem to be any good clinical trial evidence for the routine use of magnesium'

moderate sized effects were possible, the apparent large effects observed in the meta-analyses of small trials with magnesium ... should perhaps have been tempered by this general judgement. If a result appears too good to be true, it probably is'.[18] This expression of prior scepticism was echoed by Peto and colleagues,[19] who argued that the risk reduction of the initial overview was 'implausibly large', and that even when combined with the LIMIT-2 data 'still indicated an implausibly large reduction of one-third in mortality'. However, Peto reports that the ISIS-4 steering committee were convinced there would be at least some benefit, right up until they were shown the results.

Criticism of the meta-analysis

A number of authors remark that the meta-analyses may be missing studies.[15,17,20–23] Egger and Davey Smith claimed that 'selective non-publication of negative trials ... seems to be a likely explanation for the discrepant findings' since funnel plots (of treatment effect estimates against sample size or precision) are distinctly asymmetrical,[21] and they pointed out that 'if a thorough sensitivity analysis had been conducted earlier it would have become clear that the findings were less robust than suggested'.[22] Using a different argument, Pogue and Yusuf[24,25] use a frequentist stopping rule applied to the meta-analysis, designed to have high power to detect a moderate effect (15% reduction in mortality), and find 'a lack of conclusive evidence of the benefit of magnesium provided by the meta-analysis. Therefore the apparently conflicting evidence from ISIS-4 should not have been completely unexpected'. This conclusion echoes criticisms that the meta-analysis was too small to come to a reliable conclusion,[17,26] being based on only 78 events before LIMIT-2 and 286 afterwards. Antman[27–29] has discussed the choice of statistical method used for the meta-analyses, pointing out that a random effects model gives materially different results from the fixed effect model used by Teo *et al.*,[10] Teo and Yusuf[20] and in the ISIS report.[16] 'An inappropriate statistical model ... was used in combining the results of the ISIS-4 study with those of the studies reporting favourable results'.[29]

Criticism of the mega-trial

Woods[30] argued that in a mega-trial such as ISIS-4 'null bias will arise when the contrast between treatment and no-treatment, or between subgroups, is blunted either by non-protocol therapy or by inaccuracy of data'. In addition, he claims that magnesium's benefit is to prevent reperfusion injury, and yet the ISIS-4 protocol expected all patients to be given thrombolytic therapy (which tends to induce reperfusion) before randomization, and hence 'the study was not designed to test for an effect of magnesium on reperfusion injury'. Other authors[31–33] have echoed the view that in ISIS-4 'there is a distinct possibility that magnesium was administered too late'.[33] Woods claims the subgroup who did not receive thrombolytic therapy had a statistical power 'too low to detect reliably even a treatment benefit as large as that of thrombolytic therapy', and Borzak and Ridker point out that 'an important limitation of the ISIS-4 protocol was that the actual time of magnesium initiation was unknown'.[34]

Clinical explanations

Several clinical explanations for the discrepant results have been proposed,[27,29,31,32,35,36] including '(1) time of initiation of magnesium treatment after acute MI and thrombolytic therapy; (2) dosage of magnesium in the first 24 hours after acute MI; (3) duration of post-acute MI magnesium infusion and (4) differences in patient risks'.[31] Shechter *et al.* note that 'the low mortality rate in the control group in ISIS-4, the late enrolment of patients, ... and the fact that magnesium infusions were delayed 1–2 hours after thrombolytic therapy, suggest that it is possible that the majority of patients in ISIS-4 were at low risk of mortality... The lack of therapeutic effect of magnesium in ISIS-4 fully accords with the results of various experimental models'.[32] Antman has concluded, retrospectively, that no effect was to be expected in ISIS-4 were it to have the low event rate that was indeed observed.[33,37]

Objectives

Our principal aim is to investigate how a Bayesian perspective might have influenced the interpretation of the published evidence on intravenous magnesium in acute MI. We address the following fundamental questions:

(1) What degree of 'scepticism' would have been necessary in 1993 not to be convinced by the meta-analysis reported in Yusuf and colleagues?[15]
(2) How might a Bayesian analysis attempt to reconcile the results of the studies in 1995, and is there evidence that the treatment effect depends on the study size or underlying risk?

Methods

Before ISIS-4: A sceptic's view of the trials in 1993

For our first perspective on the magnesium trials we look at the data available in 1993 when the enthusiastic *Circulation* editorial by Yusuf and colleagues was published.[15] We address the first eight trials, listed at the top of Table 2, comprising the trials in the 1991 meta-analysis[10] together with the LIMIT-2 trial.[14] Our aim is to emulate the Peto fixed effect meta-analysis of these trials, but to incorporate prior scepticism through a prior distribution on the common odds ratio.

We base a sceptical *a priori* position on Pogue and Yusuf: 'Most clinically important interventions are likely to reduce the relative risk of major outcomes, such as myocardial infarction, stroke, or death, by about 10–20%'.[25] We might then consider that a reasonable degree of scepticism is to think it unlikely (only 5% chance) that magnesium would reduce the odds of mortality by more than 25%. In section (2) of the Appendix we show how this can be translated into a normal prior distribution for the common log(odds ratio), centred on 0 and with a variance of 0.03, and that this precision is equivalent to a 'trial' with 72 deaths in each group. This compares with 286 deaths actually observed. To address objective (1), we use a similar technique to design a prior distribution to produce a posterior distribution with 5% chance that there is no benefit from magnesium.

We compare the fixed effect meta-analyses with random effects meta-analyses, again based on the Peto approach. The conventional random effects model corresponds to an assumption that the treatment effects arise from a normal distribution. In Bayesian terminology this reflects an assumption of *exchangeability* about the treatment effects, which we address in more detail in the Discussion. Technical details of the analysis are presented in the Appendix (section (1), Method (i)).

Post ISIS-4: Reconciling the trial findings in 1995

For our second perspective we look at the data available towards the end of 1995. The results of the ISIS-4 mega-trial were published in this year, along with updated meta-analyses that identified a further six small trials. All 15 trials are listed in Table 2. Further trials have been published since 1995, or are underway as we write (e.g. the MAGIC trial[42]). However, these have not yet been systematically identified for inclusion in a systematic review, although a Cochrane review is currently in preparation that will bring the evidence up to date.[43] For analyses from this perspective we use a different approach to the analysis that explicitly models the binomial nature of the data and does not suffer from a bias inherent in the Peto approach.[44,45] Details are again provided in the Appendix (section (1), Method (ii)).

Egger and Davey Smith[21,22] claim that a potential problem with the initial meta-analysis[10] is bias associated with the smaller negative trials. They apply a test for asymmetry of a funnel plot which is essentially a regression of the log(odds ratio) on its estimated standard error.[23,46] This regression test suffers from inherent correlation between the dependent and the independent variable; an alternative approach is to take sample size as the independent variable.[47] Here we choose the log(sample size), due to the substantial variation in sample sizes between the smaller trials and the mega-trial.

We also investigate the dependence of treatment effect on underlying risk (as measured by the mortality risk in the control group). Naive examinations of whether benefit of treatment

Table 2 Summary data from 15 randomized trials of intravenous magnesium for acute myocardial infarction (data from Sterne *et al.*[38]) with classical meta-analyses using the Peto method[39] and DerSimonian and Laird (D-L) method[40] based on sample log(odds ratios) and a moment estimate of between-trial variance. A test of homogeneity of odds ratios (p_{het}) is as described by Woolf[41]

	Magnesium group		Control group	
Trial	Deaths r_i^M	Patients n_i^M	Deaths r_i^C	Patients n_i^C
Morton	1	40	2	36
Rasmussen	9	135	23	135
Smith	2	200	7	200
Abraham	1	48	1	46
Feldstedt	10	150	8	148
Shechter 1990	1	59	9	56
Ceremuzynski	1	25	3	23
LIMIT-2	90	1159	118	1157
Fixed effect (Peto) meta-analysis of above eight trials: OR = 0.65 (95% CI : 0.51, 0.82); p_{het} = 0.19 Random effects (D-L) meta-analysis of above eight trials: OR = 0.55 (95% CI : 0.34, 0.89)				
Bertschat	0	22	1	21
Singh	6	76	11	75
Pereira	1	27	7	27
Golf	5	23	13	33
Thogersen	4	130	8	122
Shechter 1995	4	107	17	108
Fixed effect (Peto) meta-analysis of above 14 trials: OR = 0.57 (95% CI : 0.46, 0.71); p_{het} = 0.19 Random effects (D-L) meta-analysis of above 14 trials: OR = 0.47 (95% CI : 0.33, 0.68)				
ISIS-4	2216	29 011	2103	29 039
Fixed effect (Peto) meta-analysis of above 15 trials: OR = 1.01 (95% CI : 0.95, 1.07); p_{het} = 0.0002 Random effects (D-L) meta-analysis of above 15 trials: OR = 0.53 (95% CI : 0.36, 0.77)				

depends on underlying risk can detect spurious relationships due to natural correlation between observed risks and treatment effects. Methodology has been developed that overcomes this problem.[48–51] We choose a Bayesian approach that relates the underlying log(odds ratio) to the underlying logit(risk) in the control group.[50]

We perform all analyses using WinBUGS,[52] in which all the above models are readily programmed. Code is presented in the Appendix (section 3).

Results

Before ISIS-4: A sceptic's view of the trials in 1993

Figure 1(a) shows the prior, likelihood and posterior for a Bayesian meta-analysis using our sceptical prior distribution. Table 3 (row (a)) demonstrates that, with this degree of scepticism, the analysis finds the evidence quite convincing concerning a clinically worthwhile improvement, in that there is now a 97% probability that the treatment benefit is at least 10%.

How sceptical would one need to have been in order not to have been convinced by the results? The required prior distribution would be the equivalent of having observed 515 deaths in each group prior to seeing the trials (Figure 1(b), Table 3, row (b)). Technically, the prior distribution is $N(0,0.0655^2)$ for the common log(odds ratio). The prior necessary not to have found the meta-analysis 'significant', even at a one-sided 5% probability, is clearly an unreasonably extreme form of scepticism. We therefore can reject Yusuf and Flather's claim that a reasonably sceptical approach applied to their analysis would have led to caution.

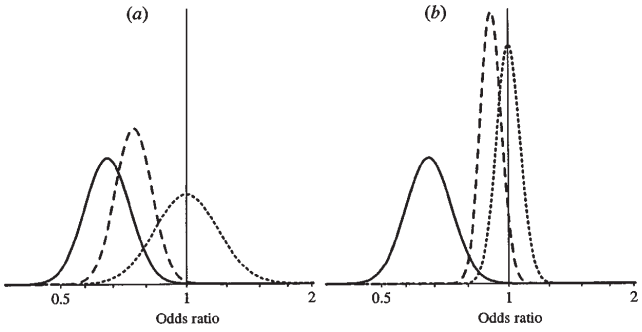

Figure 1 Prior (dotted line), likelihood (solid line) and posterior distributions (dashed line) for Peto-style meta-analyses of the first eight magnesium trials using two sceptical prior distributions. (a) A 'reasonable' expression of scepticism (equivalent to a 'trial' with 72 deaths in each group). (b) The scepticism necessary not to have found the meta-analysis 'significant' (equivalent to a 'trial' with 515 deaths in each group)

Table 3 Posterior probabilities of absolute and clinical superiority of magnesium, given two levels of sceptical prior (fixed effect model based on Peto method)

	Magnesium superior p(odds ratio <1)	Magnesium clinically superior p(odds ratio <0.9)
(a) Reasonably sceptical (72 in each group)	99.8%	97%
(b) Very sceptical (515 in each group)	95%	44%

Random effects meta-analysis of the first eight trials leads to a different conclusion. Figure 2 illustrates a Bayesian random effects analysis using the reasonably sceptical prior distribution. The 95% credibility interval for the central overall mortality odds ratio includes 1, hence yielding the cautious result sought by Yusuf. The Figure illustrates the consequences of a Bayesian analysis on the estimates from the individual trials. The LIMIT-2 results are hardly changed, whereas the smaller studies are 'shrunk' towards the cautious overall conclusion. A random effects Bayesian meta-analysis using a 'flat' reference prior distribution is also illustrated. This emulates the traditional non-Bayesian random effects meta-analysis, and gives a significant result in favour of magnesium.

Post ISIS-4: Reconciling the trial findings in 1995

In Figure 3 we illustrate the impact of adding ISIS-4 to fixed effect and random effects meta-analyses of the first 14 trials. For the fixed effect analyses there is little material difference between using a flat reference prior and using the reasonably sceptical prior, such is the overwhelming influence of ISIS-4. The mega-trial has much less impact on random effects analyses.

Does effect depend on trial size?

Table 4 shows results of meta-regressions on log(sample size) using data from the 14 smaller studies and from all 15 studies

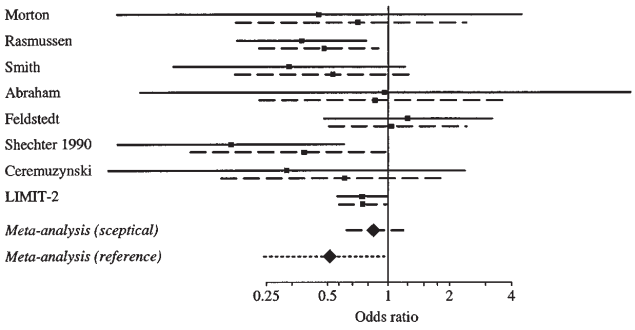

Figure 2 Random effects meta-analysis of the first eight trials of magnesium for myocardial infarct using a reasonably sceptical prior. For each study the plot shows the original point estimate (solid line) and the shrunk estimate (dashed line), both with 95% CI. A random effects meta-analysis using a 'flat' reference prior distribution is included for comparison

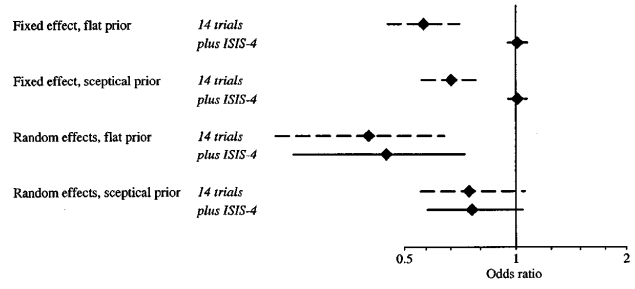

Figure 3 The impact of adding ISIS-4 to different meta-analyses of the 14 smaller studies (including LIMIT-2). 95% posterior credible intervals for the mortality odds ratio associated with magnesium, for both fixed effect and random effects analysis, with a 'flat' reference prior and the reasonably sceptical prior (5% chance of at least a 25% reduction in mortality odds)

Table 4 Estimate and 95% CI for the influence of increased sample size and underlying risk on the mortality odds ratio for magnesium treatment. A negative slope corresponds to the treatment effect becoming smaller as sample size decreases or underlying risk declines, with the treatment effect becoming zero when the covariate value is as stated in the final column

	Slope	95% CI	Covariate value for average null result
Log(sample size)			**Sample size**
Without ISIS-4	0.33	(0.02, 0.75)	4900 (520, 140 million)
With ISIS-4	0.21	(0.078, 0.38)	31 000 (3200, 10 million)
Logit(underlying risk)			**Underlying risk**
Without ISIS-4	–0.36	(–0.93, 0.34)	1.4% (0–100%)
With ISIS-4	–0.45	(–1.00, 0.17)	2.2% (0–100%)

including ISIS-4. Minimally informative reference priors are used for both analyses. The estimated slopes are positive with 95% credible intervals excluding 0, providing evidence that smaller studies had more extreme results. We also estimate the sample size at which, according to the regression equation, the average treatment effect would be an odds ratio of 1. Before ISIS-4, it might have been expected that studies of 4900 participants would produce a null result *on average*. The predicted odds ratio from this regression for a trial of 58 050 participants (the size of ISIS-4) is 2.2 with a very wide 95% interval from 0.27 to 32. The relationship is even clearer once ISIS-4 is included in the meta-regression. This analysis is illustrated in Figure 4, along with prediction intervals for trials of different sizes.

Relationship to underlying risk

Figure 5 shows the apparent relationship between observed treatment effect and underlying risk. The fitted regression line provides the results shown in Table 4, with and without the ISIS-4 data, and the relationship among all 15 trials is illustrated in Figure 5. A relationship with underlying risk is suggested before and after inclusion of ISIS-4, but with a wide interval. Before ISIS-4 the model would have predicted that the treatment would not be effective with an underlying risk below 1.4%. The underlying risk in the ISIS-4 trial was 7.2%. Our conclusions differ from those of Antman *et al.*, who predicted a null effect at a risk of around 7%. There are three differences between our analysis and Antman's. First, the data are different; second, ours is based on the binomial model whereas his is based on summary data; and third, Antman assumes a random effect for the control group logit(risks), whereas we estimate them separately within each trial. We note from sensitivity analyses that the results of this underlying risk analysis can be highly dependent on specific model assumptions (Appendix, section 4).

Discussion

We have used a Bayesian approach to examine how prior scepticism might have dampened the initial enthusiasm for magnesium based on a meta-analysis of the first eight trials. Incorporating any reasonably sceptical prior distribution in the fixed effect meta-analysis of the first eight trials would not have changed the conclusions. Random effects meta-analyses would have led to more caution.

The results of a meta-analysis of all 15 trials, including ISIS-4, depend crucially on the model chosen for the analysis. What is an appropriate model for analysis of these trials? Initial meta-analyses including the ISIS-4 data used a fixed effect approach,

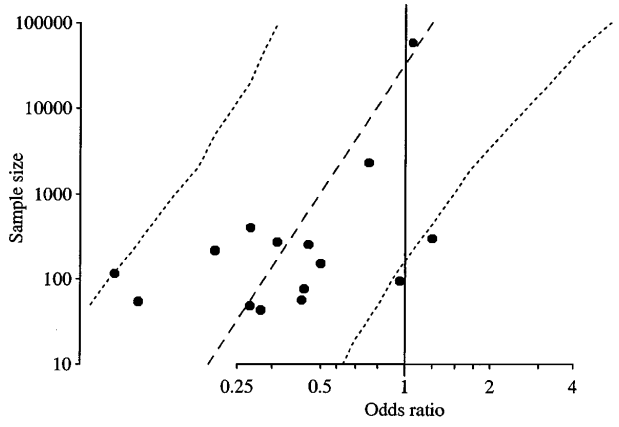

Figure 4 Funnel plot illustrating the relationship between mortality odds ratio and sample size for 15 trials including ISIS-4. The meta-regression line is plotted together with a predictive distribution around the regression line

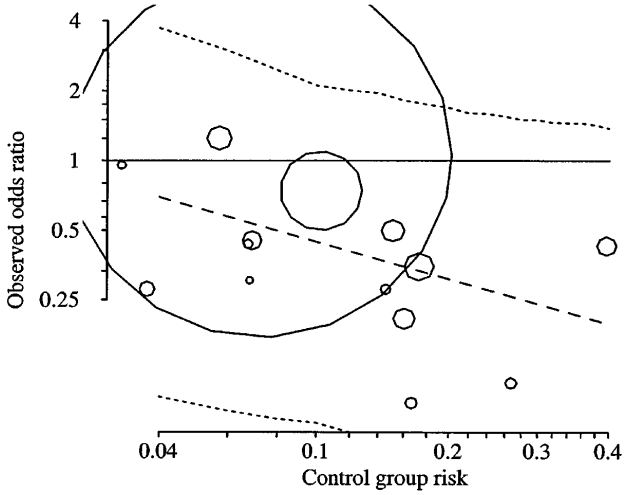

Figure 5 Relationship between mortality odds ratio and underlying risk (event rate in the control group). All 15 trials are included, with the size of each point proportional to the precision of estimation (so the point corresponding to ISIS-4 is very large). The underlying risk meta-regression line is plotted together with a predictive distribution around the regression line

based on the argument that it provides a valid test of the null hypothesis of no effect across all trials. However, estimates of effect, and particularly confidence intervals, cannot easily be interpreted from an inappropriate fixed effect analysis. A fixed effect model assumes a common odds ratio underlies each and every trial. The lower limit of a 95% confidence interval from ISIS-4 alone lies at 1, which had probability 0.002 of being exceeded in a fixed effect analysis of the early trials, even using the reasonably sceptical prior (Table 3, row (a)). Thus the fixed effect assumption would appear to be unjustified.

Exchangeability

The Bayesian justification of the random effects model reflects an *a priori* assumption of *exchangeability* about the treatment effects (odds ratios) underlying the trials, i.e. the joint distribution of the treatment effects is independent of the identity of the actual trials being considered.[53] In practice, the exchangeability assumption involves two components. First, that the odds ratios are unlikely to be identical, but are likely to be similar (such that it makes sense to combine them in a meta-analysis). Second, that there is no reason to expect the odds ratio in any specified trial to be larger than the odds ratio in another specified trial. The second component has the consequence that an *a priori* ranking of the effect sizes is not possible.

Would ISIS-4 have been considered exchangeable with the 14 smaller trials before the results were announced? Peto reports how 'the ISIS-4 steering committee was sufficiently sceptical to want large-scale randomized evidence. They knew that there might well be a negligible benefit, or even a small net hazard'.[19] When asked what size of effect they expected to see, the median 'best guess' was a reduction in mortality from 8% to 7%, a considerably smaller effect than had been observed in the meta-analyses. Wood's retrospective remarks about null bias in mega-trials may have led him also to expect a smaller effect in ISIS-4.

A by-product of a random effects meta-analysis is that it will down-weight large studies when there is evidence of heterogeneity. The effect is considerable with ISIS-4. The mega-trial has much less impact on a random effects analyses than it does on a fixed effect analysis (Figure 3). One who believed the exchangeability assumption would consider the genuine effect particular to the ISIS-4 protocol to be near the line of no effect by chance. With this view, it would be reasonable to pay less attention to the size of ISIS-4, and more to the variation across the studies, which is precisely what the random effects analysis does. It is interesting that the consequence of this view coincides with Woods' comment on ISIS-4: 'Where [non-trial] treatments replicate the effect of the intervention under test, a bias will arise ... Trial size is irrelevant to the problem of bias, except to give unjustified authority to the result'.[30]

We believe the weight of evidence is that an exchangeability assumption does not seem reasonable for the 15 magnesium trials, on the basis that the effect underlying a mega-trial may be, compared with a small trial, biased towards the null as argued by Wood. Furthermore, if a multi-centre mega-trial like ISIS-4 is viewed as a collection of separate trials in different centres, prior uncertainty regarding a pooled effect across the centres would not be the same as for a single small trial like the early magnesium trials. Since we are unwilling to assume exchangeability, we conclude that neither a fixed effect nor a simple random effects meta-analysis is appropriate. The natural progression is to examine reasons for the discrepant results,[54] which led us to address two potentially important reasons that had been raised in previous discussions of the magnesium trials: sample size and underlying risk. Inclusion of important trial-specific covariates may well make more tenable an assumption of exchangeability of residual effects.

Concluding Remarks

Our specific findings regarding magnesium and MI can be summarized as follows. First, if one had carried out a standard fixed effect analysis, one would have needed to be unreasonably sceptical not to have found the 1993 meta-analysis convincing. Second, reasonable scepticism and a random effects meta-analysis would have led to appropriate caution. Third, neither a standard fixed effect nor a random effects model is suitable for the entire group of trials including ISIS-4. Fourth, there is evidence of dependence of the treatment effect on trial size, and some suggestion of a dependence on underlying risk.

The Bayesian analysis using minimally informative priors produces similar conclusions to a traditional analysis, but the additional possibility of informative prior input allows clinical judgements to be formally incorporated and the sensitivity of the conclusions to those beliefs explored within a coherent framework. In addition, Bayesian analysis can investigate why individuals with different opinions may interpret data differently. Nevertheless, it should be emphasized that with limited numbers of studies and events, there can be considerable sensitivity to assumptions concerning 'secondary' aspects of the model, such as whether the control group logit(risks) are assumed independent or exchangeable, whether uniform priors are placed on the control group risks themselves or their logits, what prior is placed on the between-trial variance, the normality assumption for the random effects, and so on. Modern Bayesian computational methods allow all these options to be explored, but there is still a dearth of experienced guidance on appropriate models.

Finally, is it possible to draw some general conclusions from such a retrospective case study? A primary lesson must be that scepticism is a reasonable position and can be formally examined. In addition, the assumption of exchangeability between studies is a strong one and requires careful examination. Indeed, it raises the question—what is a 'study'? We could always break a large study into smaller ones to add weight: for example, ISIS-4 had 31 countries and 1086 hospitals, and it would be illuminating to investigate the heterogeneity between these centres in a structured way. It is vital to explore possible reasons for heterogeneity between trials, and a standard random effects analysis may be quite inappropriate.

Acknowledgements

An earlier version of this analysis was funded by the NHS R&D Health Technology Assessment Methodology programme, grant number 93/50/5. JH was supported by MRC grant G9815466.

References

[1] Kass RE, Greenhouse JB. Comments on 'Investigating therapies of potentially great benefit: ECMO' by JH Ware. *Statist Sci* 1989;**4:**310–17.

[2] Fletcher A, Spiegelhalter D, Staessen J, Thijs L, Bulpitt C. Implications for trials in progress of publication of positive results. *Lancet* 1993; **342:**653–57.

[3] Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomised trials (with discussion). *J R Statist Soc (A)* 1994;**157:**357–87.

[4] Fayers PM, Ashby D, Parmar MKB. Bayesian data monitoring in clinical trials. *Stat Med* 1997;**16:**1413–30.

[5] DerSimonian R. Meta-analysis in the design and monitoring of clinical trials. *Stat Med* 1996;**15:**1237–48.

[6] Heitjan DF. Bayesian interim analysis of phase II cancer clinical trials. *Stat Med* 1997;**16:**1791–802.

[7] Parmar MKB, Ungerleider RS, Simon R. Assessing whether to perform a confirmatory randomised clinical trial. *J Natl Cancer Inst* 1996;**88:**1645–51.

[8] Spiegelhalter DJ, Myles JM, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Health Technol Assess* 2000; **4(38)**. Available from http://www.ncchta.org/htapubs.htm

[9] Rasmussen HS, McNair P, Norregard P, Bacjer V, Lindeneg O, Balslev S. Intravenous magnesium in acute myocardial infarction. *Lancet* 1986;**i:**234–36.

[10] Teo KK, Yusuf S, Collins R, Held PH, Peto R. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. *Br Med J* 1991;**303:**1499–503.

[11] Shechter M, E. K, Rabinowitz B. The rationale of magnesium supplementation in acute myocardial infarction: a review of the literature. *Arch Intern Med* 1992;**152:**2189–96.

[12] Morton BC, Smith FM, McKibbon TJ, Nair RC, Poznanski WJ. Magnesium therapy in acute myocardial infarction—double blind study. *Magnesium* 1984;**3:**346–52.

[13] Morton BC, Nair RC, Smith FM, McKibbon TG, Poznanski WJ. Magnesium therapy in acute myocardial infarction. *Magnesium Bull* 1981;**1a:**192–94.

[14] Woods KL, Fletcher S, Roffe C, Haider Y. Intravenous magnesium sulphate in suspected acute myocardial infarction: results of the second Leicester Intravenous Magnesium Intervention Trial (LIMIT-2). *Lancet* 1992;**339:**1553–58.

[15] Yusuf S, Teo K, Woods K. Intravenous magnesium in acute myocardial infarction: an effective, safe, simple and inexpensive treatment. *Circulation* 1993;**87:**2043–46.

[16] ISIS-4 Collaborative Group. ISIS-4—a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium-sulphate in 58 050 patients with suspected acute myocardial infarction. *Lancet* 1995;**345:**669–85.

[17] Yusuf S, Flather M. Magnesium in acute myocardial-infarction. *Br Med J* 1995;**310:**751–52.

[18] Yusuf S. Meta-analysis of randomised trials: looking back and looking ahead. *Controlled Clin Trials* 1997;**18:**594–601.

[19] Peto R, Collins R, Gray R. Large-scale randomised evidence—large, simple trials and overviews of trials. *J Clin Epidemiol* 1995;**48:**23–40.

[20] Teo KK, Yusuf S. Role of magnesium in reducing mortality in acute myocardial infarction: a review of the evidence. *Drugs* 1993;**46:**347–59.

[21] Egger M, Davey Smith G. Misleading meta-analysis. *Br Med J* 1995; **311:**753–54.

[22] Egger M, Davey Smith G. Magnesium and myocardial-infarction. *Lancet* 1994;**343:**1285.

[23] Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Br Med J* 1997;**315:** 629–34.

[24] Pogue JM, Yusuf S. Cumulating evidence from randomised trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Controlled Clin Trials* 1997;**18:**580–93.

[25] Pogue JM, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *Lancet* 1998;**351:**47–52.

[26] Anonymous. Magnesium, myocardial infarction, meta-analysis and megatrials. *Drug Therapeutics Bull* 1995;**33:**25–27.

[27] Antman EM. Magnesium in acute MI: timing is critical. *Circulation* 1995;**92:**2367–72.

[28] Antman EM. Randomized trials of magnesium in acute myocardial infarction: big numbers do not tell the whole story. *Am J Cardiol* 1995; **75:**391–93.

[29] Seelig MS, Elin RJ, Antman EM. Magnesium in acute myocardial infarctionl still an open question. *Can J Cardiol* 1998;**14:**745–49.

[30] Woods KL. Mega-trials and management of acute myocardial-infarction. *Lancet* 1995;**346:**611–14.

[31] Seelig MS, Elin RJ. Is there a place for magnesium in the treatment of acute myocardial infarction? *Am Heart J* 1996;**132:**471–77.

[32] Shechter M, Kaplinsky E, Rabinowitz B. Review of clinical evidence—is there a role for supplemental magnesium in acute myocardial infarction in high-risk populations (patients ineligible for thrombolysis and the elderly)? *Coronary Artery Dis* 1996;**7:**352–58.

[33] Antman EM. Magnesium in acute myocardial infarction: overview of available evidence. *Am Heart J* 1996;**132:**487–94.

[34] Borzak S, Ridker PM. Discordance between meta-analyses and large-scale randomized, controlled trials: examples from the management of acute myocardial infarction. *Ann Intern Med* 1995;**123:**873–77.

[35] Casscells W. Magnesium and myocardial infarction [letter]. *Lancet* 1994;**343:**807–09.

[36] Teo KK, McAlister FA, Montague TJ. Results of randomized clinical trials of magnesium—can a consensus be reached? *Coronary Artery Dis* 1996;**7:**341–47.

[37] Antman EM, Seelig MS, Fleischmann K *et al*. Magnesium in acute myocardial infarction: scientific, statistical, and economic rationale for its use. *Cardiovasc Drugs Therapy* 1996;**10:**297–301.

[38] Sterne JAC, Bradburn MJ, Egger M. Meta-analysis in STATA. In: Egger M, Davey Smith G, Altman DG (eds). *Systematic Reviews in Health Care: Meta-analysis in Context*. London: BMJ Books, 2001, pp.347–69.

[39] Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Progr Cardiovasc Dis* 1985;**27:**335–71.

[40] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;**7:**177–88.

[41] Woolf B. On estimating the relation between blood group and disease. *Ann Hum Genet* 1955;**19:**251–53.

[42] MAGIC Steering Committee. Rationale and design of the magnesium in coronaries (MAGIC) study: a clinical trial to reevaluate the efficacy of early administration of magnesium in acute myocardial infarction. *Am Heart J* 2000;**139:**10–14.

[43] Li J, Zhang Q, Egger M. Intravenous magnesium for acute myocardial infarction (Protocol for a Cochrane review). In: The Cochrane Library, Oxford: Update Software, 2001 Issue 1.

[44] Greenland S, Salvan A. Bias in the one-step method for pooling study results. *Stat Med* 1990;**9:**247–52.

[45] Higgins JPT. *Exploiting Information in Random Effects Meta-analysis*. PhD thesis, The University of Reading, UK, 1997.

[46] Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;**53:**1119–29.

[47] Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Stat Med* 2001;**20:**641–54.

[48] Vanhouwelingen HC, Zwinderman K, Stijnen T. A bivariate approach to meta-analysis. *Stat Med* 1993;**12:**2272–84.

[49] McIntosh MW. The population risk as an explanatory variable in research synthesis of clinical trials. *Stat Med* 1996;**15:**1713–28.

[50] Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Stat Med* 1997;**16:**2741–58.

[51] Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Stat Med* 2000;**19:**3251–74.

[52] Spiegelhalter DJ, Thomas A, Best NG. *WinBUGS Version 1.3 User Manual*. MRC Biostatistics Unit, Cambridge 2000. Available from www.mrc-bsu.cam.ac.uk/bugs

[53] Bernardo JM, Smith AFM. *Bayesian Theory*. Chichester, England: John Wiley, 1994.

[54] Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* 1994;**309:**1351–55.

[55] Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;**15:**2733–49.

[56] Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995;**14:**2685–99.

Appendix

We describe here our statistical models and how we obtain a 'reasonably sceptical' prior distribution. We also illustrate how analyses were performed using WinBUGS,[52] and present some sensitivity analyses for the underlying risk meta-regression.

(1) Models

We consider two ways of summarizing within-trial treatment effects for Bayesian meta-analysis of binary outcomes: both are summarized by Higgins and Whitehead.[55] Suppose, in the i^{th} trial, r_i^M events were observed in a total of n_i^M patients in the magnesium group, and r_i^C events in a total of n_i^C patients in the control (placebo) group.

Method (i)

A Bayesian version of the Peto method takes an estimate of log(odds ratio), δ_i, $(O_i - E_i)/V_i$ with approximate variance $1/V_i$ where

$$O_i = r_i^M$$

$$E_i = \frac{r_i^C + r_i^M}{n_i^C + n_i^M}\, n_i^M$$

$$V_i = \frac{n_i^C n_i^M (r_i^C + r_i^M)(n_i^C + n_i^M - r_i^C - r_i^M)}{(n_i^C + n_i^M)^2 (n_i^C + n_i^M - 1)}.$$

The sampling model comprises the assumption that

$$\frac{O_i + E_i}{V_i} \sim \mathrm{N}(\delta_i,\, 1/V_i). \qquad (1)$$

Method (ii)

The second approach is the full binomial approach described by Smith *et al.*,[56] in which for each trial the log(odds ratio), δ_i, is modelled directly as a function of the underlying risks in the two treatment groups.

For either method, different assumptions (or prior distributions) regarding the unknown quantities (particularly the δ_i) give rise to different statistical models for the meta-analysis. The fixed effect model corresponds to the assumption that all underlying treatment effects are identical such that $\delta_i = \delta$. The conventional random effects model corresponds to an assumption that the treatment effects arise from a normal distribution: $\delta_i \sim \mathrm{N}(\mu,\tau^2)$. Meta-regression is a simple extension of the model to incorporate a trial-level covariate, x_i, acting on the log(odds ratio): $\delta_i \sim \mathrm{N}(\mu + \beta(x_i - \bar{x}),\tau^2)$. The covariate value at which treatment ceases to be beneficial on average may be estimated using $x_0 = \bar{x} - (\mu/\beta)$. A similar construction may be used to estimate the underlying risk at which the average treatment effect is zero.

(2) Expressing a sceptical prior

Our 'reasonably sceptical' prior distribution, N(0,0.03) for δ or for μ was derived as follows. We seek a normal distribution for the log(odds ratio). To describe scepticism this will be centred on 0 (an odds ratio of 1), and we prescribe that there is a 5% probability of the odds ratio being below 0.75. By symmetry, such a distribution implies a 5% probability of the odds ratio exceeding 1.333. On the log(odds ratio) scale, the distribution has 90% of its area between –0.288 and 0.288. Using simple normal theory, the standard deviation, SD, of the log(odds ratio) fulfils $1.645 \times \mathrm{SD} = 0.288$, so that the desired variance is $\mathrm{SD}^2 = 0.03$.

To determine the size of a trial that would provide evidence equivalent to this degree of scepticism, we assume a two-arm trial with equal allocation and no treatment effect, with mortality risk equal to the overall risk of patients across the control groups of the first eight trials. There were 171 deaths among 1809 patients in these control groups, yielding an overall risk of 0.095. We use the standard variance approximation,

$$\mathrm{var(logOR)} = \frac{1}{r^C} + \frac{1}{n^C - r^C} + \frac{1}{r^M} + \frac{1}{n^M - r^M}.$$

To obtain the required value of var(logOR) = 0.03, we take $r^C = r^M = 0.095n^C = 0.095n^M$, and solving yields $n^C = n^M = 761$, and therefore a total of 72 deaths in each group.

(3) Analysis using WinBUGS

WinBUGS code for the random effects version of the binomial model is as follows. Note that WinBUGS represents normal distributions using the mean and inverse-variance (precision) rather than mean and variance. The data consist of arrays `rc[]`, `nc[]`, `rm[]` and `nm[]`, and `k`, the number of trials. To specify fixed effects for control group risks, we place a uniform distribution between 0 and 1 on each π_i^C. We place a uniform prior on τ. Our 'reference' priors for δ, μ and βs are also uniform, implemented as a normal distribution with very large variance (10 000).

```
model {
   for(i in 1:k) {
      rc[i] ~ dbin(pc[i],nc[i]);            #Binomial structure
      rm[i] ~ dbin(pm[i],nm[i]);
      phi[i] <- logit(pc[i]);
      logit(pm[i]) <- phi[i] + delta[i];    #Define log(odds ratio)
      delta[i] ~ dnorm(mu, precision);      #Random effects
      pc[i] ~ dunif(0,1)                    #Prior for pc
   }

   delta.new ~ dnorm(mu, precision);        #Predicted effect

   mu ~ dnorm(0.0, 0.0001);                 #Reference prior
#  mu ~ dnorm(0.0, 32.69);                  #Sceptical prior
   tau ~ dunif(0, 100);                     #Prior for tau
   precision <- 1/(tau*tau);
   tau.sq <- 1/precision;

   less0 <- min(theta.new, 0);              #Calculate probabilities
   prob0 <- 1 - equals(less0, 0);           #in Table 2
   less10 <- min(theta.new, -0.1054);
   prob10 <- 1 - equals(less10, -0.1054);
}
```

For the Peto approach, data arrays `y[]` and `w[]` are defined to take values $(O_i - E_i)/V_i$ and V_i respectively, and the first loop is replaced by

```
   for(i in 1:k) {
      y[i] ~ dnorm(delta[i], w[i]);         #Estimated log(odds ratio)
      delta[i] ~ dnorm(mu, precision);      #Random effects
   }
```

For underlying risk meta-regression, the line defining `delta` is replaced by

```
      delta[i] <- delta0[i] + beta.r * (phi[i] - mean(phi[1:k]));
      delta0[i] ~ dnorm(mu, precision);
```

A prior distribution and the prediction interval are included at the end of the model as follows, where the additional data array `pc.pred[]` contains a sequence of `Npred` underlying risks for which the log(odds ratio) is to be predicted.

```
   beta.r ~ dnorm(0.0, 0.0001)              #Prior for beta.r
   for(j in 1:Npred) {
      phi.pred[j] <- logit(pc.pred[j]) ;
      delta0.pred[j] ~ dnorm(mu, precision);
      delta.pred[j] <- delta0.pred[j]+beta.r*(phi.pred[j]-mean(phi[1:k]));
   }
```

(4) Sensitivity analyses

To illustrate the sensitivity of the underlying risk analysis to assumptions, we present brief results from a selection of alternative analyses of the 15 trials for comparison with the second row of the second part of Table 4. Assuming a random effect for the control group logit(risks) gives an estimate of slope of –0.31 (95% credible interval –1.01, 0.51). Replacing the uniform prior on τ with an inverse gamma (0.001, 0.001) on $1/\tau^2$ gives a slope of –0.47 (–0.97, 0.13). McIntosh's approach[49] based on sample odds ratios and sample control group logit(risks) gives a slope of –0.75 (–1.45, 0.01).