STATISTICS IN MEDICINE
Statist. Med. 2002; **21**:1559–1573 (DOI: 10.1002/sim.1187)

How should meta-regression analyses be undertaken and interpreted?

Simon G. Thompson*,† and Julian P. T. Higgins

MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, U.K.

SUMMARY

Appropriate methods for meta-regression applied to a set of clinical trials, and the limitations and pitfalls in interpretation, are insufficiently recognized. Here we summarize recent research focusing on these issues, and consider three published examples of meta-regression in the light of this work. One principal methodological issue is that meta-regression should be weighted to take account of both within-trial variances of treatment effects and the residual between-trial heterogeneity (that is, heterogeneity not explained by the covariates in the regression). This corresponds to random effects meta-regression. The associations derived from meta-regressions are observational, and have a weaker interpretation than the causal relationships derived from randomized comparisons. This applies particularly when averages of patient characteristics in each trial are used as covariates in the regression. Data dredging is the main pitfall in reaching reliable conclusions from meta-regression. It can only be avoided by prespecification of covariates that will be investigated as potential sources of heterogeneity. However, in practice this is not always easy to achieve. The examples considered in this paper show the tension between the scientific rationale for using meta-regression and the difficult interpretative problems to which such analyses are prone. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: meta-analysis; meta-regression; heterogeneity; study-level covariates; ecological associations; false positive results

1. INTRODUCTION

Traditional methods of meta-analysis attempt to combine results in order to obtain a single summarized ‘effect size’. The observed effect in each study is an estimate, with some imprecision, of the true effect in that study. Statistical heterogeneity refers to the true effects in each study not being identical. Clinical and methodological diversity among the studies included in a meta-analysis necessarily leads to statistical heterogeneity [1]. In contrast to simple meta-analysis, meta-regression aims to relate the size of effect to one or more characteristics of the studies involved. For example, placebo controlled clinical trials of a drug at higher doses may yield larger observed effects of the treatment, or clinical trials of a change in diet versus no

*Correspondence to: S. G. Thompson, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, U.K.
† E-mail: simon.thompson@mrc-bsu.cam.ac.uk

Contract/grant sponsor: U.K. Medical Research Council

change may yield larger effects when the diet produces greater reductions in serum cholesterol level. The potential scientific value of such explorations of sources of heterogeneity has been emphasized in the past [1–5], so that meta-regression is now becoming a more widely used technique. Although we focus on clinical trials in this paper, the issues discussed are also relevant (and indeed may be even more problematic) for meta-analyses of observational studies.

We use the term meta-regression to indicate the use of trial-level covariates, as distinct from regression analyses that are possible when individual patient data on outcomes and covariates are available. In this paper we review recent research on statistical methods for meta-regression, and then discuss the limitations and pitfalls of the technique. In the light of this, we examine a number of practical examples where meta-regression has been used within systematic reviews. We focus on the extent to which meta-regressions influence the clinical conclusions of a systematic review, pointing out problems and areas of uncertainty.

To motivate our paper, we briefly introduce one example first. Johnson *et al.* report an investigation of the effect of different doses of aspirin in the secondary prevention of stroke, using data from eleven trials (Table I) [6]. Aspirin was administered in the different trials in widely ranging doses from 50 to 1500 mg/d. There had been disagreement about the optimal dose of aspirin in this clinical situation, and the principal aim of the paper is to undertake a single meta-regression to determine whether higher doses are associated with increased benefit. Figure 1 shows the log relative risk estimated in each trial plotted against the aspirin dose, with the area of each circle inversely proportional to the within-trial variance of the log relative risk. The purpose of the meta-regression is to quantify the relationship between log relative risk and aspirin dose, as shown by the line in Figure 1. Further discussion of this example is deferred to Section 5.1.

2. STATISTICAL METHODS FOR META-REGRESSION

Various statistical methods for meta-regression have been published. For example, fixed effect meta-regression was described originally [7], a random effects model more recently [8] and a fuller comparison of available methods made subsequently [9]. Other papers have also addressed methodology in this area [10–12]. Here we summarize the main conclusions and ramifications from this work. The points are numbered so that they can be referred to when considering the published examples of meta-regressions in Section 5:

(i) A visual presentation of a meta-regression relationship is essential. A diagram such as Figure 1 seems to be the most helpful, using symbol sizes that relate to the precision of each treatment effect estimate. If equally sized points are used, it is impossible to see which trials provide the greatest information, and how this might affect the interpretation. Confidence intervals around the points tend to pull the eye towards the small imprecise trials, which is exactly the opposite of what is wanted. A diagram such as Figure 1 emphasizes that the unit of analysis is the trial, not the individual patient. If there are few trials, even if there are many patients, meta-regression is unlikely to be scientifically useful.

(ii) Meta-regression investigates whether particular covariates (potential 'effect modifiers') explain any of the heterogeneity of treatment effects between studies. It is not

Table I. Characteristics of the trials and analyses in three meta-regression publications.

	Aspirin dose and secondary prevention of stroke [6]	Beta-blockers after myocardial infarction [30]	Aminoglycosides for bacterial infection [31]
Number of trials	11	31	19
Treatments	Aspirin versus placebo	Beta-blockers versus placebo or alternative control	Single versus multiple daily doses
Number of subjects	9629	24974	2177
Number of events	1391 strokes	2415 deaths	256 clinical failures
Average duration of treatment and follow-up (range)	2 to 4 years	0.5 to 4 years	7 to 10 days
Meta-regression technique	Fixed effect meta-regression of summary statistics	Bayesian random effects logistic regression	Fixed effect meta-regression of summary statistics
Number of reported covariates	1	4	7
Prespecification of covariates	Yes	Partial/unclear	No

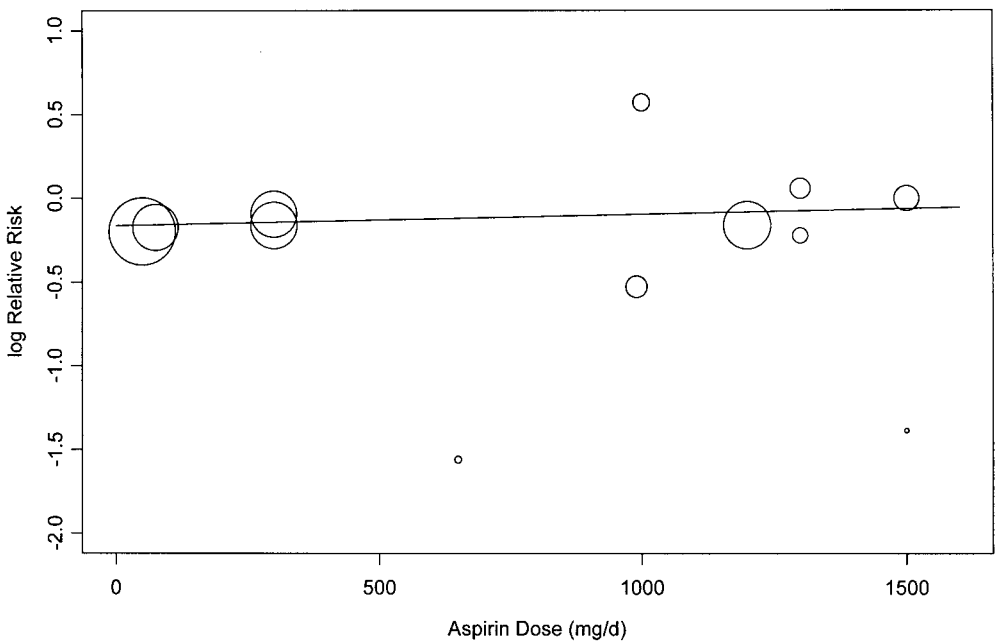

Figure 1. Log relative risk of stroke in 13 trials of aspirin versus placebo [6], according to aspirin dose, together with a summary random effects meta-regression. The area of each circle is inversely proportional to the variance of the log relative risk estimate.

reasonable to assume that all of the heterogeneity is explained, and the possibility of 'residual heterogeneity' must be acknowledged in the statistical analysis. The appropriate analysis is therefore 'random effects' rather than 'fixed effect' meta-regression. For given values of the covariates considered, a fixed effect analysis estimates the assumed common effect, whereas a random effects analysis estimates the mean of a distribution of effects across studies. If residual heterogeneity exists, a random effects analysis appropriately yields wider confidence intervals for the regression coefficients than a fixed effect analysis.

(iii) The regression should clearly be weighted, so that the more precise studies have more influence in the analysis. However, the weight for each trial should be equal to the inverse of the sum of the within-trial variance and the residual between-trial variance, in order to correspond to a random effects analysis. Taking weights equal to the inverse of the within-trial variances alone yields a fixed effect meta-regression analysis.

(iv) It is appropriate to use meta-regression to explore sources of heterogeneity even if an initial overall test for heterogeneity is non-significant. It is well known that this test often has low power [13] and therefore a non-significant result does not reliably identify lack of heterogeneity. Furthermore, the test is for general 'overdispersion' of trial results, and does not address whether heterogeneity relates to particular covariates. In any case, some would argue that such a test for heterogeneity is redundant because we know, given the diversity of trials in any meta-analysis, that heterogeneity must exist. Whether we happen to be able to detect it or not is irrelevant.

(v) Estimating the residual between-trial variance is somewhat problematic. The estimate is usually imprecise because it is based on rather a limited number of trials. Different authors have advocated different estimates, for example an empirical Bayes estimate [8] or a restricted maximum likelihood (REML) estimate [9]. Moreover, conventional random effects methods ignore the imprecision in the between-trial variance estimate. One way to allow for the imprecision is to adopt a Bayesian approach, using, for example, non-informative priors [14]. While this is preferable in principle, especially when the number of trials is small or when the between-trial variance is estimated as zero, the resulting widening of the confidence intervals is rather slight in most practical examples. Choice of 'non-informative' priors can also be somewhat problematic in a Bayesian analysis [15].

(vi) The outcome (or dependent) variable in a meta-regression analysis is usually a summary statistic, for example the observed log-odds ratio from each trial. The estimated variance of this summary statistic is assumed to be the true variance, an assumption that is less appropriate when trials are small. There may also be some bias introduced for binary outcome data because of a structural relationship between, for example, odds ratio estimates and their variance estimates based on the counts in a 2×2 table [8]. A practical alternative for binary data is often available, since many publications include the original counts of events and non-events in each treatment group in each trial. The resulting analysis, a random effects logistic regression, is preferable in principle to an analysis using summary statistics, but again it seems rare to find examples where this makes an important difference in practice [9, 11].

(vii) It is easiest to think of meta-regression in the context of a continuous covariate, as in Figure 1. Heterogeneity is however often addressed in practice by subgrouping trials with different characteristics. Such subgroup analysis is equivalent to meta-regression with a categorical trial-level covariate. Considering subgroup analysis formally as a meta-regression has advantages, since it focuses on differences between subgroups as is appropriate, rather than the effects in each subgroup separately. Also, in a random effects setting, allowance is made for residual heterogeneity not explained by the subgrouping.

Software to undertake random effects meta-regression using summary statistics is available, for example in Stata [16]. This uses an iterative method to provide REML estimates of regression parameters, their asymptotic variances, and the residual heterogeneity variance. Software for random effects logistic regression requires software explicitly designed for hierarchical models. For example, MLwiN and SAS Proc Mixed employ a classical statistical approach, and BUGS a Bayesian approach [9, 17].

3. LIMITATIONS OF META-REGRESSION

Even if appropriate statistical methods have been used for meta-regression, there are a number of limitations to the interpretation of the results. These are summarized below:

(i) The relationship described by a meta-regression is an observational association across trials. Although the original studies may be randomized trials, the meta-regression is across trials and does not have the benefit of randomization to underpin a causal

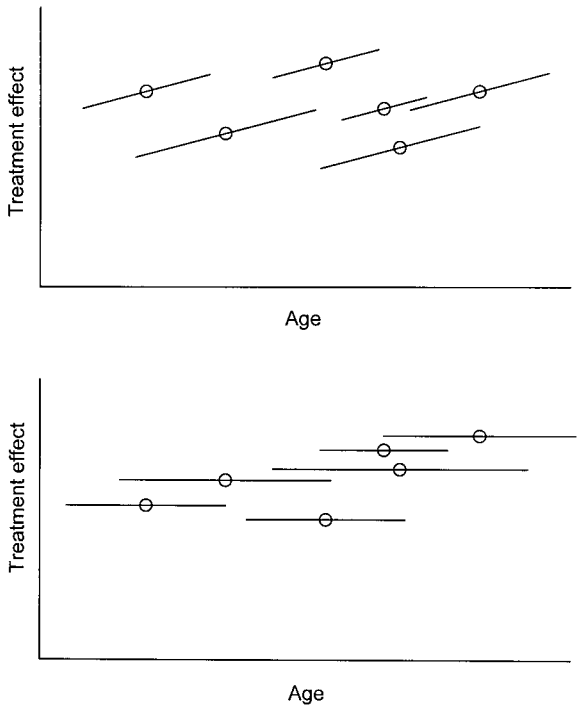

Figure 2. Hypothetical relationships between age and treatment effect both within trials (represented by the lines) and between trials (represented by the dots). See Section 3(ii) for explanation.

interpretation. One can view meta-regression as a study of the 'epidemiology of trials'; it suffers from the same disadvantages as other observational epidemiological investigations, notably bias by confounding. Thus an association identified with one trial characteristic may in reality reflect a true association with other correlated characteristics, whether these are known or unknown. This is a particular problem in meta-regression because there are many characteristics which differ between the trials in a meta-analysis, and these can be highly correlated [10].

(ii) Meta-regression is sometimes used to relate the results of the trials to published averages of patient characteristics within trials, for example average age or proportion of women. Such analyses are difficult to interpret. A generally minor issue is that such relationships may, if the patient averages are based on small samples, be attenuated by measurement error [12, 18]. More important however is that the relationship with patient averages across trials may not be the same as the relationship for patients within trials. Figure 2 shows a hypothetical example of this. In the upper diagram, treatment effect is related to age within each trial (represented by the sloping lines in the figure), but is not related to mean age across trials (represented by the dots). In the lower diagram, the opposite is the case, in that there is a relationship across trials but none within trials. Such situations can occur in practice [17], and arise through confounding at either the trial level (biasing the relationship across trials) or at the individual level

Table II. Hypothetical results for treatment effect according to sex in four trials (see Section 3(iv) for explanation). All odds ratios are assumed to be equally precise.

Trial	Sex	Treatment effect odds ratio
Trial 1	M	0.7
	F	1.0
Trial 2	M	1.0
Trial 3	F	0.4
Trial 4	M	0.7
	F	1.0

M versus F overall, 0.8 versus 0.8 (ratio 1.0).
M versus F within trials, 0.7 versus 1.0 (ratio 0.7).

(biasing the relationship within trials). The phenomenon is variously referred to as 'aggregation bias', 'ecological bias', 'ecological confounding' or the 'ecological fallacy' [19], and without individual patient data cannot be investigated.

(iii) While the impact of any measurable characteristic of the studies may theoretically be investigated using meta-regression, results are easier to interpret when the characteristic has high variability across studies compared to within studies. To illustrate this, suppose exactly the same wide range of doses had been studied among patients within each of a set of trials. A single summary of dose might be obtained for each trial by taking the average dose for patients within the trial. However, a meta-regression on average dose would be of limited use since the averages from the different trials would be similar to each other, and there would be little potential to discriminate between the trials [20]. That statistically non-significant relationships should not be equated to absence of true relationships is of course true in general, but particularly so in these circumstances.

(iv) For categorical covariates, it is sometimes possible to identify, even from published papers, the relationship with treatment effects both within trials and between trials. A hypothetical example is given in Table II. Trials 1 and 4 contain both sexes (and results are available separately for men and women), whereas trial 2 recruited only men and trial 3 only women. We assume that each observed odds ratio in Table II is equally precise. Comparing men and women overall, the average odds ratio is 0.8 in men and 0.8 in women (a ratio between sexes of 1.0). Comparing men and women only within trials, and thus using only trials 1 and 4, the average odds ratio is 0.7 in men and 1.0 in women (a ratio of 0.7). Thus a different answer is obtained if we rely only on information within trials (which is free from ecological bias across studies), or if we use all the information (which is more precise). Although this is simply another example of potential confounding across trials, it emphasizes the need to separate clearly whether the relationships being described by meta-regression are within trials, across trials, or (as is often the case) some mixture of the two.

(v) Meta-regression has sometimes been used to investigate whether treatment benefit depends on the 'underlying risk' of the patients in the trial. The underlying risk is usually measured by the risk or rate of events in each trial's control group. A conventional meta-regression analysis of this relationship is flawed by regression to the mean [21]: the measurement error in the covariate (control group risk) appears also in the

dependent variable (treatment effect) causing an artefactual negative association. More complex models, which address the dependency in the measurement errors, are needed to obtain valid results for the relationship between treatment benefit and underlying risk [22–24].

(vi) A practical limitation in meta-regression is the availability of data from the primary papers. Meta-regression requires the estimated treatment effect, its variance, and covariate values for each trial in the systematic review. A common reason for not being able to undertake meta-regression is that these are not all available. Any analysis can only be based on the subset of trials for which full information happens to be available, potentially biasing the results.

(vii) Often systematic reviews contain very few studies [25]. For example, in a study of 39 Cochrane systematic reviews, only one contained more than 15 studies with available data for the primary outcome [26]. Given such numbers of studies, the potential for robust conclusions from meta-regression analyses is clearly very limited.

4. THE PITFALLS OF META-REGRESSION

Exploring sources of heterogeneity may result in false positive conclusions through 'data dredging'. So important is this in practice that one might label it as the principal pitfall in meta-regression, even aside from the methodological aspects and interpretative limitations addressed above. This issue is summarized below:

(i) A common situation is that there are few trials in a meta-analysis but many possible trial or patient characteristics that might explain heterogeneity. Multiple analyses using each of these characteristics may be undertaken. Some analyses may be done simply because of observed patterns in the results of the trials, which suggest that a certain covariate may be important. Such multiple or *post hoc* analyses lead to data dredging and a high probability of false positive conclusions.

(ii) False positive conclusions deriving from meta-analyses are particularly hazardous. *Post hoc* conclusions from, for example, multiple subgroup analyses within a single clinical trial should be regarded as hypothesis generating rather than hypothesis testing. These hypotheses should then be investigated in other trial data sets. In a meta-analysis, which has accumulated the totality of evidence on a particular issue, there is no possibility of such external testing or validation. Thus we are left in the dangerous position of having speculative conclusions, without the ability to investigate them further (at least until many more trials have been undertaken). Indeed, concern over perceived ethical issues that stem from such 'apparently convincing' findings may prevent further research being undertaken to resolve the issue.

(iii) The only way to protect against such false positive conclusions, as in other situations [27, 28], is to prespecify which covariates are going to be investigated by subgroup analyses or meta-regressions. This raises some difficult issues. The first is the need for a protocol for the meta-analysis that identifies appropriate covariates. While Cochrane systematic reviews generally have protocols that are documented and publicly available [25], other systematic reviews and meta-analyses (for example most of those appearing in peer reviewed journals) may not. Secondly, in order for decisions about investigating heterogeneity to be truly prespecified, such a protocol should be drawn up without

knowledge of any of the relevant trials' results. Since meta-analysis summarizes studies that have already taken place and may be known to the reviewers, this is unlikely to be achieved in practice. This is especially so since subject-matter specialists should be involved in identifying suitable covariates that have a strong scientific rationale.

(iv) In addition to prespecification, it is necessary to limit the number of covariates proposed for investigation again to protect against false positive conclusions. If multiple covariates are of real scientific interest, false positive conclusions can be limited to a desired level by using a Bonferroni adjustment to the significance level for each covariate. Indeed such a strategy might encourage investigators to maintain power for a very few covariates, rather than attempt to investigate many.

Prespecification of relevant covariates proves difficult in practice. In a study of 39 systematic reviews in the Cochrane Database (CDSR), of which 28 had protocols, the covariates that were specified to be the subject of subgroup analyses or meta-regressions in the protocols were compared with the analyses which were actually undertaken in the reviews [26]. Amongst 15 that specified some such analyses in the protocol, seven undertook none, three did none that was specified but undertook some that were not specified, and three undertook ones that were not specified in addition to those that were. A mere two reviews undertook only ones that had been prespecified. This exemplifies the practical problem of specifying relevant subgroup or meta-regression analyses in advance. When faced with the reality of data to hand, either information was not available to investigate the originally identified covariates, or the investigators focused their attention on new apparently important covariates that had not been prespecified.

5. EXAMPLES

We selected three recently published papers that illustrate different applications of, and approaches to, meta-regression. The basic characteristics of the data available are summarized in Table I. The papers were chosen to exemplify relevant methodological and interpretative issues, as outlined in Sections 2 to 4, not to criticize these particular publications. Neither are all aspects relevant to the interpretation of the meta-regressions discussed below, nor do we claim that these papers are representative of the use of meta-regression in practice.

5.1. Aspirin in secondary stroke prevention

The first example, which has already been described in the introduction, relates to the secondary prevention of stroke and the dose of aspirin [6]. The statistical analysis used a weighted linear regression of the log relative risk of stroke on aspirin dose as a continuous covariate (Figure 1). The regression includes 12 data points, since one of the 11 studies randomized patients to placebo or one of two doses of aspirin. The weights used in the regression were the inverse within-trial variances of the log relative risks. Separate analyses assuming either linear or quadratic relationships between dose and log relative risk were performed, as well as a cubic spline analysis to allow for more general curvature.

A simple meta-analysis gave a Mantel–Haenszel pooled relative risk benefit of around 15 per cent (95 per cent CI 6 to 23 per cent). No statistically significant relationships with dose were identified. For the linear regression, the estimated increase in log relative risk per 100 mg/d increase in dose was reported as 0.0068 (SE 0.0096, 95 per cent CI -0.015 to

+0.028, $P = 0.49$). The predicted relative risk benefit of aspirin decreased from 15 per cent at 50 mg/d to 6 per cent at 1500 mg/d. The authors concluded that the dose–response relationship is essentially flat, a potentially valuable finding given that high doses of aspirin are associated with adverse side-effects.

A number of aspects of this analysis are interesting. The focus of the paper is on one covariate, aspirin dose, although three regressions (linear, quadratic and cubic spline) were investigated. Although it is not clear in what sense this was 'prespecified', data dredging would seem to be an unreasonable accusation in this case (Section 4(i)). The large range of aspirin dose across trials (and no variation within trials) makes meta-regression here a particularly suitable tool for investigating this as a covariate (Section 3(iii)). Meta-regression was carried out despite the lack of evidence for heterogeneity overall (reported $P = 0.23$), as is appropriate (Section 2(iv)).

Summary statistics constituted the dependent variable in the meta-regression (Section 2(vi)), being the log relative risk in each trial. A fixed effect meta-regression was used which does not allow for potential residual heterogeneity. Thus the standard error of the regression coefficient may be too small (Section 2(ii)). However, in our reanalysis based on the published information, a random effects meta-regression gave a REML estimate of zero for the residual heterogeneity variance, and so in this case the fixed effect and classical random effects meta-regression results coincide. Nevertheless a Bayesian approach, which allows for the imprecision in the zero estimate of residual variance based on only 11 trials, gives a slightly larger standard error (Section 2(v)), 0.0123 rather than 0.0096.

The estimated coefficient of +0.0068 per 100 mg/d corresponds to a proportionate increase in the relative risk of 10 per cent as aspirin dose increases from 50 to 1500 mg/d, implying decreased efficacy of aspirin with increasing dose. Conversely, the lower 95 per cent confidence limit of −0.015 corresponds to a 20 per cent decrease in relative risk. Thus while there is no evidence of a relationship with aspirin dose, the data cannot rule out a modest (and perhaps clinically relevant) increased benefit at higher aspirin doses. As always, it is tempting to accept lack of evidence for an effect as evidence of no effect, while of course this may be a false negative conclusion (Section 3(iii)).

The authors recognize in their discussion that the relationship with aspirin dose is across trials and may be confounded by other trial or patient characteristics (Section 3(i)). They point out that there was only one trial with more than one aspirin dose group (300 and 1200 mg/d); this provided the only directly randomized evidence about aspirin dose and treatment benefit (Section 3(iv)). In the meta-regression, this trial is apparently handled by including the placebo group twice, once for each aspirin dose. Appropriate inclusion of a three-arm trial in a meta-analysis (or meta-regression) requires that the placebo group data are only used once, which can be achieved by considering the original binary data directly (Section 2(vi)) rather than using log relative risks as summary statistics [29].

5.2. Beta-blockers after myocardial infarction

Freemantle *et al.* report a systematic review of 31 clinical trials of the long-term use of beta-blockers after myocardial infarction (Table I) [30]. The review aims to determine the effect on mortality of beta-blockers overall, and makes use of meta-regressions to address secondary clinical questions mainly relating to mechanisms of action of the different beta-blockers. The authors report investigations of four dichotomous characteristics. The focus was

on whether beta-blockers that act specifically on the heart (cardioselective), or specifically on the sympathetic nervous system (intrinsic sympathomimetic activity), were associated with higher efficacy in these trials, because these factors had been identified as potentially important in a previous meta-analysis. Drugs can be either cardioselective alone, sympathomimetic active alone, both, or neither. Two further characteristics of the studies were investigated: whether there was initial intravenous treatment, and publication date (before or after the median of 1982) as a proxy for the availability of additional treatment options.

The meta-analysis demonstrated a statistically significant benefit of beta-blockers on all-cause mortality (random effects odds ratio 0.77, 95 per cent CI 0.69 to 0.85), without statistically significant heterogeneity ($P = 0.16$). Random effects meta-regression was undertaken using a Bayesian approach, employing the binomial outcome data available in each trial. The estimated relative odds ratio for cardioselective drugs versus others was 1.10 (95 per cent CI 0.89 to 1.39), and for intrinsic sympathomimetic active drugs versus others was 1.19 (95 per cent CI 0.96 to 1.47). These indicate trends towards reduced benefit of beta-blockers with these characteristics. The authors concluded on the basis of the latter nearly significant finding, that drugs with intrinsic sympathomimetic activity 'should be avoided'. Neither early initial intravenous treatment nor publication date appeared to affect the efficacy of the beta-blockers.

This paper used a Bayesian random effects logistic regression analysis (Section 2(vi)), thus avoiding unduly precise inferences by accounting for imprecision in estimation of random effects (Section 2(v)). With 31 studies in the meta-analysis, random effects may be estimated with reasonable precision, and in fact non-Bayesian random effects meta-regressions give very similar results in this example. The investigation of heterogeneity employs only dichotomous covariates, and the authors have contrasted subgroups by using meta-regression as is appropriate (Section 3(vii)).

The meta-regression associations are observational in nature (Section 3(i)), and should not be interpreted as if they come from randomized comparisons. Indeed, the potential confounding between the four characteristics considered could be directly assessed, rather than carrying out four univariate meta-regressions. For example, confounding between cardioselectivity and intrinsic sympathomimetic activity could have been addressed in a meta-regression using both covariates simultaneously.

This example provides a typical situation regarding the difficulty in 'prespecification' of covariates (Section 4(iii)). Two of the covariates were identified as potentially important in a previous meta-analysis, which included over half of the trials that were in the current meta-analysis. Thus the 'prespecification' is to an extent data dependent. To what degree the other two covariates were prespecified is unclear, and indeed other covariates may have been investigated but not reported. Such a situation is typical for meta-analyses published in peer reviewed journals, in the absence of published protocols (Section 4(iii)). Given that (at least) four covariates were investigated, conclusions regarding the most significant one should be down-played (Section 4(iv)). One might therefore regard the authors' conclusions regarding beta-blockers with intrinsic sympathomimetic activity as overly strong.

5.3. Aminoglycosides for treating infections

Barza *et al.* present a systematic review of 19 randomized trials of single versus multiple (two or more) daily doses of aminoglycosides in the treatment of bacterial infections (Table I) [31]. Considerable heterogeneity of effect was discovered for the outcome of 'clinical failure' (the

original trials' definitions were used for this), and the authors used meta-regression as an exploratory tool to try and explain it. Seven characteristics were investigated, covering aspects of patients (proportion of bacteria isolated from patients that were found to be of pseudomonas species, which is more resistant to commonly used antibiotics), interventions (specific aminoglycoside, number of multiple doses, duration of treatment, presence of concurrent antibiotic medication) and other characteristics (publication date, control group event rate).

In a random effects meta-analysis, the single daily dose regimen was associated with a non-significant decrease in risk of clinical failure, with an overall relative risk of 0.83 (95 per cent CI 0.57 to 1.21). The test for heterogeneity was significant ($P = 0.03$). A fixed effect meta-regression was used, with log relative risk as the summary statistic response variable. The efficacy of single versus multiple doses of aminoglycosides was found to be unrelated to six of the covariates. However, for the 13 trials that provided information on pseudomonas, a significant relationship was observed: a 1 per cent increase in the prevalence of pseudomonas was associated with a 4.1 per cent increased benefit (95 per cent CI 1.6 per cent to 6.6 per cent, $P = 0.001$) of single dose over multiple dose aminoglycosides.

As in the first example, the fixed effect meta-regression does not allow for residual heterogeneity (Section 2(ii)) and the regression coefficients are too precisely estimated. The published paper does not contain the pseudomonas data, so the meta-regression cannot be represented visually here (Section 2(i)), and the analysis cannot be repeated using the more appropriate random effects method (Section 2(iii)).

The finding regarding pseudomonas is difficult to interpret. First, the analysis is a *post hoc* investigation, which is prone to produce false positive results (Section 4(i)). The finding may not be a real effect even though the significance level appears extreme. Second, the pseudomonas covariate is not available in six out of the 19 trials (Section 3(iv)). Third, the large number of covariates investigated strongly increases the chance of false positive results (Section 4(iv)). The seven characteristics investigated represent one for every three studies. Fourth, the association is observational across trials (Section 3(i)). Yet some exploratory findings may of course be real. If the finding regarding pseudomonas were real, there would likely be direct clinical consequences for treatment of patients with pseudomonas, and ethical restrictions on obtaining further randomized evidence. The unresolved challenge is to know which exploratory findings are real, and which are not.

The pseudomonas covariate is a summary of patient characteristics rather than a specific attribute of the trial (Section 3(ii)). The finding is an ecological association and not an association that can necessarily be extrapolated to individual patients. In other words, trials in which the patients had a higher prevalence of pseudomonas showed a greater treatment benefit, rather than patients with pseudomonas necessarily being the actual patients who are more likely to benefit. The authors recognize this and point out that 'this does not necessarily mean that patients with pseudomonas infection did better with a single daily dose'. A final point is that one of the covariates investigated was control group risk, and for this the use of simple meta-regression techniques is incorrect (Section 3(v)).

6. DISCUSSION

The three papers discussed above illustrate different objectives of meta-regression. The first aims to answer new questions by examining the influence of one key characteristic on the

size of treatment effect; the second attempts to gain additional insight into treatments by supplementing a meta-analysis with investigations of important clinical differences between trials; and the third generates new questions by trying to determine causes of variation in results of different trials. The examples are no doubt not typical in a number of respects, not least in that they include a generally large number of studies, individuals and events. Problems relating to data dredging, in essence over-fitting the available data, will be exacerbated when there are fewer studies even when covariates are prespecified. This is compounded by the (natural) tendency of authors to over-interpret results in published papers in order to render them apparently exciting or novel. Furthermore, data dredging is often hidden from the reader.

The same methodological issues and interpretative problems we discuss are relevant in undertaking meta-regressions of the findings in observational studies. Indeed some of the problems are more severe [32]. Observational studies are more variable in design than randomized trials, and heterogeneity in their results may reflect design differences rather than clinical diversity. Selection and other biases often hinder the interpretation of both meta-analyses and meta-regressions. The variables adjusted for in statistical analyses to reduce confounding within studies are almost always different (or differently handled) in each study. The binary nature of outcome data (Section 2(vi)) cannot sensibly be used for published observational studies, since adjustment for confounders would almost always be considered the more important issue. Measurement error affects the strength of associations, and its extent and impact may vary across studies. Getting individual patient data from observational studies is even more difficult than for randomized trials, while the effects of publication bias in the available literature may be more extreme. For all these reasons, the results from meta-regression of observational studies are even harder to interpret than those from randomized trials.

There are additional issues in meta-regression that we have not emphasized in this paper. For example, the linearity of the regressions for continuous covariates is usually assumed without comment (although the aspirin example above is an exception in this regard). Within the three examples presented, no multiple regressions were undertaken to investigate confounding, although some examples do exist [33]. The multivariate exploration of both trial and patient characteristics in meta-analysis has been expressed as an aspiration [5], but lack of data usually prevents anything other than very modest progress towards this ideal in practice. None of our three examples commented on the extent to which the heterogeneity is explained by the covariates considered, although this idea is part of the motivation of meta-regression. The proportion of between-trial variance explained will usually of course only be imprecisely estimated.

The description of weighted regression methods in published papers is often ambiguous. A phrase such as 'the regression was weighted by the inverse of the within-study variances' does not indicate whether the weights were taken *equal* to the inverse variances (resulting in a fixed effect meta-regression) or *proportional* to the inverse variances (resulting in a multiplicative rather than additive adjustment for residual heterogeneity) [9]. There is little to motivate the use of a multiplicative variance adjustment factor in meta-regression, since the within-study variances are known, although this is what is achieved by the conventional use of weighted regression programs in most statistical software. An additive component for the residual variance is more reasonable in both meta-regression [9] and other contexts [34], although the results from using multiplicative or additive components can be similar in practice. We and others [35, 36] strongly advocate the use of such random effects models, but recognize that some authors would disagree with us [37].

In general, no protocol is available for meta-analyses presented in published papers, and lack of space limits both description of methods and interpretation of results. One solution is that the Web could be used to give further details, as is starting to be done for some journals. Another practical issue is whether aspects of study quality, such as randomization concealment, should be used as a covariate before any variables representing clinical diversity are considered. Statistical issues for future research into meta-regression methods include how the number of covariates that can reliably be included depends on the number of trials (and their imprecisions) [38], the handling of multi-arm trials when individual patient data are not available, the appropriate use of regression diagnostics and sensitivity analyses [39, 40], and whether there are biases in using derived statistics measured after baseline as covariates. Examples of the latter include surrogate markers, such as extent of serum cholesterol reduction [29] or CD4 count increase [18], in considering disease event rates, or the standard error of the treatment effect as a covariate to discern publication bias [41].

Individual patient data, both of outcomes and covariates, can alleviate some of the problems in meta-regression. In particular within-trial and between-trial relationships can be more clearly distinguished, and confounding by individual level covariates can be investigated. Nevertheless many of the problems remain, not least those related to prespecification and data dredging. Statistical methods for individual patient data meta-regression also need exposition and development [17]. At the present time, however, most meta-analyses (and meta-regressions) are based on published information. Heterogeneity inevitably remains a difficult issue in meta-analysis but, as this paper has pointed out, both meta-analysts and consumers of their research need to be aware of the special hazards in the interpretation of meta-regressions.

ACKNOWLEDGEMENTS

We thank Dr Melissa Wake and Dr Harriet Hiscock for providing clinical background to the published meta-regressions discussed, and two referees for their constructive comments. Both authors are supported by the U.K. Medical Research Council.

REFERENCES

1. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *British Medical Journal* 1994; **309**:1351–1355.
2. Rubin D. A new perspective. In *The Future of Meta-analysis*, Wachter KW, Straf ML (eds). Russell Sage Foundation: New York, 1990; 155–165.
3. Berlin JA. Benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *American Journal of Epidemiology* 1995; **142**:383–387.
4. Davey Smith G, Egger M, Phillips AN. Meta-analysis: beyond the grand mean? *British Medical Journal* 1997; **315**:1610–1614.
5. Lau J, Ioannidis JP, Schmid CH. Summing up evidence: one answer is not always enough. *Lancet* 1998; **351**: 123–127.
6. Johnson ES, Lanes SF, Wentworth CE, Satterfield MH, Abebe BL, Dicker LW. A metaregression analysis of the dose-response effect of aspirin on stroke. *Archives of Internal Medicine* 1999; **159**:1248–1253.
7. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews* 1987; **9**:1–30.
8. Berkey CS, Hoaglin DC, Mosteller F, Colditz GA. A random effects regression model for meta-analysis. *Statistics in Medicine* 1995; **14**:395–411.
9. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine* 1999; **18**:2693–2708.
10. Berlin JA, Antman EM. Advantages and limitations of meta-analytic regressions of clinical trial data. *Online Journal of Clinical Trials* 1994; **3**:document 134.

11. Schmid CH. Exploring heterogeneity in randomised trials via meta-analysis. *Drug Information Journal* 1999; **33**:211–224.
12. Greenwood CMT, Midgley JP, Matthew AG, Logan AG. Statistical issues in a meta-regression analysis of randomized trials: impact on the dietary sodium intake and blood pressure relationship. *Biometrics* 1999; **55**: 630–636.
13. Hardy RJ, Thompson SG. Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* 1998; **17**:841–856.
14. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random effects meta-analysis: a comparative study. *Statistics in Medicine* 1995; **14**:2685–2699.
15. Natarajan R, Kass RE. Reference Bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association* 2000; **95**:227–237.
16. Sharp SJ. Meta-analysis regression. *Stata Technical Bulletin* 1998; **42**:16–22.
17. Higgins JPT, Whitehead A, Turner RM, Omar RZ, Thompson SG. Meta-analysis of continuous outcome data from individual patients. *Statistics in Medicine* 2001; **20**:2219–2241.
18. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* 1997; **16**:1965–1982.
19. Morgenstern H. Uses of ecological analysis in epidemiologic research. *American Journal of Public Health* 1982; **72**:127–130.
20. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *Journal of Clinical Epidemiology* 2002;**55**:86–94.
21. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. *British Medical Journal* 1996; **313**:735–738.
22. Thompson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* 1997; **16**:2741–2758.
23. McIntosh M. The population risk as an explanatory variable in research synthesis of clinical trials. *Statistics in Medicine* 1996; **15**:1713–1728.
24. Sharp SJ, Thompson SG. Analysing the relationship between treatment effect and underlying risk in meta-analysis: comparison and development of approaches. *Statistics in Medicine* 2000; **19**:3251–3274.
25. The Cochrane Collaboration. *Cochrane Database of Systematic Reviews*. Update Software: Oxford, 2000, Issue 2.
26. Higgins JPT, Thompson SG, Altman DG, Deeks JJ. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *Journal of Health Services Research and Practice* 2002:**7**:51–61.
27. Oxman AD, Guyatt GH. A consumer's guide to subgroup analysis. *Annals of Internal Medicine* 1992; **116**: 78–84.
28. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analyses and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; **355**:1064–1069.
29. Thompson SG. Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Statistical Methods in Medical Research* 1993; **2**:173–192.
30. Freemantle N, Cleland J, Young P, Mason J, Harrison J. β blockade after myocardial infarction: systematic review and meta-regression analysis. *British Medical Journal* 1999; **318**:1730–1737.
31. Barza M, Ioannidis JPA, Cappelleri JC, Lau J. Single or multiple daily doses of aminoglycosides: a meta-analysis. *British Medical Journal* 1996; **213**:338–345.
32. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *British Medical Journal* 1998; **316**:140–145.
33. Sterne JAC, Egger M, Davey Smith G. Investigating and dealing with publication and other biases. In *Systematic Reviews in Health Care*: *Meta-analysis in Context*, Egger M, Davey Smith G, Altman DG (eds). BMJ Books: London, 2001; 189–208.
34. Goldstein H. *Multilevel Statistical Models*. Edward Arnold: London, 1995.
35. National Research Council. *Combining Information*: *Statistical Issues and Opportunities for Research*. National Academy Press: Washington DC, 1992.
36. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. *Health Technology Assessment* 1998; **2**:report 19.
37. Peto R. Why do we need systematic overviews of randomised trials? *Statistics in Medicine* 1987; **6**:233–240.
38. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 1996; **49**:1373–1379.
39. Hedges LV, Olkin I. *Statistical Methods for Meta-analysis*. Academic-Press: London, 1985; 257–263.
40. Olkin I. Diagnostic statistical procedures in medical meta-analyses. *Statistics in Medicine* 1999; **18**:2331–2341.
41. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 1997; **315**:629–634.

