RESEARCH METHODS & REPORTING

Interpretation of random effects meta-analyses

Richard D Riley,[1] Julian P T Higgins,[2] Jonathan J Deeks[1]

[1]Department of Public Health, Epidemiology and Biostatistics, Public Health Building, University of Birmingham, Birmingham B15 2TT, UK
[2]MRC Biostatistics Unit, Institute of Public Health, Cambridge CB2 OSR, UK
Correspondence to: R D Riley
r.d.riley@bham.ac.uk

bmj.com Previous articles in this series
❯ Economic evaluation using decision analytical modelling: design, conduct, analysis, and reporting (*BMJ* 2011;342:d1766)
❯ Economic evaluation alongside randomised controlled trials: design, conduct, analysis, and reporting (*BMJ* 2011;342:d1548)
❯ Strengthening the reporting of genetic risk prediction studies: the GRIPS statement (*BMJ* 2011;342:d631)
❯ Correlation in restricted ranges of data (*BMJ* 2011;342:d556)

Summary estimates of treatment effect from random effects meta-analysis give only the average effect across all studies. Inclusion of prediction intervals, which estimate the likely effect in an individual setting, could make it easier to apply the results to clinical practice

Meta-analysis is used to synthesise quantitative information from related studies and produce results that summarise a whole body of research.[1] A typical systematic review uses meta-analytical methods to combine the study estimates of a particular effect of interest and obtain a summary estimate of effect.[2] For example, in a meta-analysis of randomised trials comparing a new treatment with placebo, researchers will collect the estimates of treatment effect for each study, as measured by a relevant statistic such as a risk ratio, and then statistically synthesise them to obtain a summary estimate of the treatment effect.
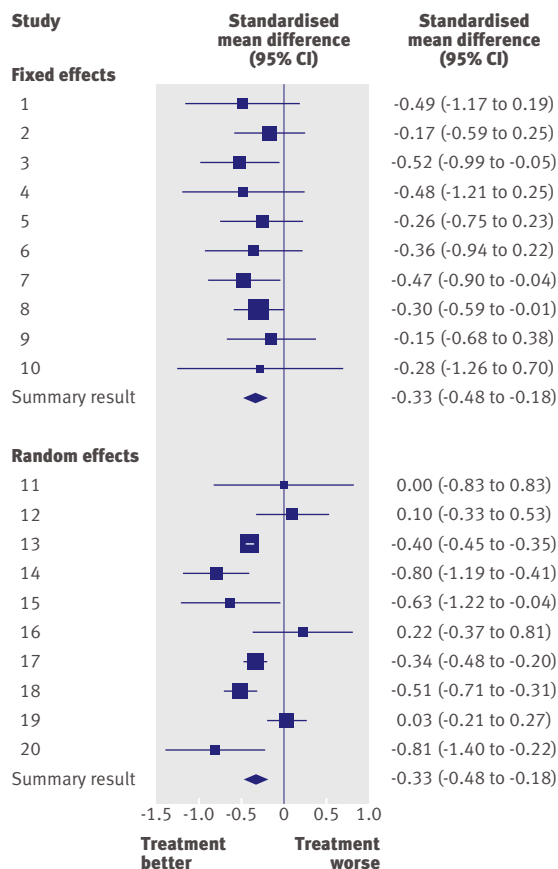
Meta-analyses use either a fixed effect or a random effects statistical model. A fixed effect meta-analysis assumes all studies are estimating the same (fixed) treatment effect, whereas a random effects meta-analysis allows for differences in the treatment effect from study to study. This choice of method affects the interpretation of the summary estimates. We examine the differences and explain why a prediction interval can provide a more complete summary of a random effects meta-analysis than is usually provided.

Difference between fixed effect and random effects meta-analyses

Figure 1 shows two hypothetical meta-analyses, in which estimates of treatment effect are computed and synthesised from 10 studies of the same antihypertensive drug. Each study provides an unbiased estimate of the standardised mean difference in change in systolic blood pressure between the treatment group and the control group. Negative estimates indicate a greater blood pressure reduction for patients in the treatment group than the control group.

The two meta-analyses give identical summary estimates of treatment effect of −0.33 with a 95% confidence interval of −0.48 to −0.18, but the first uses a fixed effect model and the second a random effects model. In the following two sections we explain why the summary result should be interpreted differently in these two examples because of the different meta-analysis models they use.

Study	Standardised mean difference (95% CI)	Standardised mean difference (95% CI)
Fixed effects		
1		-0.49 (-1.17 to 0.19)
2		-0.17 (-0.59 to 0.25)
3		-0.52 (-0.99 to -0.05)
4		-0.48 (-1.21 to 0.25)
5		-0.26 (-0.75 to 0.23)
6		-0.36 (-0.94 to 0.22)
7		-0.47 (-0.90 to -0.04)
8		-0.30 (-0.59 to -0.01)
9		-0.15 (-0.68 to 0.38)
10		-0.28 (-1.26 to 0.70)
Summary result		-0.33 (-0.48 to -0.18)
Random effects		
11		0.00 (-0.83 to 0.83)
12		0.10 (-0.33 to 0.53)
13		-0.40 (-0.45 to -0.35)
14		-0.80 (-1.19 to -0.41)
15		-0.63 (-1.22 to -0.04)
16		0.22 (-0.37 to 0.81)
17		-0.34 (-0.48 to -0.20)
18		-0.51 (-0.71 to -0.31)
19		0.03 (-0.21 to 0.27)
20		-0.81 (-1.40 to -0.22)
Summary result		-0.33 (-0.48 to -0.18)

-1.5 -1.0 -0.5 0 0.5 1.0
Treatment better   Treatment worse

Fig 1 | Forest plots of two distinct hypothetical meta-analyses that give the same summary estimate (centre of diamond) and its 95% confidence interval (width of diamond). In the fixed effect meta-analysis (top) the summary result provided the best estimate of an assumed common treatment effect. In the random effects meta-analysis (bottom) the summary result gives the average from the distribution of treatment effects across studies

---

SUMMARY POINTS

Meta-analysis combines the study estimates of a particular effect of interest, such as a treatment effect

Fixed effect meta-analysis assumes a common treatment effect in each study and variation in observed study estimates is due only to chance

Random effects meta-analysis assumes the true treatment effect differs from study to study and provides an estimate of the average treatment effect

Interpretation of random effects meta-analysis is aided by a prediction interval, which provides a predicted range for the true treatment effect in an individual study

How to calculate a prediction interval

A random effects meta-analysis combines the study estimates of a parameter of interest (eg, a treatment effect) in order to estimate the average value (denoted by μ) and the standard deviation of the parameter (denoted by τ) across studies. The term random denotes that the μ value is not common to all studies, and that the parameter value in an individual study can vary randomly about the average value due to unexplained heterogeneity. Methods to estimate μ and τ are discussed elsewhere.[2]

After a random effects meta-analysis, a prediction interval can be calculated to give a range for the predicted parameter value in a new study. Assuming the random effects (that is, the individual study parameter values) are normally distributed with between-study standard deviation (τ), then the prediction interval is approximately[6]

$$\hat{\mu} - t_{k-2}\sqrt{\hat{\tau}^2 + SE(\hat{\mu})^2} \; , \quad \hat{\mu} + t_{k-2}\sqrt{\hat{\tau}^2 + SE(\hat{\mu})^2}$$

where $\hat{\mu}$ is the estimate of the average parameter value across studies; $SE(\hat{\mu})$ is the standard error of $\hat{\mu}$; $\hat{\tau}$ is the estimate of between study standard deviation; t_{k-2} is the $100(1-\alpha/2)$ percentile of the t distribution with $k-2$ degrees of freedom, where k is the number of studies in the meta-analysis and α is usually chosen as 0.05, to give a 5% significance level and thus 95% prediction interval. A t distribution, rather than a normal distribution, is used to help account for the uncertainty of $\hat{\tau}$. The correct number of degrees of freedom for this t distribution is complex, and we use a value of $k-2$ largely for pragmatic reasons.

As an example computation, for the fibromyalgia syndrome meta-analysis (fig 3) the effect of interest is the standardised mean difference in pain (for the antidepressant group minus the control group) and $\hat{\mu} = -0.425$, $SE(\hat{\mu}) = 0.063$, $\hat{\tau} = 0.18$, and a $t_{k-2}^{\alpha} = t_{20}^{0.05} = 2.086$, leading to a 95% prediction interval of -0.83 to -0.02.

Finally, it is important to work on a scale that helps meet the normal assumption for the random effects. In particular, when the parameter under investigation is a ratio measure (such as a relative risk or odds ratio), then $\hat{\mu}$, $\hat{\tau}$ and the subsequent prediction interval are best derived on the natural log scale and the results subsequently exponentiated back to the original scale.

Fig 2

Fixed effect meta-analysis

Use of a fixed effect meta-analysis model assumes all studies are estimating the same (common) treatment effect. In other words, there is no between study heterogeneity in the true treatment effect. The implication of this model is that the observed treatment effect estimates vary only because of chance differences created from sampling patients. Hypothetically, if all studies had an infinite sample size, there would be no differences due to chance and the differences in study estimates would completely disappear.

I^2 measures the percentage of variability in treatment effect estimates that is due to between study heterogeneity rather than chance.[3] I^2 is 0% in our fixed effect meta-analysis example, suggesting the variability in study estimates is entirely due to chance. This is visually evident by the narrow scatter of effect estimates with large overlap in their confidence intervals (fig 1, top). The summary result of -0.33 (95% confidence interval of -0.48 to -0.18) in our example thus provides the best estimate of a common treatment effect, and the confidence interval depicts the uncertainty around this estimate. As the confidence interval does not contain zero, there is strong evidence that the treatment is effective.

Random effects meta-analysis

A random-effects meta-analysis model assumes the observed estimates of treatment effect can vary across studies because of real differences in the treatment effect in each study as well as sampling variability (chance). Thus, even if all studies had an infinitely large sample size, the observed study effects would still vary because of the real differences in treatment effects. Such heterogeneity in treatment effects is caused by differences in study populations (such as age of patients), interventions received (such as dose of drug), follow-up length, and other factors.

In the random effects example in figure 1, I^2 is 71%, suggesting 71% of the variability in treatment effect estimates is due to real study differences (heterogeneity) and only 29% due to chance.[3] This is visually evident from the wide scatter of effect estimates with little overlap in their confidence intervals, in contrast to the fixed effect example (fig 1). The random effects model summary result of -0.33 (95% confidence interval -0.48 to -0.18) provides an estimate of the average treatment effect, and the confidence interval depicts the uncertainty around this estimate. As the confidence interval does not contain zero, there is strong evidence that on average the treatment effect is beneficial.

Use and interpretation of meta-analysis in practice

Unfortunately, meta-analysis results are often interpreted in the same manner regardless of whether a fixed effect or random effects model is used. We reviewed 44 Cochrane reviews that each reported a random effects meta-analysis and found that none correctly interpreted the summary result as an estimate of the average effect rather than the common effect.[4] Furthermore, only one indicated why the summary result from a random effects meta-analysis was clinically meaningful,[5] arguing that, although real study differences (heterogeneity) in treatment effects existed (because of different doses), the studies were reasonably clinically comparable as the same drug was used and patient characteristics were similar.

Another problem is that a fixed effect meta-analysis model is often used even when heterogeneity is present. We examined 31 Cochrane reviews that did not use a random effects model and found that 26 had potentially moderate or large heterogeneity between studies ($I^2 > 25\%$ as a guide[3]) yet still used a fixed effect model, without justifying why.[4] Ignoring heterogeneity leads to an overly precise summary result (that is, the confidence interval is too narrow) and may wrongly imply that a common treatment effect exists when actually there are real differences in treatment effectiveness across studies.

Benefits of using prediction intervals

After a random effects meta-analysis, researchers usually focus on the average treatment effect estimate and its confidence interval. However, it is important also to consider the potential effect of treatment when it is applied within an individual study setting, as this may be different from the average effect. This can be achieved by calculating a prediction interval (fig 2).[6]

Intervals akin to prediction intervals are commonly used in other areas of medicine. For example, when considering the blood pressure of an individual or the birthweight of an infant, we not only compare it with the average value but also with a reference range (prediction interval) for blood pressure or birthweight across the population. In the meta-analysis setting, our measures are treatment effects, and we work at the study level (rather than the individual level) with a population of study effects. We therefore can report the range of effects across study settings, providing a more complete picture for clinical practice. For instance, consider the random effects analysis in figure 1 again, for which the 95% prediction interval is -0.76 to 0.09. Although most of this interval is below zero, indicating the treatment will be
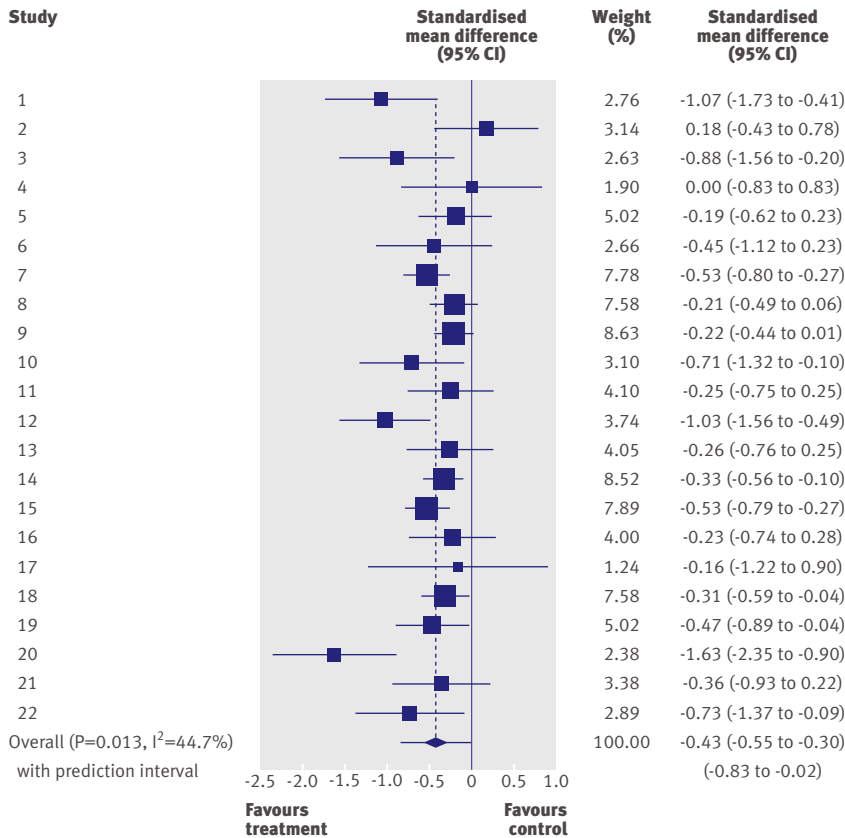
Study	Standardised mean difference (95% CI)	Weight (%)	Standardised mean difference (95% CI)
1		2.76	-1.07 (-1.73 to -0.41)
2		3.14	0.18 (-0.43 to 0.78)
3		2.63	-0.88 (-1.56 to -0.20)
4		1.90	0.00 (-0.83 to 0.83)
5		5.02	-0.19 (-0.62 to 0.23)
6		2.66	-0.45 (-1.12 to 0.23)
7		7.78	-0.53 (-0.80 to -0.27)
8		7.58	-0.21 (-0.49 to 0.06)
9		8.63	-0.22 (-0.44 to 0.01)
10		3.10	-0.71 (-1.32 to -0.10)
11		4.10	-0.25 (-0.75 to 0.25)
12		3.74	-1.03 (-1.56 to -0.49)
13		4.05	-0.26 (-0.76 to 0.25)
14		8.52	-0.33 (-0.56 to -0.10)
15		7.89	-0.53 (-0.79 to -0.27)
16		4.00	-0.23 (-0.74 to 0.28)
17		1.24	-0.16 (-1.22 to 0.90)
18		7.58	-0.31 (-0.59 to -0.04)
19		5.02	-0.47 (-0.89 to -0.04)
20		2.38	-1.63 (-2.35 to -0.90)
21		3.38	-0.36 (-0.93 to 0.22)
22		2.89	-0.73 (-1.37 to -0.09)
Overall (P=0.013, I^2=44.7%) with prediction interval		100.00	-0.43 (-0.55 to -0.30) (-0.83 to -0.02)

-2.5 -2.0 -1.5 -1.0 -0.5 0 0.5 1.0

Favours treatment   Favours control

Fig 3 | Random effects meta-analysis of 22 studies that examine the effect of antidepressants on reducing pain in patients with fibromyalgia syndrome[8]

treatment effects (often denoted by the Greek letter τ), and the uncertainty in the between study standard deviation estimate itself.[6] It can be calculated when the meta-analysis contains at least three studies, although the interval may be very wide with so few studies. A prediction interval will be most appropriate when the studies included in the meta-analysis have a low risk of bias.[7] Otherwise, it will encompass heterogeneity in treatment effects caused by these biases, in addition to that caused by genuine clinical differences.

Examples

Antidepressants for reducing pain in fibromyalgia syndrome

Hauser and colleagues report a meta-analysis of randomised trials to determine the efficacy of antidepressants for fibromyalgia syndrome, a chronic pain disorder associated with multiple debilitating symptoms.[8] Twenty two estimates of the standardised mean difference in pain (for the antidepressant group minus the control group) were available from the included trials (fig 3), with negative values indicating a benefit for antidepressants. Studies used different classes of antidepressants, and other clinical and methodological differences also existed, resulting in large between study heterogeneity in treatment effect (I^2=45%; between study standard deviation estimate=0.18). The authors therefore used a random effects meta-analysis and obtained a summary result of −0.43 (95% confidence interval −0.55 to −0.30), concluding that "antidepressant medications are associated with improvements in pain."

The summary result here relates to the average effect of antidepressants across the trials. As the confidence interval is below zero, it provides strong evidence that on average antidepressants are beneficial; however, it does not indicate whether antidepressants are always beneficial. The authors acknowledge the heterogeneity of treatment effects but conclude that "although study effect sizes differed, results were mostly consistent." This can be quantified more formally by a 95% prediction interval, which we calculated as −0.83 to −0.02 (fig 2). This interval is entirely below zero and shows that antidepressants will be beneficial when applied in at least 95% of the individual study settings, an important finding for clinical practice.

Inpatient rehabilitation in geriatric patients

Bachmann and colleagues did a random effects meta-analysis of 12 randomised trials to summarise the effect of inpatient rehabilitation compared with usual care on functional outcome in geriatric patients (fig 4).[9] The summary odds ratio estimate is 1.36 (1.07 to 1.71), which indicates that the average effect of the intervention is to make the odds of functional improvement 1.36 times higher than usual care. As the confidence interval is above one, it provides strong evidence that the average intervention effect is beneficial.

However, there is large between study heterogeneity in intervention effect (I^2=51%; between study standard deviation estimate=0.27), possibly because of differences in the type of intervention used (such as general or orthopaedic rehabilitation) and length of follow-up, among other factors. Responding to the heterogeneity, the authors state: "Pooled effects should be interpreted with caution because the true differences in effects between studies might be due to uncharacterised or unexplained underlying factors or

beneficial in most settings, the interval overlaps zero and so in some settings the treatment may actually be ineffective. This finding was masked when we focused only on the average effect and its confidence interval.

A prediction interval can be provided at the bottom of a forest plot (fig 3). It is centred at the summary estimate, and its width accounts for the uncertainty of the summary estimate, the estimate of between study standard deviation in the true

Study	Odds ratio (95% CI)	Weight (%)	Odds ratio (95% CI)
1		6.08	1.11 (0.51 to 2.39)
2		15.94	0.97 (0.78 to 1.21)
3		11.38	1.13 (0.73 to 1.72)
4		4.60	1.08 (0.42 to 2.75)
5		5.75	0.88 (0.39 to 1.95)
6		8.48	1.28 (0.71 to 2.30)
7		9.07	1.19 (0.69 to 2.08)
8		4.03	3.82 (1.37 to 10.60)
9		9.58	1.06 (0.63 to 1.79)
10		7.59	2.95 (1.54 to 5.63)
11		7.00	2.36 (1.18 to 4.72)
12		10.51	1.68 (1.05 to 2.70)
Overall (P=0.021, I^2=51.2%) with prediction interval		100.00	1.36 (1.08 to 1.71) (0.70 to 2.64)

0.5 0.75 1   2   5   10
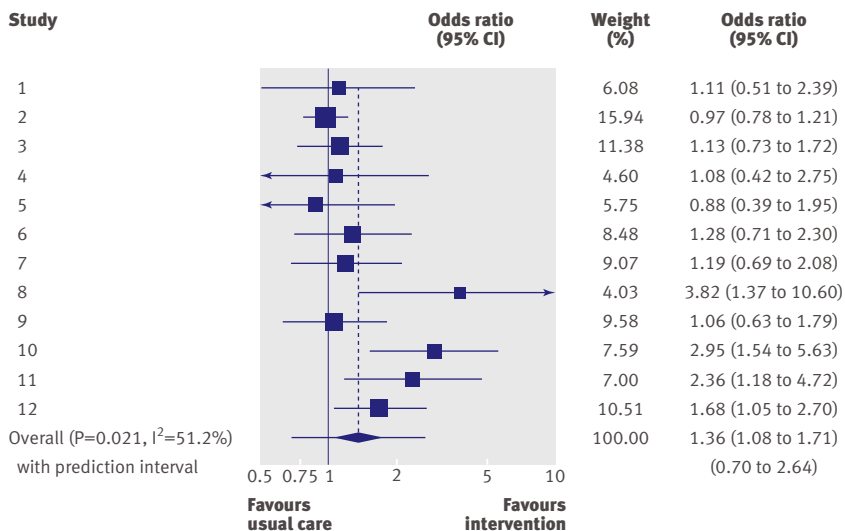
Favours usual care   Favours intervention

Fig 4 | Random effects meta-analysis of 12 trials that examine the effect of inpatient rehabilitation designed for geriatric patients versus usual care on improving functional outcome[9]

the variability of outcome measures on functional status."[9] This cautionary note can be quantified by presenting a 95% prediction interval, which we calculate as 0.70 to 2.64. This interval contains values below 1 and so, although on average the intervention seems effective, it may not always be beneficial in an individual setting. Further research is needed to identify causes of the heterogeneity, in particular the subtypes of geriatric rehabilitation programmes that work best and the subgroups of patients that benefit most.

Discussion

Between study heterogeneity in treatment effects is a common problem for meta-analysts. Although it is desirable to identify the causes of heterogeneity (by using meta-regression or subgroup analyses, for example),[10] this is often not practically possible.[11] [12] For instance, there may be too few studies to examine heterogeneity reliably; no prespecified idea of what factors might cause heterogeneity; or a lack of necessary information (such as no individual participant data[13]). Even when factors causing heterogeneity are identified, unexplained heterogeneity may remain. Thus random effects meta-analysis, which accounts for unexplained heterogeneity, will continue to be prominent in the medical literature. Including a prediction interval, which indicates the possible treatment effect in an individual setting, will make these analyses more useful in clinical practice and decision making.[14] Although our examples focused on the synthesis of randomised trials, prediction intervals can also be used in other meta-analysis settings such as studies of diagnostic test accuracy[15] and prognostic biomarkers.[16]

1 Egger M, Davey Smith G. Meta-analysis: Potentials and promise. BMJ 1997;315:1371-4.
2 Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Methods for Meta-analysis in medical research. John Wiley, 2000.
3 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
4 Riley RD, Gates SG, Neilson J, Alfirevic Z. Statistical methods can be improved within Cochrane pregnancy and childbirth reviews. J Clin Epidemiol 2010 Nov 24. [Epub ahead of print].
5 King J, Flenady V, Cole S, Thornton S. Cyclo-oxygenase (COX) inhibitors for treating preterm labour. Cochrane Database Syst Rev 2005;2:CD001992.
6 Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A 2009;172:137-59.
7 Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions. John Wiley, 2008.
8 Häuser W, Bernardy K, Üceyler N, Sommer S. Treatment of fibromyalgia syndrome with antidepressants: a meta-analysis. JAMA 2009;301:198-209.
9 Bachmann S, Finger C, Huss A, Egger M, Stuck AE, Clough-Gorr KM. Inpatient rehabilitation specifically designed for geriatric patients: systematic review and meta-analysis of randomised controlled trials. BMJ 2010;340:c1718.
10 Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. BMJ 1994;309:1351-5.
11 Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002;21:1559-74.
12 Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. J Health Serv Res Policy 2002;7:51-61.
13 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: conduct, rationale and reporting. BMJ 2010;340:c221.
14 Ades AE, Lu G, Higgins JPT. The interpretation of random-effects meta-analysis in decision models. Med Decis Making 2005;25:646-54.
15 Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982-90.
16 Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: the evolution of evidence from single studies to meta-analysis, and beyond. Br J Cancer 2009;100:1219-29.

The problem with publications

It is drummed into medical students and doctors across the world that, to get ahead, you need to get published. "Getting published" is often used as a catch-all term to refer to any writing that might boost academic credentials—presentations, leaflets, full papers, abstracts, guidelines, and so on.

In my first foundation year, I was no different from any other junior struggling to pull myself out of that cohort of trainees who wouldn't get a training post. I was perhaps an overenthusiastic house officer, and I suggested to my then supervisor that I write up a case for a journal.

He didn't see the publication potential that I did. So I went it alone, eventually presenting the case report at a national conference. I had exceeded my supervisor's expectations, and naturally I was pleased. It was exciting; the case was fairly unusual, and I really felt that I was changing the face of medicine from my very small platform.

As is so often the case, however, my happiness didn't last. It seems there is such a thing as too much publication: the case was so interesting that the conference organisers suggested sending out a press release to the national media. Naively, I found all the media talk very exciting,

but it didn't take long for me to realise the full implications of what was suggested. The sensitivity of the case combined with the high level of interest and exposure could destroy the anonymity of the patient whose case I had described.

I felt out of my depth, and it was only after seeking medicolegal advice, rigorous scrutiny of the original patient consent form, and gaining the opinion of a colleague who had found himself in a similar position that the situation was eventually rectified. My case was left alone, but I had come perilously close to things going terribly wrong—not just for myself, but also for the patient involved.

It's all too easy to publish findings from research and forget the information source. What does it mean for those patients affected? It doesn't hurt to take a cautionary glance both ways before stepping into the limelight. As is often advocated, prevention is better than cure.

Esohe C Omoregie foundation trainee, East Kent University Hospital Trust, Canterbury, UK
ecomoregie@gmail.com

Cite this as: BMJ 2010;341:c6252