Bandolier bias guide

The *Bandolier* bias guide is reproduced from *Bandolier* 80 because people have asked for a one-off printable version.

Bandolier has been struck of late, "many a time and oft'", by the continuing and cavalier attitude towards bias in clinical trials. We know that the way that clinical trials are designed and conducted can influence their results. Yet people still ignore known sources of bias when making decisions about treatments at all levels.

What is bias?

A dictionary definition of bias is "a one-sided inclination of the mind". In our business it defines a systematic disposition of certain trial designs to produce results consistently better or worse than other trial designs.

Garbage in, garbage out

For the avoidance of doubt, the clinical bottom line is that wherever bias is found it results in a large over-estimation of the effect of treatments. Poor trial design makes treatments look better than they really are. It can even make them look as if they work when actually they do not work.

This is why good guides to systematic review suggest strategies for bias minimisation by avoiding including trials with known sources of bias. They further suggest performing sensitivity analysis to see whether different trial designs are affecting results in a systematic review.

But this advice is ignored more often than not. It is ignored in reviews, and it is ignored in decision-making. The result is that decisions are being made on incorrect information, and they will be wrong.

Bandolier bias guide

Bandolier has therefore decided to revisit some of the words written on bias in these pages and elsewhere, and collect them into one handy reference guide. The guide can be used when examining a systematic review, or a single clinical trial. The guide is not to be used for observational studies, or for studies of diagnostic tests.

Randomisation

The process of randomisation is important in eliminating selection bias in trials. If the selection is done by a computer, or even the toss of a coin, then any conscious or subconscious attitude of the researcher is avoided.

Some of the most influential people in evidence-based thinking showed how inadequate design exaggerated the effect measured in a trial (Table). They compared trials in which the authors reported adequately concealed treatment allocations with those in which randomisation was either inadequate or unclearly described, as well as examining the effects of exclusions and binding.

The results were striking and sobering, as the Table on page 2 shows. Odds ratios were exaggerated by 41% in trials where treatment allocation was inadequately concealed, and by 30% when the process of allocation concealment was not clearly described.

Figure 1: Effect of randomisation on outcome of trials of TENS in acute pain

www.ebandolier.com
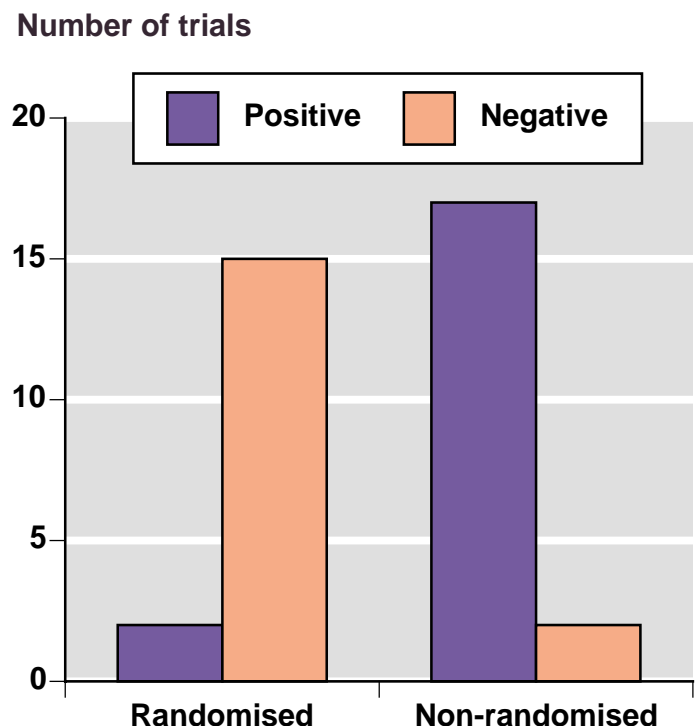
Table: Examples of known bias in trials of treatment efficacy

Source of bias	Effect on treatment efficacy	Size of the effect	References
Randomisation	Increase	Non-randomised studies overestimate treatment effect by 41% with inadequate method, 30% with unclear method	KF Schultz, I Chalmers, RJ Hayes, DG Altman. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Journal of the American Medical Association 1995 273: 408-12.
Randomisation	Increase	Completely different result between randomised and non-randomised studies	Carroll D, Tramèr M, McQuay H, Nye B, Moore A. Randomization is important in studies with pain outcomes: systematic review of transcutaneous electrical nerve stimulation in acute postoperative pain. British Journal of Anaesthesia 1996; 77: 798-803.
Blinding	Increase	17%	KF Schultz, I Chalmers, RJ Hayes, DG Altman. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Journal of the American Medical Association 1995 273: 408-12.
Blinding	Increase	Completely different result between blind and non-blind studies	Ernst E, White AR. Acupuncture for back pain: A meta-analysis of randomised controlled trials. Arch Int Med 1998, 158: 2235-2241.
Reporting quality	Increase	About 25%	Khan KS, Daya S, Jadad AR. The importance of quality of primary studies in producing unbiased systematic reviews. Arch Intern Med 1996,156 :661-6. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of interventi
Duplication	Increase	About 20%	Tramèr M, Reynolds DJM, Moore RA, McQuay HJ. Effect of covert duplicate publication on meta-analysis; a case study. BMJ 1997, 315: 635-40.
Geography	Increase	May be large for some alternative therapies	Vickers A, Goyal N, Harland R, Rees R. Do certain countries produce only positive results? A systematic review of controlled trials. Control Clin Trial 1998, 19: 159-166.
Size	Increase	Small trials may overestimate treatment effects by about 30%	Moore RA, Carroll D, Wiffen PJ, Tramèr M, McQuay HJ. Quantitative systematic review of topically-applied non-steroidal anti-inflammatory drugs. BMJ 1998, 316: 333-8. Moore RA, Gavaghan D, Tramèr MR, Collins SL, McQuay HJ. Size is everything - large amounts of information are needed to overcome random effects in estimating direction and magnitude of treatment effects. Pain 1998, 78: 217-220.
Statistical	Increase	Not known to any extent, probably modest, but important especially where vote-counting occurs	Smith LA, Oldman AD, McQuay HJ, Moore RA. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. Pain 2000, 86: 119-132.
Validity	Increase	Not known to any extent, probably modest, but important especially where vote-counting occurs	Smith LA, Oldman AD, McQuay HJ, Moore RA. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. Pain 2000, 86: 119-132.
Language	Increase	Not known to any extent, but may be modest	Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German, Lancet 1997 350: 326-329.
Publication	Increase	Not known to any extent, probably modest, but important especially where there is little evidence	M Egger, G Davey Smith. Under the meta-scope: potentials and limitations of meta-analysis. In M Tramèr, Ed. Evidence Based Resource in Anaesthesia and Analgesia. BMJ Publications, 2000.

Many systematic reviews exclude non-randomised trials because of the amount of bias arising from failure to randomise. *Bandolier* believes that restricting systematic reviews to include only randomised studies makes sense for reviews of treatment efficacy. The reason is the many, many examples where non-randomised studies have led reviews to come to the wrong conclusion.

Examples abound. A classic example (*Bandolier* 37) is a review of transcutaneous nerve stimulation (TENS) for post-operative pain relief (Figure 1). Randomised studies overwhelmingly showed no benefit over placebo, while non-randomised studies did show benefit. Particularly where a review counts votes (a study is positive or negative) rather than combines data in a meta-analysis the randomisation effect is strong. It applies particularly to studies in alternative therapies.

Blinding

The importance of blinding is that it avoids observer bias. If no-one knows which treatment a patient has received, then no systematic over-estimation of the effect of any particular treatment is possible.

Non-blinded studies over-estimate treatment effects by about 17% (Table). In a review of acupuncture for back pain (Figure 2), including both blinded and non-blinded studies changed the overall conclusion (*Bandolier* 60). The blinded studies showed 57% of patients improved with acupuncture and 50% with control, a relative benefit of 1.2 (95% confidence interval 0.9 to 1.5). Five non-blinded studies showed a difference from control, with 67% improved with acupuncture and 38% with control. Here the relative benefit was significant at 1.8 (1.3 to 2.4).

Figure 2: Effect of blinding on outcome of trials of acupuncture for chronic back pain

Percent with short term improvement
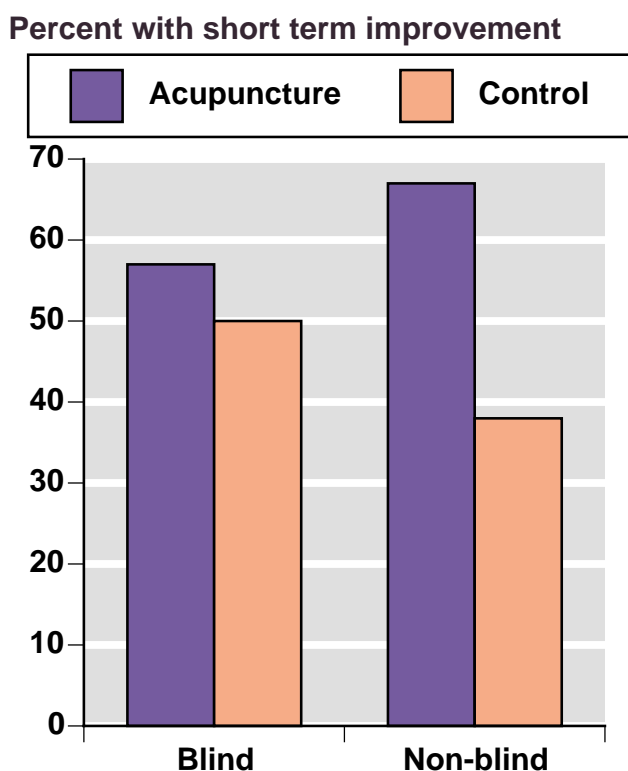
Reporting quality

Because of the large bias expected from studies which are not randomised or not blind, a scoring system [1] that is highly dependent on randomisation and blinding will also correlate with bias. Trials of poor reporting quality consistently over estimate the effect of treatment (Table). This particular scoring system has a range of 0 to 5 based on randomisation, blinding and withdrawals and dropouts. Studies scoring 2 or less consistently show greater effects of treatment than those scoring 3 or more.

Duplication

Results from some trials are reported more than once. This may be entirely justified for a whole range of reasons. Examples might be a later follow up of the trial, or a re-analysis. Sometimes, though, information about patients in trials is reported more than once without that being obvious, or overt, or referenced. Only the more impressive information seems to be duplicated, sometimes in papers with completely different authors. A consequence of covert duplication would be to overestimate the effect of treatment (Table).

Geography

In *Bandolier* 71 we reported on how geography can be a source of bias in systematic reviews. Vickers and colleagues (Table) showed that trials of acupuncture conducted in east Asia were universally positive, while those conducted in Australia/New Zealand, north America or western Europe were positive only about half the time. Randomised trials of therapies other than acupuncture conducted in China, Taiwan, Japan or Russia/USSR were also overwhelmingly positive, and much more so than in other parts of the world. This may be a result of an historical cultural difference, but it does mean that care should be exercised where there is a preponderance of studies from these cultures. Again, this is particularly important for alternative therapies.

Size

Clinical trials should have a power calculation performed at the design stage. This will estimate how many patients are needed so that, say, 90% of studies with X number of patients would show a difference of Y% between two treatments. When the value of Y is very large, the value of X can be small. More often the value of Y is modest, or small. In those circumstances, X needs to be larger, and more patients will be needed in trials for them to have a hope of showing a difference.

Yet clinical trials are often ridiculously small. *Bandolier*'s record is a randomised study on three patients in a parallel group design. But when are trials so tiny that they can be ignored? Many folk take a pragmatic view that trials with fewer than 10 patients per treatment arm should be ignored, though others may disagree.

There are examples where sensitivity in meta-analysis has shown small trials to have a larger effect of treatment than

Figure 3: Trials of ibuprofen in acute pain that are randomised, double blind, and with the same outcomes over the same time in patients with the same initial pain intensity

Adequate pain relief with ibuprofen 400 mg (%)

Adequate pain relief with placebo (%)
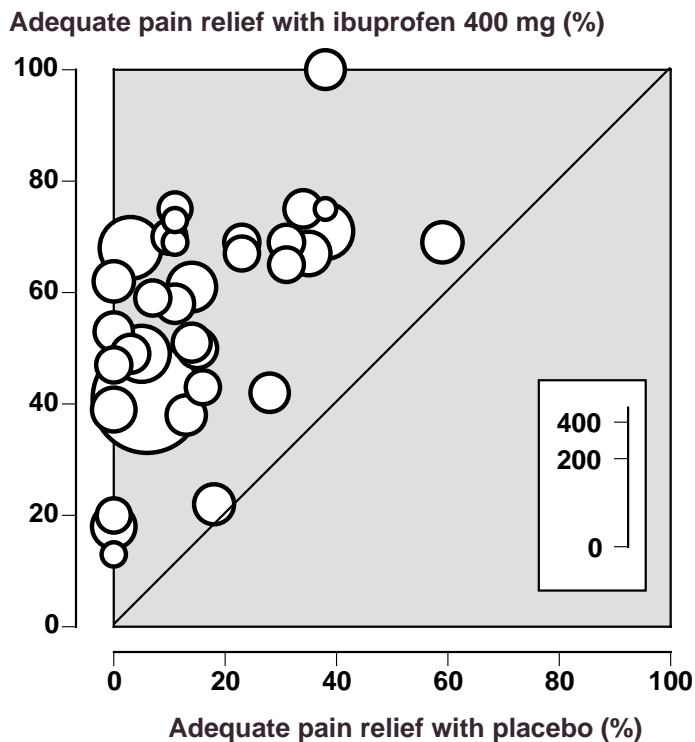
Figure 4: Computer model of trials of ibuprofen in acute pain. Intensity of colour matches probability of outcome of a single trial
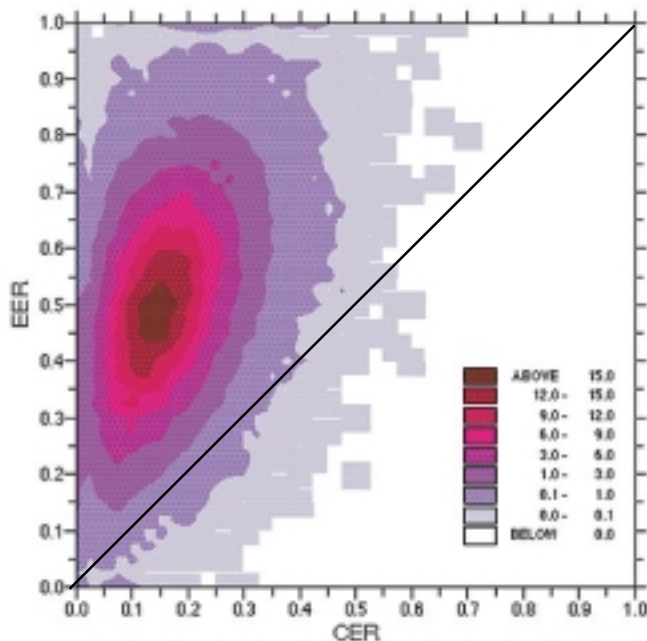
larger trials (Table). The degree of variability between trials of adequate power is still large, because trials are powered to detect that there is a *difference* between treatments, rather than *how big* that difference is.

The random play of chance can remain a significant factor despite adequate power to detect a difference. Figure 3 shows the randomised, double blind studies comparing ibuprofen 400 mg with placebo in acute postoperative pain. The trials had the same patient population, with identical initial pain intensity and with identical outcomes measured in the same way for the same time using standard measuring techniques. There were big differences in the outcomes of individual studies.

Figure 4 shows the results of 10,000 studies in a computer model based on information from about 5,000 individual patients [2]. Anywhere in the gray area is where a study could occur just because of the random play of chance. And for those who may think that this reflects on pain as a subjective outcome, the same variability can be seen in other trial settings, with objective outcomes.

Statistics, data manipulation and outcomes

Despite the best efforts of editors and peer reviewers, some papers are published that are just plain wrong. Wrong covers a multitude of sins, but two are particularly important.

Statistical incorrectness can take a variety of guises. It may be as simple as the data presented in a paper as statistically significant not being significant. It can often take the form of inappropriate statistical tests. It can be data trawling, where a single statistical significance is obtained and a paper written round it. Reams could be written about this, but the simple warning is that readers or reviewers of papers have to be cautious of results of trials, especially where vote-counting is being done.

But also beware the power of words. Even when statistical testing shows no difference, it is common to see the results hailed as a success. While that may sound silly when written down, even the most cynical of readers can be fooled into drawing the wrong conclusion. Abstracts are notorious for misleading in this way.

Data manipulation is a bit more complicated to detect. An example would be an intervention where we are not told what the start condition of patients is, nor the end, but that at some time in between the rate of change was statistically significant by some test with which we are unfamiliar. This is done only to make positive that which is not positive, and the direction of the bias is obvious (Table). Again, crucially important where vote counting is being done to determine whether the intervention works or not.

Outcomes reported in trials are an even more sticky problem. It is not infrequent that surrogate measures are used rather than an outcome of real clinical importance. Unless these surrogate measures are known unequivocally to correlate with clinical outcomes of importance, then an unjust sense of effectiveness could be implied or assumed.

Validity

Do individual trials have a design (apart from issues like randomisation and blinding) that allows them to adequately measure an effect? What constitutes validity depends on the circumstances of a trial, but studies often lack validity. A validity scoring system applied to acupuncture for back and neck pain demonstrated that trials with lower validity were more likely to say that the treatment worked than those that were valid (Table).

Language

Too often the search strategy for a systematic review or meta-analysis restricts itself to the English language only. Authors whose language is not English may be more likely to publish positive findings in an English language journal, because these would have a greater international impact. Negative findings would be more likely to be published in non-English language journals (Table).

Publication

Finally there is the old chestnut of publication bias. This is usually thought to be the propensity for positive trials to be published and for negative trials not to be published. It must exist, and there is a huge literature about publication bias.

Bandolier has some reservations about the fuss that is made, though. Partly this stems from the failure to include assessments of trial validity and quality. Most peer reviewers would reject non-randomised studies, or those where there are major failings in methodology. These trials will be hard to publish. Much the same can be said for dissertations or theses. One attempt to include theses [3] found 17 dissertations for one treatment. Thirteen were excluded because of methodological problems, mainly lack of randomisation, three had been published and were already included in the relevant review, and one could be added. It made no difference.

Bandolier is also sceptical that funnel plots are in any way helpful. One often quoted, of magnesium in acute myocardial infarction [4], can more easily be explained by the fact that trials in a meta-analysis were trivially small to detect any effect and should never have been included in a meta-analysis in the first place.

But these are quibbles. If there is sufficient evidence available, large numbers of large, well conducted trials, then publication bias is not likely to be a problem. Where there is little information, small numbers of low quality trials, then it becomes more problematical.

Comment

This is but a brief review of some sources of bias in trials of treatment efficacy. Others choose to highlight different sources of potential bias. That bias is present, and exists in so many different forms is why we have to be vigilant when reading about a clinical trial, and especially when taking the results of a single trial into clinical practice.

But systematic reviews and meta-analyses also suffer from quality problems. They should consider potential sources of bias when they are being written. Many do not, and will therefore mislead. If systematic reviews or meta-analyses include poor trials or have poor reporting quality, then, just like individual trials, they too have a propensity a greater likelihood of a positive result [4,5].

There is no doubt that meta-analyses can mislead. If they do, then it is because they have been incorrectly assembled or incorrectly used. The defence, indeed the only defence, is for readers to have sufficient knowledge themselves to know when the review or paper they are reading should be confined to the bin.

References:

1   Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, McQuay HJ. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trial 1996, 17: 1-12.
2   Moore RA, Gavaghan D, Tramèr MR, Collins SL, McQuay HJ. Size is everything - large amounts of information are needed to overcome random effects in estimating direction and magnitude of treatment effects. Pain 1998, 78: 217-220.
3   A Vickers, C Smith. Incorporating data from dissertations in systematic reviews. Int J Technol Assess Health Care 2000 16:2: 711-713.
4   Jadad AR, McQuay HJ. Meta-analysis to evaluate analgesic interventions: a systematic qualitative review of the literature. J Clin Epidemiol 1996, 49:235-243.
5   Smith L, Oldman A. Acupuncture and dental pain. Br Dent J 1999, 186: 158-159.