

## EVIDENCE AND MIGRAINE TRIALS

"Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients." [1]

This quotation from Dave Sackett and his colleagues is as good a place as any to start thinking about evidence and migraine trials and treatments. The full article goes beyond this definition and includes patient and societal values. The main issue, though, is about where the practitioner goes to find "current best evidence". It could be using local or national guidelines, as for instance produced by organisations like the National Institute of Clinical Excellence, or those produced by eminent bodies. Some people will remain sceptical, though, and will (and should) satisfy themselves that the evidence on which guidelines are based is sound.

## Information, knowledge and wisdom

In the past that task was difficult. With millions of papers being published each year (there is said to be about 30,000 medical journals), trying to find information, especially all the information was a heroic task. Now it is much easier. We can search PubMed online or visit electronic journals like BioMed, or electronic versions of paper journals like the BMJ. The Cochrane Library, available online or on CD for a small subscription has not only many good reviews, but also has over 250,000 controlled trials found by hand-searching the literature.

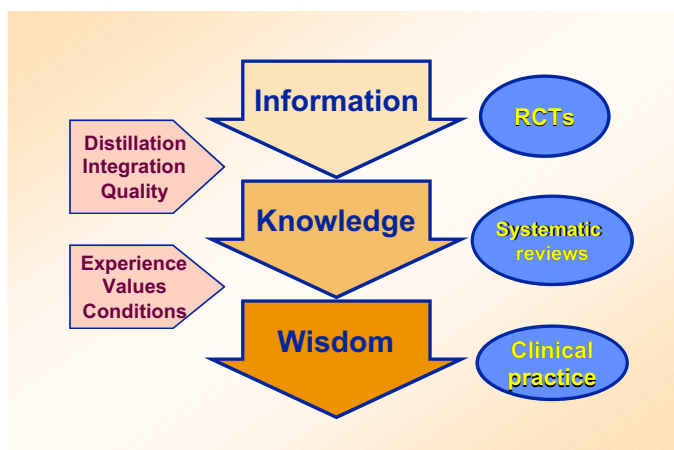
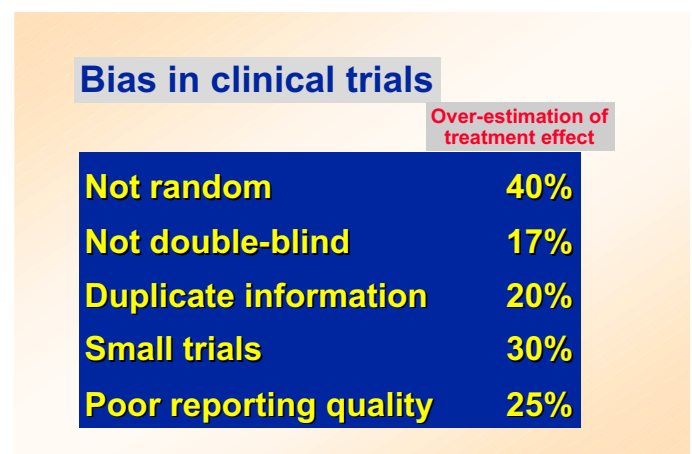
Good systematic reviews are increasingly available, where someone has asked a clinical question, and then summarised all the known information into a solid piece of knowledge. In doing so they will distill the information, perhaps integrate different types of information, and use quality fil-

ters so that only the most reliable information is used and that unreliable information is discarded.

How that knowledge is used depends on the practitioner making the conscientious, explicit and judicious use of the knowledge, in terms of the unique biology of the patient, incorporation patient concerns, their own experience and local knowledge, the values of society and the conditions in which they are working. The same piece of knowledge will play differently in Cardiff or Calcutta. That's the wisdom bit. That is why evidence-based approaches have nothing to do with rules, but should be seen as tools to allow practitioners to be better, and patients to be better informed.

## Bias in clinical trials

One of the things we have learned through doing systematic reviews (also called meta-analysis when we pool data and do some sums) has been that certain types of study architectures are likely to produce results that are more favourable to a new treatment than they should be [2]. This is called bias, and many forms of bias have been discovered. We know that trials that are not randomised overestimate the size of a treatment effect, as do trials that are not blind, or where information from patients is duplicated [3], or where trials are small [4], or where they have poor reporting quality [5,6].



We can be much more specific. For instance, in a study of transcutaneous electrical nerve stimulation in postoperative pain, 17 of 19 trials that were not randomised came up with a positive result, while 15 of 17 randomised trials came up with the completely opposite result, that it did not work [7].

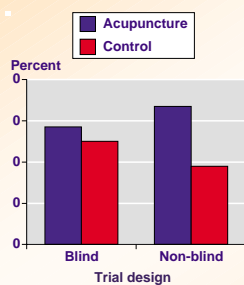
## TENS in Postoperative Pain

	Analgesic result	
	Positive	Negative
Randomised	2	15
Inadequate or not randomised	17	2

Carroll et al BJA 1996; 77: 798-803

In a review of acupuncture in back pain, lumping together all randomised trials, whether blinded or open, came up with the result that acupuncture worked for back pain. When you look at the open studies, where people making the assessments knew who had true acupuncture and who did not, there was a striking difference. But when you look at only the blinded studies, where people making the assessments did not know the treatment used, there was no difference at all. Acupuncture does not work.

## Acupuncture in back pain



- Randomised studies
- Compared with sham acupuncture
- Validity of acupuncture checked independently
- 12 trials in total, 9 with outcomes
- 4 were blind (250 patients)
- 5 were not blind (204 patients)

Ernst & White, Arch Intern Med 1998 158:2235-41

So attending to bias is an important issue in systematic reviews or meta-analysis of treatments. Where bias is known or likely to exist, then we may come up with the wrong overall result. To be sure of what we conclude in terms of best evidence, we have to use knowledge that is the very best. If we use poor quality knowledge, we may end up doing the wrong thing.

There are also some important issues around trial validity [8], especially for acupuncture.

## Outcomes (issue, consequence, result)

Another thing we have learned from doing systematic reviews is how many different outcomes people have used in clinical trials. Some are simple, like death. Some may be objective, and can be measured, like the level of a chemical in a person's blood. Others, like pain relief, may be subjective, where we have to ask, and trust, the patient. In many areas of medicine, though, it is much, much more complex.

Some of the issues around migraine outcomes will be dealt with later, but for now it is worth noting that those of us reading clinical trial reports or systematic reviews have a duty to ask ourselves, and satisfy ourselves, that the outcomes reported are meaningful to our patients, to us, or to the healthcare systems in which we work. There may be many outcomes - of benefit, or harm through adverse effects, or economic - in a single trial or review. Our job is to refuse to be blinded by science and ask if the outcome being reported is an important one.

We must avoid making what is measurable important, and find ways to make the important measurable.

Robert Macnamara

## Output (quantity turned out, or data after being processed by a computer)

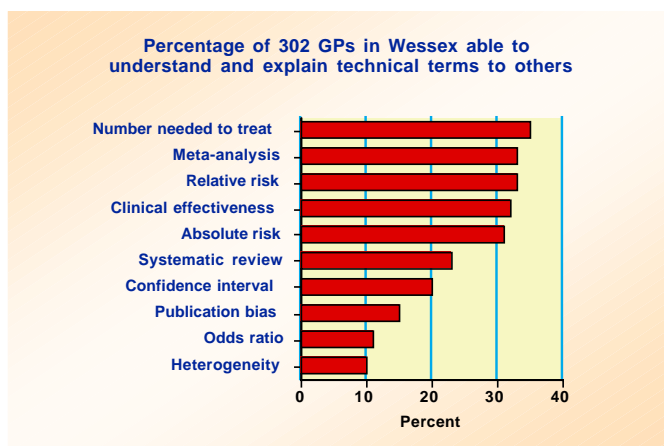
The way in which results of trials or systematic reviews are reported is of major consequence. Often we get some statistical output. Now statistics are important, so let's not forget that every study should have a proper statistical tick so that we know that the results are meaningful. But the statistical tick is not the result. It is a mathematical way of saying that things are different one from the other.

For a difference to be a difference it has to make a difference

Attributed to Gertrude Stein, ca 1920

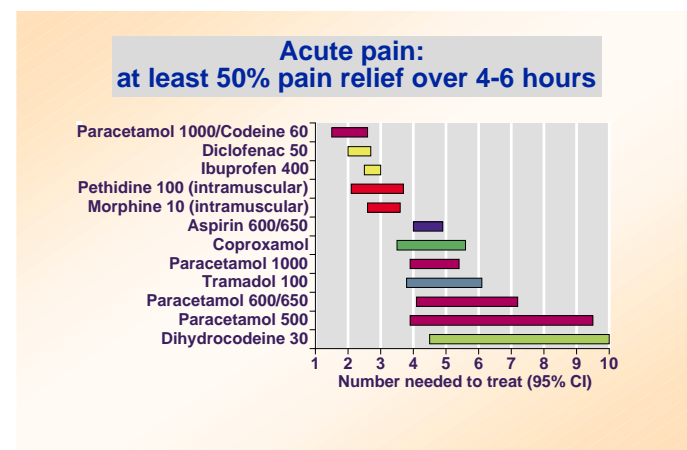
When we have the statistical tick, then we have to make up our minds whether the result makes a difference, and to do that we have to understand not just the outcome, but also how much of that outcome the intervention or treatment is delivering. It is on that that we can make our clinical judgement about whether to use it, and it is on that that we can explain the benefit of treatment to our patients.

One of the problems with many systematic reviews and meta-analyses is that they only give us the statistical output, or some derivative. This may be an odds ratio, or a relative risk, or a hazard ratio, or a weighted mean difference, or, God forbid, and effect size. Don't expect a detailed explanation of what an effect size is here. We work on the basis that our time on this earth is too short for things like that, and we need to get on with our lives. So we want simpler, more human, outputs. And we are not alone. A survey of GPs in Wessex in 1997 [9] showed that they were puzzled by outputs like odds ratios, and the one they were most likely to understand was the number needed to treat, or NNT [10]. New readers who want NNT explained in full can go to the Bandolier "what is?" download site. An NNT calculations sheet can be found on the Bandolier Internet site.



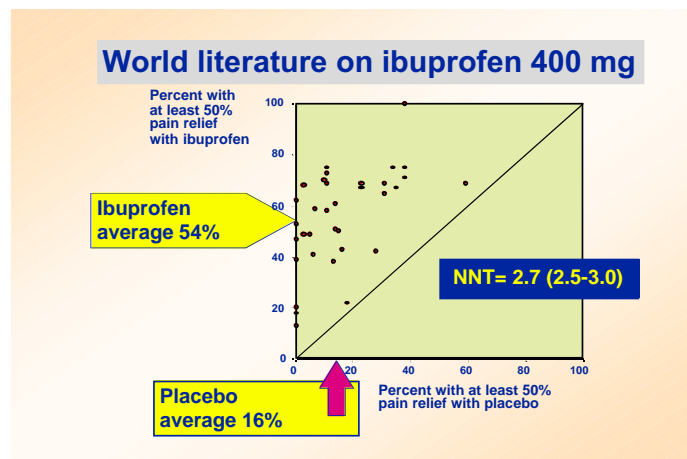
NNT is treatment specific. It describes the difference between active treatment and control in achieving a particular clinical outcome. Low NNTs indicate high treatment-specific efficacy. An NNT of 1 says that a favourable outcome occurs in every patient given the treatment but in no patient in a comparator group, the 'perfect' result in, say, a therapeutic trial of an antibiotic compared with placebo with a sensitive organism. NNTs of 2 to 5 are indicative of high efficacy, as with analgesics in acute pain.

We can compare NNTs when there is a common comparator (placebo, for instance), where there is the same outcome measured over the same period of time, when patients in the trials are the same, with the same condition and severity, and where the trials are all of high quality so that bias is minimised. An example is the acute pain league table.



## Size (bigness, magnitude)

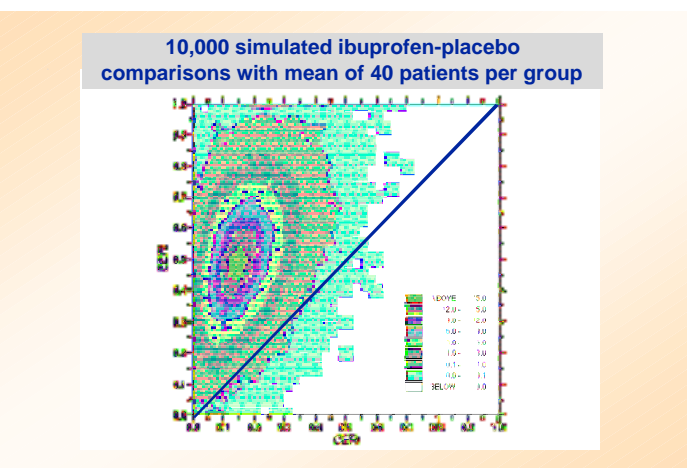
We also have to be sure that we have sufficient information on which to base a conclusion. The figure below looks at all the literature available on properly randomised, double-blind trials comparing ibuprofen with placebo in acute post-operative pain. They were impeccable trials, all using the same patients with the same initial degree of pain, and used the same outcomes over the same period of time.



Each point represents a trial, and we plot the percentage with at least 50% pain relief with placebo on the bottom, and the percentage with at least 50% placebo with ibuprofen 400 mg on the Y axis. All are above the line of equality, showing that ibuprofen is a better analgesic than placebo, which is encouraging. We can even see that the NNT of 2.7 means that ibuprofen is an effective analgesic.

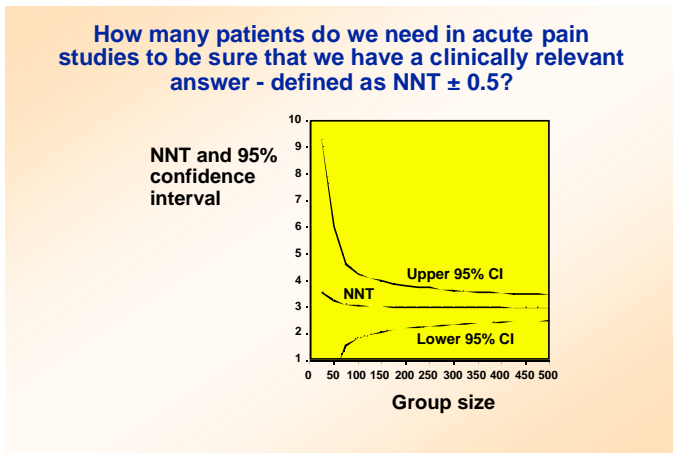
But why do we have such a scatter of points if all these trials are supposed to be the same. Is it because some were conducted in Welsh wimps and others in Scottish stoics, perhaps? Actually, no. These trials were all done to show that ibuprofen is better than placebo. They had about 40 patients per treatment group to do this. They were not done to show how much better ibuprofen is than placebo, a subtly different question, and one that needs far more patients to answer accurately.

Because we know how over 5,000 individual patients perform in these trials, we can mathematically model the effects of the random play of chance on these trials. In the representation below [11], anywhere in the grey area is



where a trial comparing ibuprofen 400 mg with placebo could fall just by chance. It is more likely to be in the redder areas, but the spread we see because of chance is at least as big as that we saw in practice with all the randomised comparisons of ibuprofen with placebo. So we don't need to seek abstruse reasons for differences between single trials until the effects of random chance have been eliminated. Only numbers will do that.

Two ideas stem from these considerations. The first is that we should beware the single trial reflex. However good a single trial is, unless it is very large these chance effects may still mislead us. Below we plot the effect of numbers on the NNT for ibuprofen from this modelling exercise. We know it is about 3, but the confidence interval is very wide until we have large numbers. If we want to be sure of the NNT within  $\pm 0.5$ , we need as many as 1000 patients in a trial. Trials that big just aren't done, so this is another reason why it is a good idea to use systematic reviews and meta-analysis that pull all quality data together.



How much information we need also depends on how big is an effect. We need less information (fewer patients) when the effect is big than when it is small.

### NNT accuracy - percentage of 10,000 simulated NNTs within $\pm 0.5$ of the true NNT

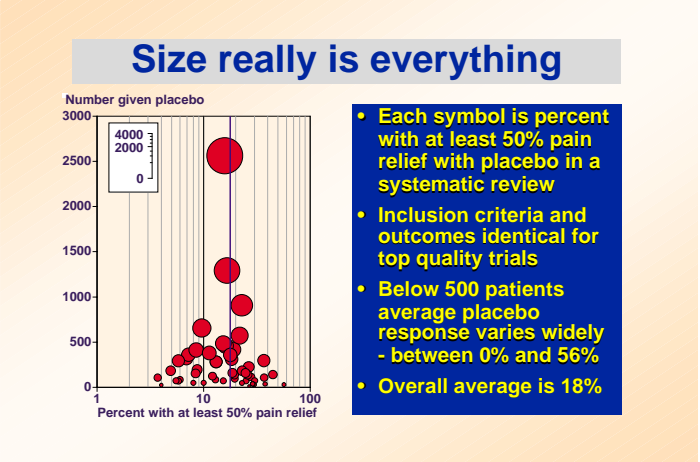
Group size	Number needed to treat		
	4.00	3.00	2.30
25	26	37	57
50	28	51	73
100	38	61	88
200	55	81	96
300	63	89	99
400	71	93	99
500	74	95	100

At least  
 50% within  $\pm 0.5$   
 80% within  $\pm 0.5$   
 95% within  $\pm 0.5$

Moore et al, Pain 1998 78: 209-16.

Just to finish off the business of size, and to emphasise again how important it is, the slide below is probably unique in that it draws together information from of 50 meta-analyses. Each blob represents the response rate found with placebo. We are plotting the rate or people achieving half pain relief with placebo against the number of patients given

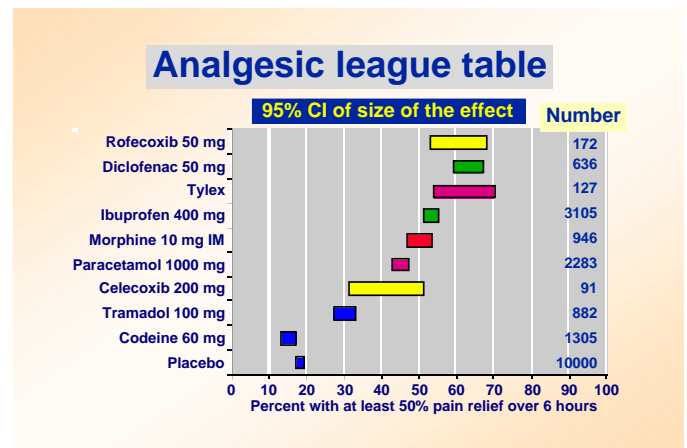
placebo. In total there are 12,000 such patients, and the blue vertical line represents the overall response rate of 18%. Only when the number of patients with placebo in the meta-analysis is large (of the order of 1000), is the overall rate accurately measured. This emphasises that size is everything.



## Utility (usefulness: the power to satisfy the wants of people in general)

The final consideration is that of being useful to people. If statistics represents the first tick, and issues around outcomes and validity a second tick, perhaps NNTs make up the third tick. But even if a third of Wessex GPs can stand up and explain to others how to calculate and use an NNT, that means that two-thirds cannot. And even the one-third will be busy, or harassed, and will want some simpler way of understanding for themselves and others.

So if there is a fourth tick, it has to be presenting results in a way that is useful - immediately useful without having to engage too many neurones. That could be as simple as telling us what proportion of people get the outcome with the treatment. An example in acute pain follows:



What this does is show us what proportion achieves half pain relief in properly done, immaculate randomised double-blind studies of the same types of patients with the same outcome over the same period of time. We've put the numbers of patients at the right edge. Some, like ibuprofen, have large numbers. Some, like Tylenol, have small numbers (but

there may be other data to support this, as a review points out, and they have been around for ages). Others, like the new coxibs, are new, and the numbers are small. In any event, here is a representation of immediate relevance.

One source of that relevance is that we concentrate not just on those achieving the outcome, but those who do not - that part of the graph to the right of the bars. If people do not achieve the outcome, then something more has to be done. That may be complex or expensive (as with reflux treatments), or may just be giving another dose of analgesic. In any event, people not achieving an outcome are really important because they represent those for whom we have to do more.

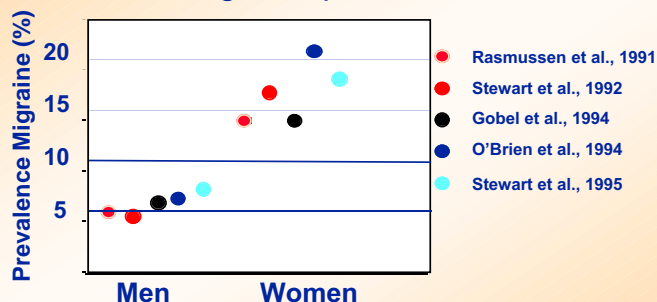
## Migraine background

Migraine is common. It affects about 1 adult woman in 5 and 1 adult man in 20. On average someone with migraine will have three attacks a month, and lose much time out of their lives. Yet only a small proportion seek help from the doctors, choosing mostly either to treat with medicines bought from pharmacies, or not to treat at all.

### Migraine prevalence

- 20% in women aged 20-64
- 7% in men
- Average 36 attacks in one year
- Average of 5 days in bed a year
- Average bed rest is 5-6 hours
- Only 1 in 5 sufferers seek help
- Incidence increasing

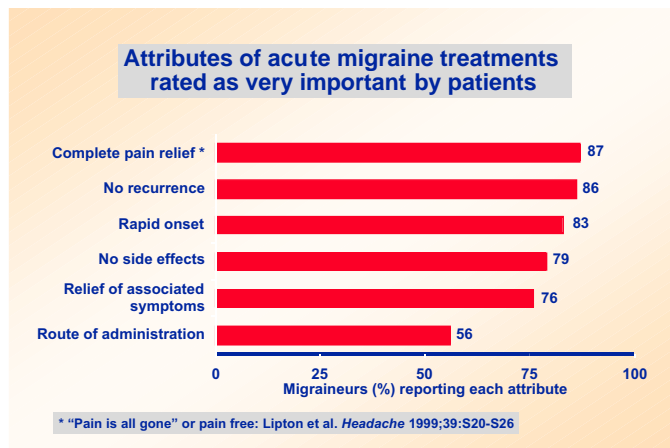
### Gender specific estimates of migraine prevalence



Migraine is also an expensive business. Health economists might argue over how much lost productivity there is in the economy, but the costs involved in whatever study one looks at, just because of time lost at home or at work, are large. It has been estimated that the cost of migraine in the USA is \$14 billion a year. We have summarised a number of studies around the health economics of migraine.

## Outcomes in migraine trials

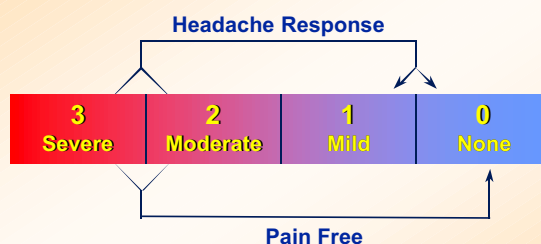
What is it that people who suffer from migraine want from their treatments? A full summary about what people want from migraine treatment is on the site. It is summarised in the slide below: they want their headache pain relieved, totally, quickly with no adverse effects, and they don't want the pain to come back.



What happens in migraine trials that can allow us to answer these questions? Firstly, patients have to score their pain (most migraine trials measure pain as the primary outcome), scoring it using the words no pain, mild pain, moderate pain, or severe pain. In order to enter the trial they have to have pain of moderate or severe intensity. This is all pretty standard stuff. Pain measurements like this have been used successfully for decades, and we know in other settings like acute pain, that if patients have only mild pain we could have insensitive studies (how can you measure an analgesic effect when there's no pain?).

The pain is then scored by the patient at hourly or half hourly intervals until, say, six hours, or even as long as 24 hours. The outcomes chosen by trialists (rather than patients) has been the headache response at two hours. This is a headache that starts as moderate or severe but where the pain has declined to mild or no pain by two hours. The headache response could be measured at any time point. Preferred now is pain free at two hours, though pain free at any time could be measured.

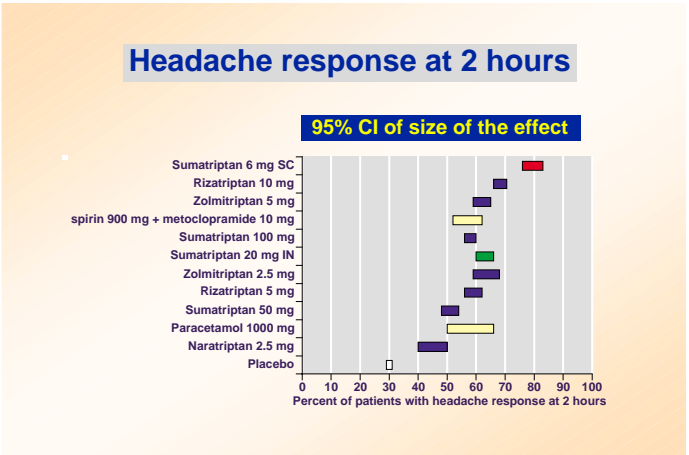
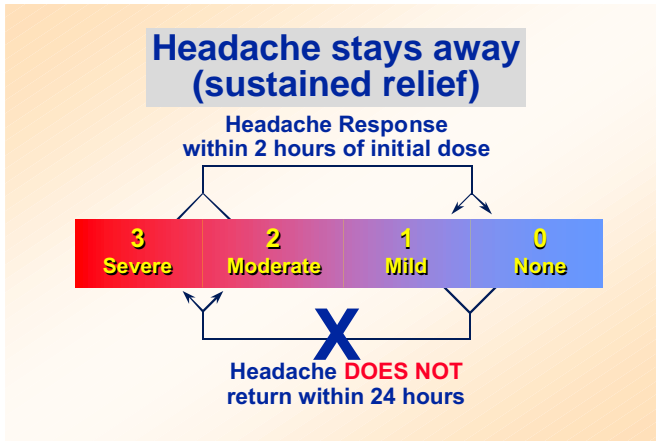
### Headache response and pain free



- Pain free at 2 hours is now the preferred IHS primary endpoint for clinical trials

Perhaps what we should be measuring is something different - namely the pain gone at two hours and not returned within the following day. So far this has not been done, but results like this should be appearing soon.

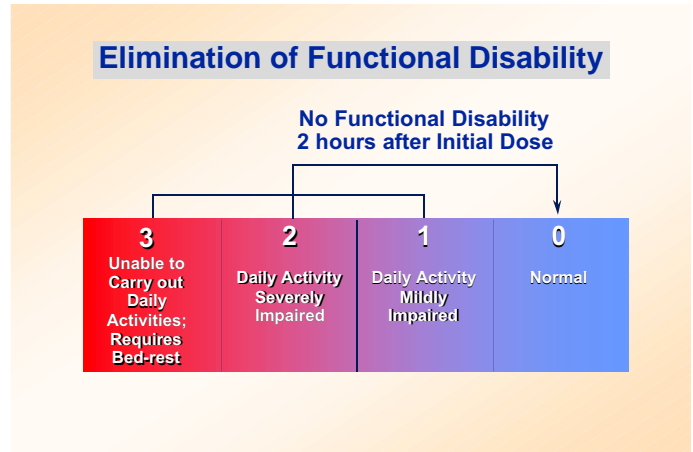
sults in league tables, both as numbers needed to treat (NNTs), or simply as percentages of patients having the outcome. Don't take these figures at face value, though, and read the comments in the league table page about the dangers of over-interpreting league tables.



The point is that we have a number of possible outcomes from recent trials of high quality and validity (discussed here). Older studies (as with ergotamine) often fail to match up to modern standards, and have many different methods of reporting outcomes.

One thing that is worth remembering is that there are many other possible outcomes of benefit in migraine trials, like reduction of nausea and vomiting, or photophobia, or phonophobia, or restrictions of daily living. We don't have reviews of these yet, but they are worth doing and worth reporting when available. An example below is for a scale that measures functional disability with migraine.

- Migraine trial outcomes**
- 1 Headache response at 1 hour
  - 2 Headache response at 2 hours
  - 3 Pain free at 2 hours
  - 4 Headache response at 2 hours plus no headache up to 24 hours
  - 5 Pain free at 2 hours plus no headache up to 24 hours

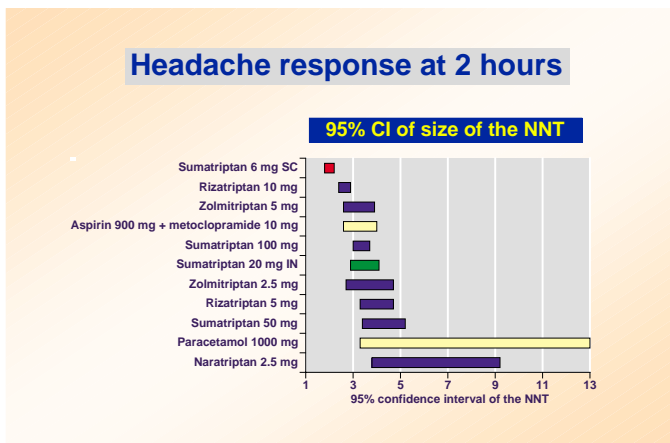


## Results of migraine trials

When we begin to look at migraine trials, using those of high quality and high validity, we find that there are many. Most treatments have been summarised as reviews, or reviewed by us, in these web pages. We can combine the re-

## Adverse events

These are proving really quite difficult to get to grips with. Part of the problem is the way that adverse events are measured and reported generally (Bandolier 85). In migraine trials there has been a particular difficulty, in that benefits (pain relief) have been measured over 24 hours, while adverse events have been collected over 10 days. The result is that at present it is difficult to find much of interest or value to say.



## Conclusions

Systematic review can help us in a number of different ways, especially when thinking about issues like placebo effects, the quality and utility of outcomes, and the validity of trials. Their real place is not only an archaeological rummage

through trials of yore, but rather a learning process to ensure that trials and research in the future reach a much higher standard, and are more immediately useful to professionals and to patients.

## References:

---

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *British Medical Journal* 1996;312:71-2.
2. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995, 273: 408-12.
3. M Tramèr M, DJM Reynolds, RA Moore, HJ McQuay. Effect of covert duplicate publication on meta-analysis; a case study. *British Medical Journal* 1997 315: 635-40.
4. RA Moore, D Carroll, PJ Wiffen, M Tramèr, HJ McQuay. Quantitative systematic review of topically-applied non-steroidal anti-inflammatory drugs. *British Medical Journal* 1998 316: 333-8.
5. 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.
6. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998, 352 :609-613.
7. 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.
8. 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.
9. A McColl, H Smith, P White, J Field. General practitioners' perceptions of the route to evidence based medicine: a questionnaire survey. *BMJ* 1998 316:361-5.
10. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. *Ann Intern Med* 1997, 126: 712-720.
11. Moore RA, Gavaghan D, Tramèr MR et al. 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.

©Bandolier: 08-Jan-2002