Bandolier Professional

Independent evidence-based health care

Outputs and utility

Introduction

This is the first of a series of articles aimed at helping to understand clinical trials and systematic reviews and meta-analyses. Understanding involves not just how trials are conducted, but what constitutes a good or bad trial, and how to comprehend and use results from a trial. Any time we read a clinical trial report or a systematic review need to think not about one or two factors, but many.

Deciding where to start the process is not easy. As a good a place as any is the output from a trial, or how the trial results are described, and how useful are different ways of describing trials. We have to comprehend the results before we can use them, and how those results are reported affects our comprehension.

It is worth remembering two things

There is probably no single "best" way of reporting results so that everyone will comprehend them. We are all very different in our tastes one from the other. It would not be surprising that we are likely to vary in the facility with which our minds grasp and manipulate different concepts. Some of us will be most happy with complex statistical concepts. Others, and ***Bandolier*** is one of those, are happiest with the most simple representations. In this essay we will try to avoid being judgemental about different outputs, but will point out where there can be difficulties.

There are two important efforts to improve the reporting of randomised trials and systematic reviews. These have produced the CONSORT guidelines [1] for randomised trials and the QUORUM statement for systematic reviews [2]. Not all journals subscribe to these, but if studies have statements about these reporting guidelines they should be better than those that do not.

The meaning of words

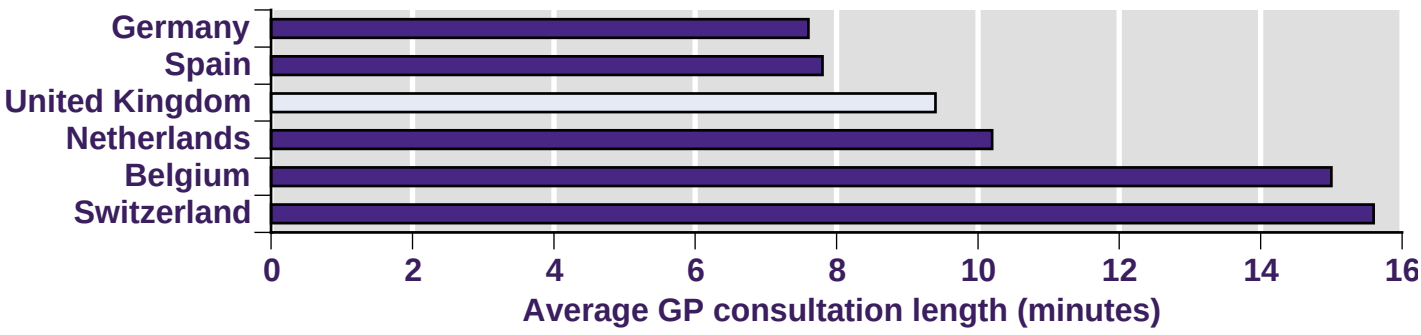
Output - "quantity produced or turned out; data after processing by a computer". This dictionary definition is the first clue to what this note is all about. If our best evidence comes from randomised trials and systematic reviews, then one of the most important things is how that trial or review gives us the result. From a clinical point of view we want to know how much benefit (or harm) a treatment will produce, and results of data processing might not be much help.

Utility - "usefulness: the power to satisfy the wants of people in general". This is the other part of the clue. If systematic reviews are to be useful, and therefore used, they have to present results in ways that are immediately accessible to ordinary professionals. When the average UK GP has just nine minutes to see a patient, and the duration of a consultation varies from 7.6 minutes in Germany to 15.6 minutes in Switzerland (Figure 1) [3], rapid understanding becomes important. Trying to work out what a hazard ratio or an effect size means to treating the patient in front of us will not make for an easy morning surgery.

Defining outputs

Most of the outputs that we use for reporting trials and reviews have their origins in epidemiology, the world where we look for small effects in large populations – things like aspirin after a heart attack, or reducing cholesterol. Most of

Figure 1: Average GP consultation times in European countries



the activity of medicine is conversely about large effects in small populations, like hip replacements for osteoarthritic joints, or anaesthesia, or pain relief, or antibiotics for infection. So our example is one most of us will be familiar with.

Table 1 is a hypothetical trial of ibuprofen in acute pain. Not worrying too much at this stage about any other features or even the result itself, we will use this trial to present some of the more common definitions for presentation of results where information is available in dichotomous form. Dichotomous means the patient had the outcome or did not, and we have the numbers for each. In this trial, for instance, 22 of 40 patients given ibuprofen had adequate pain relief compared with only 7 of 40 given placebo. The term experimental event rate (EER) is used to describe the rate that good events occur with ibuprofen (22/40, or 55%) and control event rate (CER) to describe the rate that good events occur with placebo (7/40 or 18%).

Odds ratios

This Table shows first how to compute odds. Odds refers to the ratio of the number of people ***having*** the good event to the number ***not having*** the good event, so the experimental event odds are 22/18 or 1.2. The odds ratio is the ratio of the odds with experimental treatment and that of control, or here 1.2/0.21 = 5.7. There are lots of different ways of computing odds ratios that give slightly different answers in different circumstances. Values greater than 1 show that experimental is better than control, and if a 95% confidence interval is calculated, statistical significance is assumed if the interval does not include 1.

Some would change this around and compute the odds ratios from the point of view of the patients **not** having adequate pain relief. The experimental event odds would be 18/22 or 0.82, and the control event odds would be 33/7 or 4.7. The odds ratio then would be 0.82/4.7 = 0.17.

For ibuprofen versus placebo the odds ratio is 5.7 or 0.17. Pick the bones out of that. How would you use that, other than knowing that an odds ratio that was far from 1 meant that ibuprofen was better than placebo.

Relative risk or benefit

Relative risk is a bit easier on the brain. It is simply the ratio of EER to CER, here 0.55/0.18 (or 55/18 for percentages), and is 3.1. Again values greater than 1 show that experimental is better than control, and if a 95% confidence interval is calculated, statistical significance is assumed if the interval does not include 1. Odds ratios and relative risk often give the same numerical value when they are low,

Table 1: Hypothetical acute pain trial

Results of hypthetical randomised trial

Treatment	Total number of patients treated	Number who achieved at least 50% pain relief	Number who did not achieve at least 50% pain relief
Ibuprofen 400 mg	40	22	18
Placebo	40	7	33

Calculations made from these results

Experimental event rate (EER, event rate with ibuprofen)	22/40 = 0.55 or 55%
Control event rate (CER, event rate with placebo)	7/40 = 0.18 or 18%
Experimental event odds	22/18 = 1.2
Control event odds	7/33 = 0.21
Odds ratio	1.2/0.21 = 5.7
Relative risk (EER/CER)	0.55/0.18 =3.1
Relative risk increase (100(EER-CER)/CER) as a percentage	100((0.55-0.18)/0.18) = 206%
Absolute risk increase or reduction (EER-CER)	0.55 - 0.18 = 0.37 (or 37%)
NNT (1/(EER-CER))	1/(0.55 - 0.18) = 2.7

but not when high. There is disagreement between eminent statisticians about which of these is "best". We use relative risk, but wouldn't pick a fight with someone who preferred odds ratios.

Again, knowing that the relative risk is 3.1 is not intuitively useful. Both relative risk and odds ratio are important ways of ensuring that there is statistical significance in our result. Unless there is statistical significance, we should not be using a treatment except in exceptional circumstances. So whatever else we do in the way of data manipulation, one or other of these tests has primacy for giving us the right to move on.

Relative risk reduction or increase

The relative risk reduction is the difference between the EER and CER (EER-CER) divided by the CER, and usually expressed as a percentage. In Table 1 the relative risk increase is 206%. If the number of events is smaller with treatment, then the relative risk reduction is calculated by subtracting the CER from EER in the equation.

Absolute risk increase or reduction

If we subtract the CER from the EER (EER-CER) then we have the absolute risk increase (ARI), the effect due solely to ibuprofen, and nothing else. The language here doesn't quite work because it was originally taken from the world of epidemiology where reducing risk (cholesterol lowering etc) is all. The absolute risk reduction (ARR) is CER-EER, when events occur more often with control than they do with treatment.

Number needed to treat (NNT)

For every 100 patients with acute pain treated with ibuprofen, 37 (55 - 18) will have adequate pain relief because of the ibuprofen we have given them. Clearly then, we have to treat 100/37, or 2.7 patients with ibuprofen for one to benefit because of the ibuprofen they have been given. That's what NNT is (Table 1). This has immediate clinical relevance because we immediately know what clinical and other effort is being made to produce one result with a particular intervention.

The best NNT would be 1, where everyone got better with treatment and nobody got better with control, and NNTs close to 1 can be found with antibiotic treatments for susceptible organisms, for instance. Higher NNTs represent less good treatment, and the NNT is a useful tool for comparing two similar treatments. When doing so the NNT must always specify the comparator (e.g., placebo, no treatment, or some other treatment), the therapeutic outcome, and the duration of treatment necessary to achieve that outcome. If these are different, you probably should not be comparing NNTs. It is also worth mentioning that prophylactic interventions that produce small effects in large numbers of patients will have high NNTs, perhaps 20-100. Just because an NNT is large does not mean it will not be a useful treatment.

We can use the same methods for adverse events, when numbers needed to treat become numbers needed to harm (NNH). Here small numbers are bad (more frequent harm) and larger numbers good. When making comparisons between treatments, the same provisos apply as for NNT, especially that for definition.

For both NNT and NNH we should recognise that we are working with an unusual scale which runs from 1 (everyone has outcome with treatment and none with control) to –1 (no-one has outcome with treatment and everyone has it with control), with infinity as the mid point where we divide by zero when EER equals CER. Once NNTs are NNHs are much above 10 the upper confidence interval gets closer to infinity and the upper and lower intervals look unbalanced.

Other outputs

There are masses of other outputs that people use for trials and epidemiological studies. These include effect size, relative risk reductions and so on. We don't find these useful, but there will always be circumstances in which they are the appropriate outputs.

Understanding different outputs

In order to see how the different outputs look, we can return again to a hypothetical trial of analgesic and placebo, though for convenience we use 1000 patients per group because we want to explore a range of possible results. In Table 2 we show six possible results in which the absolute proportion of patients benefiting with ibuprofen and placebo varies wildly, but the relative proportions stay the same.

Table 2 shows that the relative risk and relative risk increase can stay almost exactly the same in the face of huge changes in NNT, absolute risk increase, and the absolute percentages of patients benefiting. It explains why statistical outputs like relative risk may be great for measuring statistical significance, but lack utility in everyday practice.

These are clearly different trials or conditions, where the event rate with placebo varies between 1% and 36% in 1000 patients. Indeed, if these trials were part of a systematic review, this would be an obvious case where we would expect clinical heterogeneity in the condition (diagnosis), or its severity, or in patients recruited for the trials, or outcomes measured or time over which the outcomes were measured. What is interesting is that using *statistical* tests for heterogeneity, this is a "perfect" result, giving a p value of 1.0. It underlines the fact that most statistical tests for heterogeneity are usually wrong, and even the one that is right is useless at detecting whether a group of trials is homogeneous or not [4].

Table 2: Hypothetical trials using different outputs

	Hypothetical trials with 1000 patients each given analgesic or placebo							
	At least 50% pain relief							
Trial	analgesic	placebo	EER (%)	CER (%)	Relative risk	Relative risk increase (%)	Absolute risk increase (%)	NNT
1	800	360	80	36	2.2	122	44	2.3
2	400	180	40	18	2.2	122	22	4.6
3	200	90	20	9	2.2	122	11	9.1
4	100	45	10	5	2.2	100	5	18.2
5	50	23	5	2	2.2	150	3	37.0
6	20	9	2	1	2.2	100	1	90.9

Some real examples

Rizatriptan for acute migraine

A real example is an individual-patient meta-analysis of all the randomised single-dose trials of rizatriptan compared with placebo in acute migraine [5]. This is interesting because it looks at four different outcomes for migraine trials, and because it allows us to compute all the different outputs (Table 3). For clarity the confidence intervals have been omitted, but rizatriptan was significantly better than placebo for all four outcomes.

Both odds ratios and relative benefit tell us that rizatriptan 10 mg is better than placebo, but the numerical values are not helpful. The relative risk increase looks very impressive for pain free responses compared with headache response over two or 24 hours, despite fewer patients actually achieving this outcome. Absolute risk increase and NNT are a more helpful, while conveying the same information in slightly different ways. Best of all for some of us is the absolute percentage of patients who will get each outcome if given 10 mg rizatriptan when they have a migraine.

Prophylaxis for NSAID-induced gastrointestinal problems

A Cochrane review [6] examines the efficacy of misoprostol, histamine antagonists (H2A), and proton pump inhibitors (PPI) to reduce the propensity of NSAIDs to cause gastric and duodenal ulcers. Its main focus is on ulcers seen endoscopically (because that's were most of the evidence is), and on adverse events, particularly diarrhoea caused by misoprostol. This is a good review, but like many Cochrane reviews, one that concentrates on statistical outputs.

The review actually uses odds ratios for much of the reporting of results, but for completeness Table 4 shows the results for prevention of endoscopically-detected NSAID-induced gastroduodenal ulcers for each treatment. All are significantly better than placebo, and confidence intervals have been omitted for clarity.

Odds ratios and relative risks were lower (better), at 0.2-0.3 for PPI, double doses of histamine antagonists, and 800µg misoprostol. Relative risk reduction was higher (better) for these same three treatments. Misoprostol and standard dose histamine antagonists were less good with odds ratios and relative risks about 0.4-0.6. But looking at the NNT shows that despite these statistical results, only PPI and double-dose histamine antagonists had the lower (better) value.

But looking at the percentage figures, we find that only 2-6% of patients given misoprostol had endoscopic ulcers, about 10% for PPI and standard dose histamine antagonist, while the worst result was for double-dose histamine antagonist with 15% of patients with ulcers.

Are you confused? Join the club. To help ourselves out the first thing to notice is what happens when we do nothing, in this case give placebo. With no active prophylactic treatment there is a huge variation in what happened in the review – 12% of patients developed gastroduodenal ulcers in trials with 800 µg misoprostol, while 36% developed them with double-dose histamine antagonists. Part of this is to do with size and it's effects on variability. We'll deal with this in a future essay, but for now Figure 2 shows what happens with placebo in individual trials.

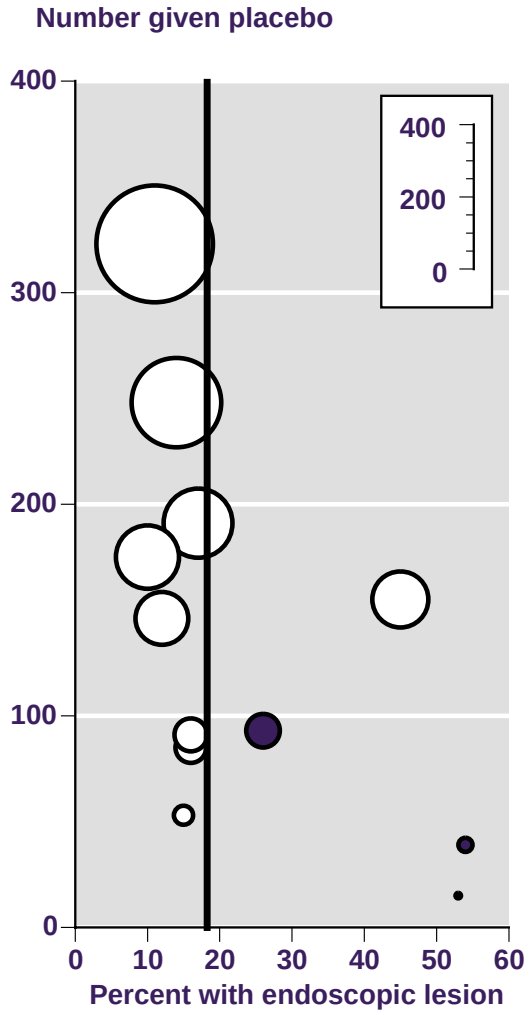
Table 3: Outputs of meta-analysis of rizatriptan 10 mg versus placebo for acute migraine

	Number of patients with the outcome		Output					
Outcome	Treatment (2056 patients)	Placebo (1249 patients)	Odds ratio	Relative benefit	Relative risk increase (%)	Absolute risk increase (%)	NNT	Percent of patients with the outcome
Headache response at 2 hours	1460	475	3.9	1.9	87	33	3.0	71
Pain free at 2 hours	843	125	4.5	4.1	310	31	3.2	41
Headache response over 24 hours	761	225	2.5	2.1	106	19	5.3	37
Pain free over 24 hours	514	87	3.4	3.6	257	18	5.5	25
Bigger or smaller numbers better			Bigger	Bigger	Bigger	Bigger	Smaller	Bigger

Table 4: Outputs of meta-analysis of prophylaxis for NSAID-induiced gastroduodenal ulcer

	Number of patients with the outcome/total		Output						
Treatment	Treatment	Placebo	Odds ratio	Relative risk	Relative risk reduction (%)	Absolute risk reduction (%)	NNT	Percent with treatment	Percent with placebo
Standard dose H2A	48/494	75/487	0.6	0.6	33	5	17	10	15
Double dose H2A	22/151	53/147	0.3	0.4	58	21	4.7	15	36
400 ug misoprostol	21/357	49/366	0.4	0.4	54	7	13	6	13
800 ug misoprostol	8/380	45/376	0.2	0.2	83	10	10	2	12
PPI	49/443	98/331	0.2	0.3	63	19	5.4	11	30

Figure 2: Endoscopic gastroduodenal lesions with placebo in individual trials



The filled circles were the trials with double-dose histamine antagonist, and the vertical line the average result (18%) from all trials. Clearly the high-dose histamine antagonist trials had two small with very high rates of ulcer development without prophylactic treatment, and with no obvious reason for it. The high rate at which things happen with placebo with small numbers of patients is why the odds ratios, relative risk and relative risk reduction were so impressive, being relative to what happened with placebo. Even absolute risk reduction and NNT are dependent on what happened with placebo, because high rates with placebo give more scope for an effect of treatment. Actually, there was not much difference between double-dose and standard dose H2A, or 400 µg and 800 µg misoprostol, or PPI on the basis of these figures.

Of course, the outcome here is gastroduodenal ulcers seen on endoscopy, which is not the same as clinically relevant ulcers. Perhaps we shouldn't get too excited about it these results, except that they are our main guidance on what to prescribe. Let's look at the reviewers' own conclusions:

"Misoprostol, PPIs, and double dose H2As are effective at preventing chronic NSAID related endoscopic gastric and duodenal ulcers. Lower doses of misoprostol are less effective and are still associated with diarrhea. Only Misoprostol 800µg/day has been directly shown to reduce the risk of ulcer complications."

So standard-dose H2A has been dismissed, despite providing lower rates of gastroduodenal ulcers than double-dose H2A, and equivalent to that obtained with PPI. Lower dose misoprostol is dismissed despite providing lower absolute rates of gastroduodenal ulcers than double-dose H2A and PPI.

Diarrhoea with misoprostol

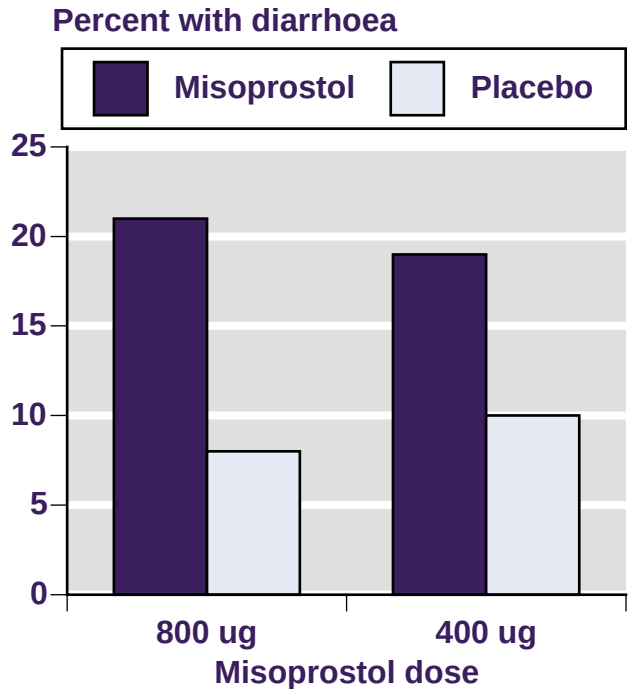
Now let's look at what the same review said about diarrhoea with misoprostol:

"Both misoprostol doses were associated with a statistically significant risk of diarrhea. However, the risk of diarrhea with 800 µg/day (RR=3.05; 95% CI:2.42-3.83) was significantly higher than that seen with 400 µg/day (RR=1.92; 95% CI:1.64-2.26) (p=0.0012)"

This is impressive stuff, with high statistical significance. The actual results are shown in Figure 3, where 21% of patients on misoprostol 800 µg had diarrhoea, compared with 19% on misoprostol 400 µg. The NNHs were 7 (95% confidence interval 6 to 9) for 800 µg and 12 (9 to 16) for 400 µg. We can be reasonably sure of these results, because in each case they were obtained from over 3,000 patients. The result is true, there was a ***statistically*** significant increase in diarrhoea with the higher dose, and the NNHs were different. But was the increase in incidence from 19% to 21% ***clinically*** significant?

These examples serve to demonstrate several important points. First, it ain't easy. But it becomes much easier if we look at all the possible outputs before making any decisions. Relative statistics can misled by over-emphasising statistical over clinical relevance. Absolute outputs tell us what happens absolutely, and the NNT tells us about the therapeutic effort needed to produce one clinically relevant result.

Figure 3: Incidence of diarrhoea with misoprostol



Different outputs do matter

There has been research on the interpretation of numerical information and how that depends on the presentation of the information. Technically this is known as "framing", and the effects of framing have been examined in a systematic review [7] of 12 studies published up to about 1998. Relative risk reduction or increase were outputs that viewed most positively by doctors, but formal meta-analysis was not possible. Some selected studies are examined below, but the review concluded that the effects of framing on clinical practice are unknown.

Purchasers and presentation

The importance of the way in which information is presented is emphasised by Fahey et al [8], who gave 182 health authority members results from a randomised trial on breast cancer screening and results from a systematic review on cardiac rehabilitation. The results were presented to them (Table 5) in four different ways:

- relative risk reduction
- absolute risk reduction
- proportion of event-free patients
- numbers of patients treated to prevent one death

From the 140 questionnaires returned the willingness to fund either programme was influenced significantly by the way in which results were presented. Relative risk reduction produced significantly higher inclination to purchase, followed by NNT. Intriguingly only three respondents, "all non-executive members claiming no training in epidemiology" said that they realised that all four sets of data summarised the same results.

Doctors and presentation

It is not only members of health authorities who are susceptible to altered perceptions of effectiveness according to the way in which the results of studies are presented to them. Two studies have looked at the effects of presentation on decisions by doctors in teaching hospitals in Canada [9] and on GPs in Italy [10]. Both used data from the Helsinki heart study.

Hospital doctors

In the first of these studies, David Naylor and colleagues [9] compared clinicians' ratings of therapeutic effectiveness by looking at different end-points presented as percent reductions in relative risk, absolute risk, and numbers-needed-to-treat. The study was conducted using random allocation of questionnaires using relative data or absolute data, each with NNT, among doctors of various grades at Toronto teaching hospitals. They used an 11-point scale anchored at "no effect" and running from -5 "harmful" to +5 "very effective".

Relative presentation consistently showed a tendency to higher scores - that is the intervention was interpreted as being more effective (Figure 4). Where data from a single end point, for any myocardial infarction, was examined, both relative and absolute comparison was scored consist-

Figure 4: Scoring effectiveness on "any myocardial infarction" by method of presentation

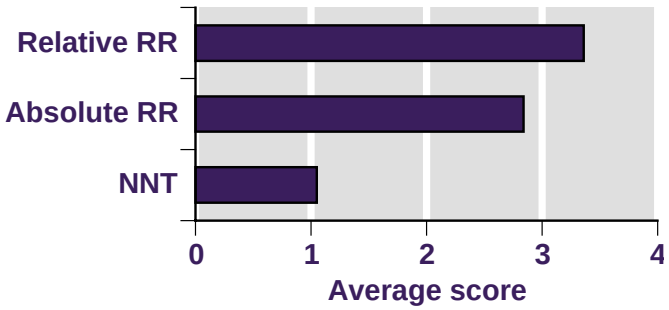
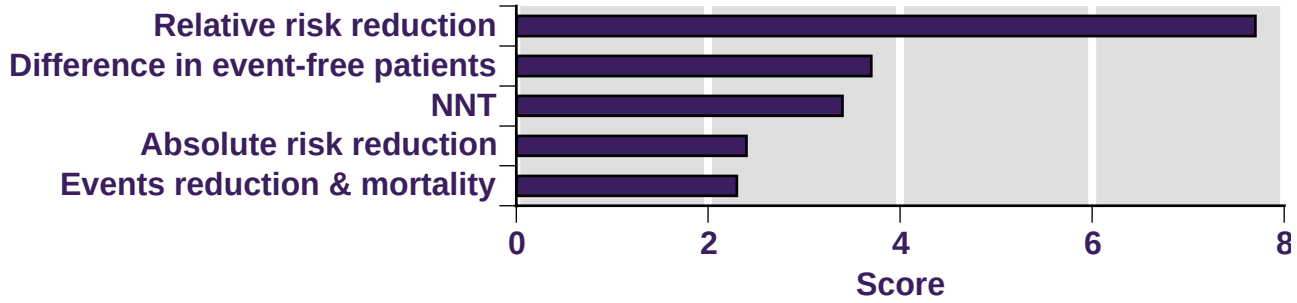


Table 5: Presentation of results on mammography and cardiac rehabilitation

Information presentation	Manmmography	Cardiac rehabilitation
Relative risk reduction	34%	20%
Absolute risk reduction	0.06%	3%
Percent of event-free patients	99.82% vs 99.80%	84% vs 87%
Number needed to treat	1592	31

Figure 5: GPs willingness to prescribe scores, by method of presenting data



ently higher than NNT presentation of the same data. NNT reporting of the same information produced a reduction of about two points in the effectiveness scale, reducing the judgement from quite effective to one of only slight effect.

General practitioners

This second study [10] presented information to 148 GPs using information from the trial as if it referred to five different drugs. The presentations were:-

- relative risk reduction
- absolute risk reduction
- difference in event-free patients
- NNT to prevent one event
- events reduction and mortality

For each statement about effects, the GPs were asked to mark a 10 cm line labelled "I would definitely not prescribe this drug" on the left and "I would definitely prescribe this drug" on the right. The statements were presented in random sequences. The results are shown in Figure 5. Presentation as relative risk reduction produced a very large tendency towards prescribing with a mean score of 7.7 out of 10. All other presentations produced scores of between about 2.5 and 3.5.

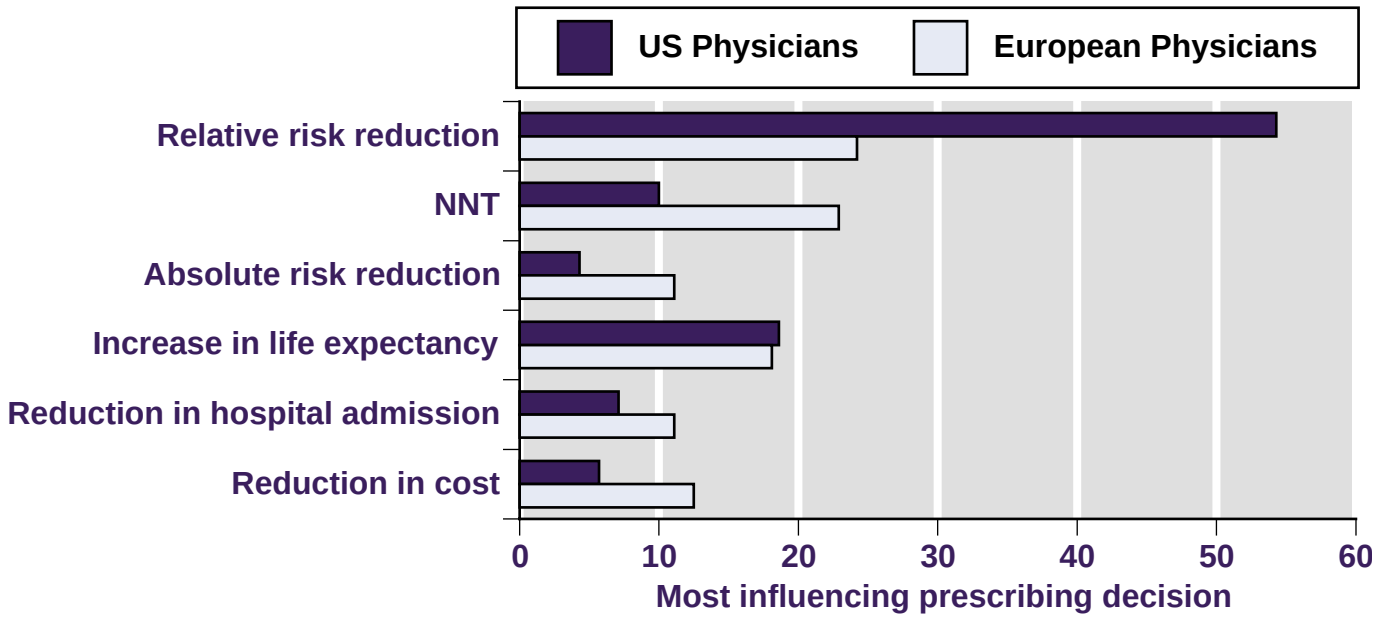
US and European physicians and pharmacists

A more recent study looked at US and European physicians and US pharmacists and examined data presentation related to their willingness to prescribe a drug [11]. The same information was presented in three different ways, and overwhelmingly respondents chose data presented as relative risk reduction as that most likely to make them prescribe (Figure 6).

But three "distracter" statements about life expectancy, cost and hospital admission rate also attracted significant attention as first-choice determinants, and about 40% of respondents preferred those over clinical trial results.

Interestingly, determinants were slightly different for European physicians from US physicians. European physicians were much more influenced by NNT and absolute risk reduction (and cost and hospital admission), and they were much less influenced by relative risk reduction. Despite this, relative risk reduction remains a potent influence on decisions, as a UK study confirmed [12].

Figure 6: Choice of information for determining prescribing decision



What GPs understand

Most consumer-oriented companies spend a lot of time finding out what their customers want, and giving it to them. There's little attention paid to the needs and wants of ordinary healthcare professionals trying to provide an excellent service despite many difficulties. In one of the few such studies regarding outputs from research GPs in Wessex were asked about their knowledge of evidence-based terms in 1996 [13].

The questionnaire asked some penetrating questions about GPs' knowledge of technical terms used in evidence-based medicine (things like odds ratios, heterogeneity and the like). They used a very high hurdle - whether respondents understood the term and could explain it to others.

Of all the terms, the one which came out top of this stiff test was the number needed to treat (NNT), with 35% of GPs being able to understand it and explain it to others (Figure 7). Relative risk was also reasonably well understood, but odds ratios were not.

It is instructive to look at this from the other perspective of how many people do ***not*** understand these terms. Ninety percent of GPs have no idea how to describe or use an odds ratio. But 65-70% don't know how to describe or use an NNT or relative risk, at least according to the high standard set by this study. And that's the best result from a savvy group of GPs.

A similar study in Australia came to very similar conclusions, while also exploring the difficulties faced by GPs in putting evidence into practice. [14]. What is also interesting is that GPs were able to spot that heterogeneity tests weren't worth thinking about, that odds ratios lack intuitive utility, and that publication bias is something best left to academic pointy heads to argue about because it can never be proved or disproved. All we can say is that tests to detect publication bias don't work [15].

How randomised trials report results

Five top English-language journals were examined in 1989, 1992, 1995 and 1998 for how randomised trials reported their results [16]. Of 359 articles in total, eight (2.2%) reported NNTs and 18 (5.0%) reported absolute risk reduction.

Comment

There's no one answer to what output works best for everyone, or anyone, or for particular circumstances. Graphical representations may be better than numbers in aiding comprehension of clinical trial results [17]. The main thing is to be sure that you know and understand whatever output you choose, and especially not to be swayed by things like relative risk, or odds ratios, or relative risk reduction, or whatever, when some of these can be highly statistically significant but clinically irrelevant.

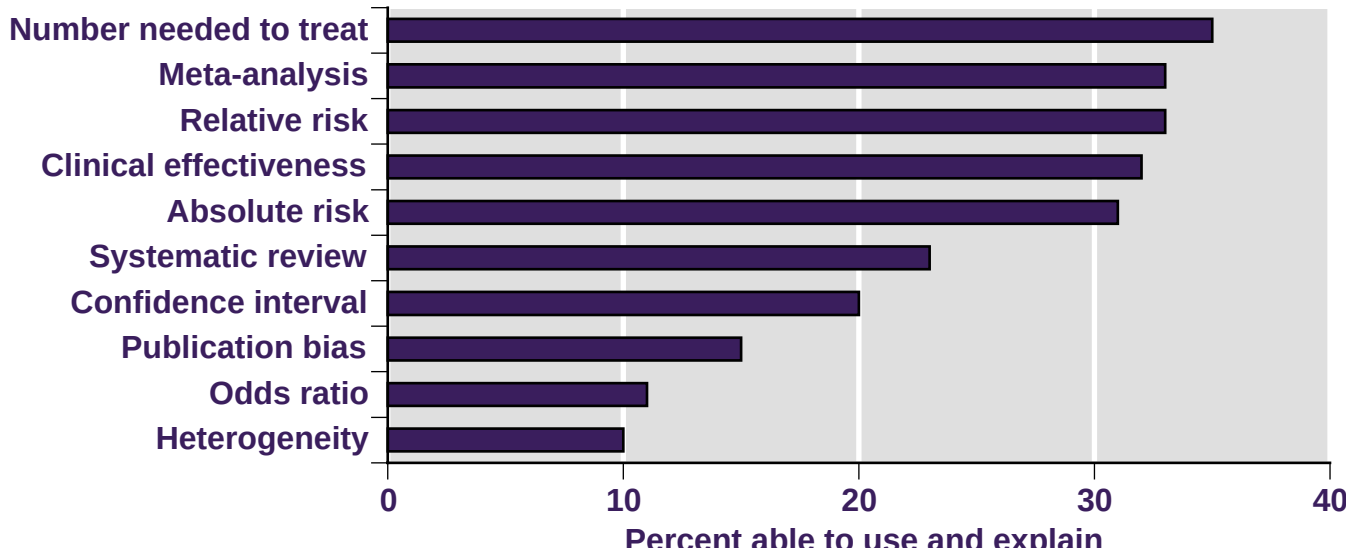
Few clinical trials and few systematic reviews will give you results in the way that you want them, so there's no escaping doing some work yourself.

We have found it most useful to follow the following procedures when looking at outputs from systematic reviews and meta-analyses:

1 First check on the statistical result.
2 **IF** statistically significant, proceed to calculate an NNT. Use NNT to estimate the ***treatment specific therapeutic effort*** needed for one outcome and puts some clinical relevance on the result.
3 **IF** this seems sensible, look at what percentage of patients benefit (or are harmed by) with treatment, and use this figure for every day work because this is immediately clinically relevant every time.

It seems to work, if only because we can remember and use percentages quite easily, and because that is what is relevant in everyday practice. The last page of this article contains ***Bandolier***'s NNT calculation sheet. It can be used for clinical trials or systematic reviews. In getting to the NNT it

Figure 7: GPs' ability to explain and used evidence-based medicine terms



makes us look at the real information in a trial or review, and that is probably the most useful part of trying to understand what's going on, and whether there are real benefits for patients.

References

1. DG Altman et al. The revised CONSORT statement for reporting randomised trials. Annals of Internal medicine 2001 134: 663-694.
2. D Moher et al. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement. Lancet 1999, 354:1896-1900.
3. M Deveugele et al. Consultation length in general practice: cross sectional study in six European countries BMJ 2002 325: 472-5.
4. DJ Gavaghan et al. An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. Pain 2000 85:415-24.
5. MD Ferrari et al. Meta-analysis of rizatriptan efficacy in randomised controlled clinical trials. Cephalalgia 2001 21:129-36.
6. A Rostom et al. Prevention of NSAID-induced gastroduodenal ulcers (Cochrane Review). In: *The Cochrane Library,* Issue 3, 2002. Oxford: Update Software.
7. P McGettigan et al. The effects of information framing on the practices of physicians. Journal of General Internal Medicine 1999 14: 633-642.
8. T Fahey et al. Evidence-based purchasing: understanding results of clinical trials and systematic reviews. BMJ 1995 311: 1056-60.
9. CD Naylor et al. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Annals of Internal Medicine 1992 117: 916-21.
10. M Bobbio et al. Completeness of reporting trial results: effect on physicians' willingness to prescribe. Lancet 1994 343: 1209-11.
11. CR Lacy. Impact of presentation of research results on likelihood of prescribing medications to patients with left ventricular dysfunction. American Journal of Cardiology 2001 87: 203-207.
12. M Cranney, T Walley. Same information, different decisions: the influence of evidence on the management of hypertension in the elderly. British Journal of General Practice 1996 46:661-3.
13. A McColl et al. General practitioners' perceptions of the route to evidence based medicine: a questionnaire survey. BMJ 1998; 316:361-5.
14. JM Young, JE Ward. Evidence-based medicine in general practice: beliefs and barriers among Australian GPs. Journal of Evaluation in Clinical Practice 2001 7: 201-210.
15. JA Sterne et al. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. Journal of Clinical Epidemiology 2000 53(11):1119-29.
16. J Nuovo et al. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. JAMA 2002 287: 2813-2814.
17. LS Elting et al. Influence of data display formats on physician investigators' decisions to stop clinical trials: prospective trial with repeated measures. BMJ 1999 318: 1527-1531.

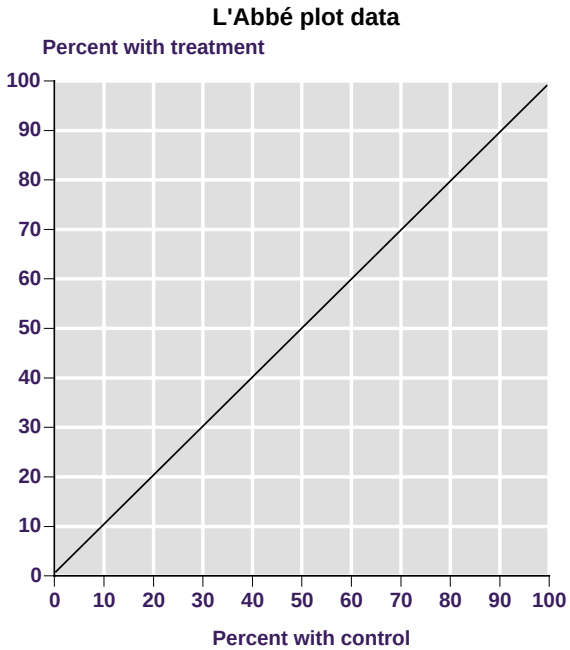
Bandolier's NNT worksheet

A number needed to treat (NNT) is defined by a number of characteristics. This worksheet is designed as an aide memoir for working out NNTs from papers and systematic reviews. First fill in the answers to the questions, where appropriate, graph the data on the L'Abbé plot, and finally do the NNT calculation.

	Question/Action	Answer
A	What is the intervention (ie drug dose & frequency)?	
B	What is the intervention for?	
C	What is the successful outcome (and when or over what time did it occur)?	
D	How many had the intervention?	
E	How many had successful outcome with the intervention?	
F	Express this as a percentage (100 x E/D) and as a proportion (E/D)	
G	What is the control or comparator?	
H	How many people had the control?	
I	How many had successful outcome with the control?	
J	Express this as a percentage (100 x I/H) and as a proportion (I/H)	

Now graph the percentages for the trial on the graph from the ***percentages*** from F and J. This can be done for different outcomes of a trial, or individual trials in a systematic review or meta-analysis.

Now calculate the NNT using the ***proportions*** from F and J.



$$\text{NNT} = \frac{1}{F - J} = \frac{1}{\quad - \quad}$$

$$\text{NNT} = \frac{1}{\quad} = \quad$$