CALCULATING AND USING NNTS

Not everyone is comfortable with calculating and using numbers needed to treat (NNTs). Surveys of GPs, for instance [1], suggest that perhaps under half are confident with them. That means that more than half are not. Some form of aide memoir should be helpful, and this essay includes worked examples and an NNT worksheet that can be copied and used for clinical trials and systematic reviews.

The examples are from the use of proton pump inhibitors (PPIs) and H2-antagonists (H2As) in the short-term healing and long-term maintenance of reflux oesophagitis, from a systematic review done by *Bandolier* in the mid-1990s. The results should be uncontroversial, because now it is widely accepted that proton pump inhibitors are effective. We also look at antibiotics for dog bite infections, and nurse interventions for smoking cessation.

There are, of course, many nuances to all this. *Bandolier* recommends a new book from David Sackett & colleagues - *Evidence-based Medicine: how to practice and teach EBM-* as a cheap and worthwhile acquisition for any thinking doctor, nurse, scientist or manager in the NHS [2].

NNTs can be calculated from raw data using a formula, from odds ratios, or from relative risk reduction and expected prevalence. The first and most important point is that NNTs should not be calculated mindlessly. One of the key issues in systematic review and meta-analysis is whether we are combining similar outcomes measured in similar ways at similar times on similar patients being treated in the same way. This is called clinical homogeneity, and is the first thing to look at, ideally using a L'Abbé plot.

L'Abbé plots

A paper [3] by Kristen L'Abbé and colleagues written ten years ago is regarded by *Bandolier* as one of the most sensible and understandable ever written on systematic reviews. The authors suggest a simple graphical representation of the information from trials. Each point on a L'Abbé scatter plot is one trial in the review. The proportion of patients achieving the outcome with the experimental intervention is plotted against the event rate in controls. Even if a review does not show the data in this way, you can do it yourself if the information is in the review.

For **treatment,** trials in which the experimental intervention was better than the control will be in the upper left of the plot, between the y axis and the line of equality. If experimental was no bet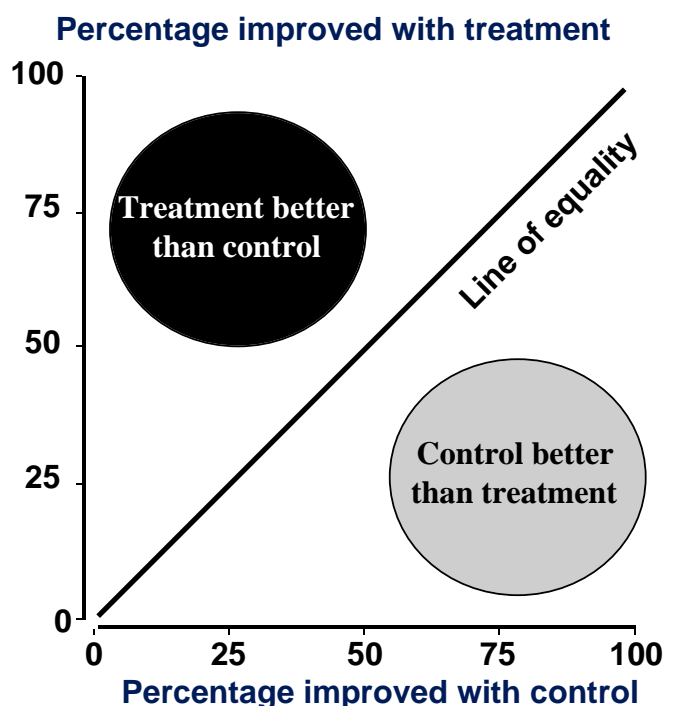ter than control then the point will fall on the line of equality, and if control was better than experimental then the point will be in the lower right of the plot, between the x axis and the line of equality.

For **prophylaxis** this pattern will be reversed. Because **prophylaxis** reduces the number of bad events - such as death after myocardial infarction by the use of aspirin - we expect a smaller proportion harmed with treatment than with control. So if experimental is better than control the trial results should be between the x axis and the line of equality.

These plots give a quick indication of the level of agreement among trials. If the points are in a consistent cloud, that gives some confidence that what we are seeing is a homogeneous effect. But if points are spread all over the graph, and particularly if they cross the line of equality, then we should be concerned about the intervention, or the patients being treated and their condition. This can also be called heterogeneity.

The important point about a L'Abbé plot is that it shows *all* of the extant data on one piece of paper. When combined with numbers in the trial, and a summary measure like NNT, it is a neat way to summarise lots of information.

Figure 1: Idealised L'Abbé plot

Percentage improved with treatment

Treatment better than control

Line of equality

Control better than treatment

Percentage improved with control

www.ebandolier.com

1 Calculating NNTs

The NNT calculation is given below. We need to distinguish between treatments, such as aspirin as an analgesic, and preventative measures, such as aspirin preventing further cardiac problems after myocardial infarction. Using the number outputs from systematic reviews is different depending on which you are looking at. The distinction is between **treatment** and **prophylaxis**. For **prophylaxis**, where fewer events occur in the treated group, the calculation shown will produce negative NNTs. You can use those (the number will be correct), or you can switch the active and control groups around to provide NNTs with a positive sign.

The NNT for **prophylaxis** is given by the equation 1/(proportion benefiting from control intervention minus the proportion benefiting from experimental intervention), and for **treatment** by 1/(proportion benefiting from experimental intervention minus the proportion benefiting from control intervention).

NNTs for **treatment** should be small. We expect large effects in small numbers of people. Because few treatments are 100% effective and because few controls - even placebo or no treatment - are without some effect, NNTs for effective treatments are usually in the range of 2 - 4. Exceptions might be antibiotics. The NNT for Helicobacter pylori eradication with triple or dual therapy, for instance, is 1.2 (Bando*lier* 12).

NNTs for **prophylaxis** will be larger, few patients affected in large populations. So the difference between treatment and control will be small, giving large NNTs. For instance, use of aspirin to prevent one death at five weeks after myocardial infarction had an NNT of 40 (Bandolier 17).

Using absolute risk reduction

The absolute risk reduction (ARR) is the difference between the event rate in the experimental group and the event rate in the control group. It is the denominator in the NNT calculation. Many reviews and trials provide this information, so if you have it and convert it into a proportion, then you can get the NNT by dividing 1 by the ARR:

NNT = 1/ARR

Confidence Intervals

The 95% confidence intervals of the NNT are an indication that 19 times out of 20 the 'true' value will be in the specified range. An NNT with an infinite confidence interval is then but a point estimate; it includes the possibility of no benefit or harm. It may still have clinical importance as a benchmark until further data permits finite confidence intervals, but decisions must take this into account. A method for calculating confidence intervals was given in *Bandolier* 18.

$$NNT = \frac{1}{(IMP_{act} / TOT_{act}) - (IMP_{con} / TOT_{con})}$$

where:

IMP_{act} = number of patients given active treatment achieving the target
TOT_{act} = total number of patients given the active treatment
IMP_{con} = number of patients given a control treatment achieving the target
TOT_{con} = total number of patients given the control treatment

2 Using odds ratios

When it is legitimate and feasible to combine data the odds ratio is the accepted statistical test to show that the experimental intervention works significantly better than control. If a quantitative systematic review produces odds ratios but no NNTs, you can derive NNTs from the Table below [4].

A caveat here is that odds ratios should be interpreted with caution when events occur commonly, as in **treatments**, and odds ratios may over-estimate the benefits of an effect when event rates are above 10%. Odds ratios are likely to be superseded by relative risk reduction because relative risk reduction provides better information in situations where event rates are high [4, 5].

Odds Ratios (OR)

			Preventive											Treatment						
		0.5	0.55	0.6	0.65	0.70	0.75	0.8	0.85	0.9		1.5	2	2.5	3	3.5	4	4.5	5	10
	0.05	41	46	52	59	69	83	104	139	209		43	22	15	12	9	8	7	6	3
Control	0.1	21	24	27	31	36	43	54	73	110		23	12	9	7	6	5	4	4	2
Event	0.2	11	13	14	17	20	24	30	40	61		14	8	5	4	4	3	3	3	2
Rate	0.3	8	9	10	12	14	18	22	30	46		11	6	5	4	3	3	3	3	2
(CER)	0.4	7	8	9	10	12	15	19	26	40		10	6	4	4	3	3	3	3	2
	0.5	6	7	8	9	11	14	18	25	38		10	6	5	4	4	3	3	3	2
	0.7	6	7	9	10	13	16	20	28	44		13	8	7	6	5	5	5	5	4
	0.9	12	15	18	22	27	34	46	64	101		32	21	17	16	14	14	13	13	11

Odds ratios are on the top line and control event rates (CER) down the left hand side. NNTs are in the boxes. So if you have an odds ratio (eg 0.6) and a CER (eg 0.5), then the NNT will be found where they cross (NNT = 8).

NOMOGRAM FOR NNTS
(From [6])

3 Relative risk reduction

Chatelier and colleagues published a useful NNT nomogram in the BMJ [6]. Relative risk reduction - the percentage reduction in risk between the experimental and control group - is used to calculate the NNT for any group in whom the risk of an event happening was known.

This is probably most likely to be used in prophylaxis. If you have a review or paper which gives a RRR (in percent) and you know the susceptibility of your patient for a bad outcome (usually called the 'patient expected event rate', or PEER), then you can find out the NNT of an intervention.

RRR is calculated by dividing the difference between the rate of events in experimental and control group by the rate of events in the control group. So if 10% of patients have a bad event in controls, and only 9% with some intervention, the RRR is (10-9)/10 = 10%. Relative risk reductions happen in **prophylaxis**. With **treatments** we have relative risk increase because we expect more good events. The method works either way.

Say the RRR is 50%, and the PEER is 50%. Then the NNT from the nomogram is 4. But if the RRR is 10% and PEER is 10% then the NNT is about 90.
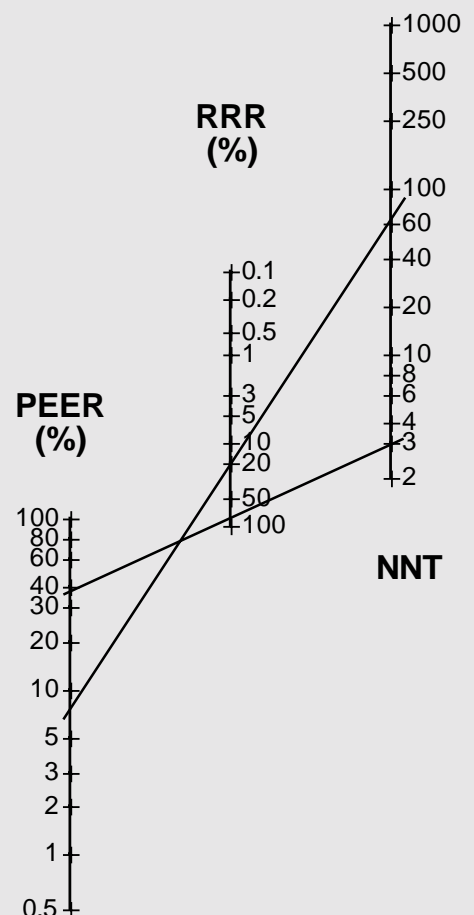
1 REFLUX OESOPHAGITIS

Gastro-oesophageal reflux is the process of reflux of stomach contents into the oesophagus. The consequence is a chemical insult from acid and enzymes. Reflux happens commonly but infrequently in many people, and it does not cause major harm because the natural peristalsis of the oesophagus clears the refluxate back into the stomach. In others where acid reflux from the stomach is persistent, the result is damage to the oesophagus causing symptoms or macroscopic oesophageal damage, and here gastro-oesophageal reflux disease (GORD) can be said to be present.

GORD produces a characteristic set of symptoms, though significant oesophagitis can be present without symptoms. Heartburn is most common; it is often described as gnawing or burning pain behind the sternum, and it may be severe enough to radiate to the arm or jaw. Usually occurring within an hour or so of a meal it can be made worse by lying down. Heartburn can wake the patient at night, and is most frequent in those with the most severe disease. Heartburn occurs occasionally in many people after a fatty or spicy meal, but in GORD the symptoms occur frequently after any sort of meal. Alcohol and coffee also induce symptoms. Antacids relieve symptoms.

Oesophageal damage

The lining of the oesophagus is ill-equipped to resist stomach acids. The stomach has cells which produce a bicarbonate-mucus barrier which protects them from stomach acid: the oesophagus does not have this protective barrier. The result of refluxed stomach acid is to damage the lining of the oesophagus. This damage can be microscopic, but is often macroscopic and seen on endoscopy.

Endoscopic oesophageal damage is graded on a scale of 1-4 with increasing severity of damage.

- Grade 0 is given to normal oesophagus with no macroscopic damage.

- Grade 1 describes an oesophagus with a few areas of erythema, mucosal friability and contact bleeding. These are minor changes regarded as normal by some gastroenterologists.

- Grade 2 oesophagitis has small superficial linear erosions. These tend to lie on the crests or tops of the mucosal folds and may have some surface exudate.

- Grade 3 describes the condition when these erosions coalesce and join around the circumference of the oesophagus. A cobblestone appearance is created by islands of oedematous tissue between the erosions.

- Grade 4 is characterised by extensive mucosal damage with deep ulcers. Strictures may develop, and where this happens there may be less damage above the stricture because the stricture forms a barrier to stomach acids.

The correlation between endoscopic grading and symptoms is not good. Severe symptoms can occur with low grade oesophagitis, and conversely severe oesophageal damage can sometimes occur with few symptoms.

Prevalence

Surveys in the USA have indicated that 44% of the adult population has heartburn at least once a month. Six out of ten of these never consult a GP about it. About 13% of the adult population take some type of indigestion aid at least twice a week. There seem to be few reliable figures on the numbers of patients who present to GPs with reflux symptoms, but a health authority with an adult population of 470,000 will have an estimated 7,500 patients seeing GPs with dyspepsia and almost 3,000 having an endoscopy, over half for oesophageal or gastrointestinal problems, including suspected ulcer.

Treatment options

These include lifestyle change and use of antacid or alginates; none of these is particularly effective. Mucoprotective agents and motility stimulants may be used, but suppression of acid secretion with H2A or PPI is the most common form of treatment. Antireflux surgery is said to be useful in some patients, and to be effective.

Systematic review

Reports were sought of comparisons between any proton pump inhibitor (PPI) and any histamine-2 antagonist (H2A) in reflux oesophageal disease with endoscopic healing as the outcome measure. Both short-term healing and long-term maintenance were included. Papers were included in the systematic review of effectiveness if they fulfilled the following criteria:-

- Full journal publication
- Randomised trial
- Compared PPI with H2A, or
- Compared either PPI or H2A with placebo
- Examined gastro-oesophageal reflux disease (GORD), erosive oesophagitis or gastritis, or reflux oesophagitis
- Had endoscopic healing as an outcome or
- Had adverse event outcomes
- Had short-term outcomes at 4 and/or 8 weeks, or
- Had long-term maintenance outcomes at 6 or 12 months

Short-term healing

Twenty-three reports with 5,118 patients fulfilled the inclusion criteria. One report had no endoscopic healing data, but did have adverse event information. Of the reports with endoscopic healing, ten (1393 patients) compared omeprazole with ranitidine, two reports (339 patients) omeprazole with cimetidine and three reports (525 patients) lanzoprazole with ranitidine. Quality scores were high using a validated scale from 1 to 5 [7]. Four studies had a score of 2, three of 3, eleven of 4 and five of 5. The median score was 4. *Bandolier* 33 showed how it is important to use stud-

Table 1: Short-term endoscopic healing of reflux oesophagitis with omeprazole and ranitidine

| Trial	Total in trial	EER	CER	NNT	Relative Risk	Odds Ratio
Klinkenberg-Knol et al, 1987	51	88	38	2.0	2.3	8.0
Sandmark et al, 1988	152	75	43	3.1	1.7	3.7
Havelund et al, 1988	162	86	60	3.9	1.4	3.8
Vantrappen et al, 1988	61	81	50	3.2	1.6	3.8
Zeitoun et al, 1989	156	87	56	3.2	1.5	4.4
Lundell et al, 1990	98	86	38	2.1	2.3	7.7
Bate et al, 1991	283	69	38	3.2	1.8	3.4
IRSG, 1991	172	73	62	8.3	1.2	1.7
Bianchi-Porro et al, 1992	60	77	33	2.3	2.3	5.6
Green et al, 1995	196	55	25	3.3	2.2	3.5
Overall	**1391**	**78**	**44**	**3.3**	**1.7**	**3.7**

Summary data from randomised controlled trials comparing omeprazole and ranitidine for endoscopic healing of erosive oesophagitis after eight weeks. EER is 8-week endoscopic healing rate with omeprazole and CER the 8-week endoscopic healing rate with ranitidine.

ies of high quality (score of 2 or more on this scale) to avoid over-estimating treatment effects.

Information from randomised controlled trials which compared ranitidine and omeprazole with endoscopic healing of erosive oesophagitis after eight weeks is shown in Table 1. The overall NNT for omeprazole compared with ranitidine was 3.3. This means that for every three patients with erosive reflux oesophagitis treated with omeprazole, one will be healed who would not have been healed if they had been treated with ranitidine.

Odds ratios, relative risk and NNTs have been calculated, and the rate of healing in the omeprazole and ranitidine groups is shown in Table 1 for each trial.

Figure 2: Short-term endoscopic healing of reflux oesophagitis with omeprazole and ranitidine (percent in each trial)
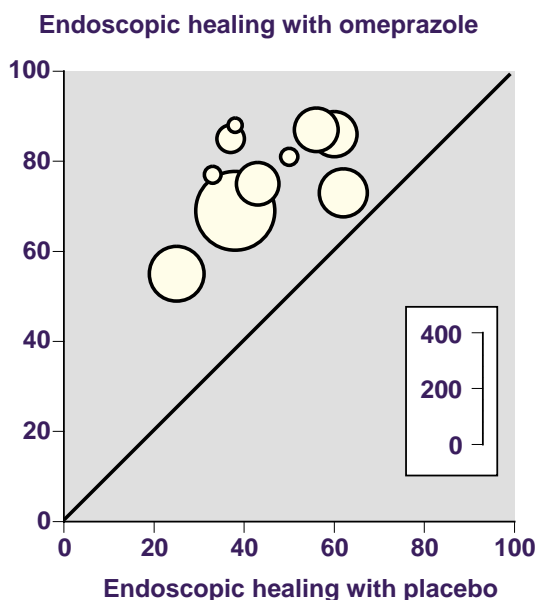
L'Abbé plot

Figure 2 shows that all the studies are well to the upper left of the line of equality meaning that in all trials omeprazole was better than ranitidine. The L'Abbé plot also provides extra information that the NNT does not. While the NNT gives us the *treatment-specific* benefit of omeprazole over ranitidine, the L'Abbé plot shows us the overall effect of treatment. So we can see that about 80% of patients are healed with omeprazole while only about 45% are healed with ranitidine.

This is useful information for patients, since it also conveys information the patient most wants - how likely am I to get better (or be harmed) with this treatment. For an individual patient this is an absolute which includes both treatment-specific and non-specific effects of treatment. Here the non-specific effect is contained within the results for ranitidine, but if we had been examining a treatment compared with placebo, the L'Abbé plot will give us that also.

NNT from ARR

For the overall results, the proportion getting benefit with omeprazole was 78%, or 0.78, and for ranitidine it was 44% or 0.44. So the NNT calculation becomes:

$$NNT = 1/ARR = 1/(0.78-0.44) = 1/0.34 = 3$$

which is close to the 3.33 calculated from raw data in the review.

NNT from OR

Looking at the Table of odds ratios and NNTs, if we go to the column with the odds ratio nearest the overall of 3.7, and track along the control event rate nearest that in our review for ranitidine of 0.4, we obtain an NNT of 3. Again close to the overall NNT of 3 calculated from raw data in the review.

Table 2: Long-term endoscopic healing of reflux oesophagitis with omeprazole and ranitidine

Trial	Total in trial	E E R	C E R	N N T	Relative Risk	Odds Ratio
Lundell Om 20 vs Ran 300	68	50	9	2.4	5.5	5.7
Smith Om 20 vs Ran 300	257	61	21	2.5	2.8	5.0
Dent Om 20 vs Ran 300	104	89	25	1.6	3.5	13.0
Hallerback Om 20 vs Ran 300	259	72	45	3.9	1.6	3.0
Vigneri Om 20 vs Ran 450	70	80	49	3.2	1.7	3.9
Overall	**7 5 8**	**7 0**	**3 0**	**2.8**	**2.1**	**4.2**

Summary data from randomised controlled trials comparing omeprazole and ranitidine for maintenance of endoscopic healed erosive oesophagitis after time point nearest to one year. EER is one-year still-healed rate (on endoscopy) with omeprazole and CER the one-year still-healed rate (on endoscopy) with ranitidine.

NNT from RRR

The relative risk increase from the table is $(78-44)/44 = 77\%$. The PEER is 44%, and our NNT from the nomogram is 3.

Long-term maintenance

Seven reports with 1,635 patients fulfilled the inclusion criteria. Four reports (1094 patients) compared omeprazole with ranitidine, one omeprazole with placebo, one lanzoprazole with placebo and one omeprazole and ranitidine alone and in combination with cisapride. The most commonly used doses were omeprazole 20 mg and ranitidine 300 mg daily. Two studies had a quality score of 3, four of 4 and one of 5.

Figure 3: Long-term endoscopic healing of reflux oesophagitis with omeprazole and ranitidine (percent in each trial)
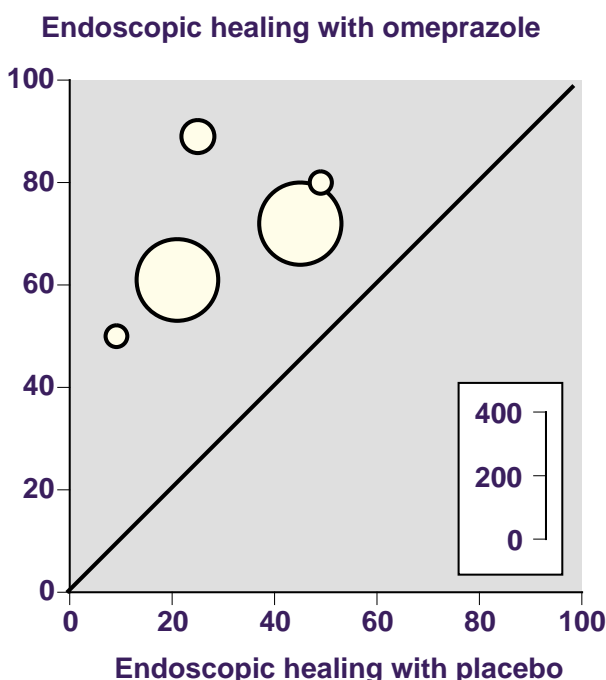
Information from randomised controlled trials which compared ranitidine and omeprazole with endoscopic healing of erosive oesophagitis after eight weeks is shown in Table 2. The overall NNT for omeprazole compared with ranitidine was 2.8. This means that for every three patients with healed erosive reflux oesophagitis treated with omeprazole, one more will still be healed after one year who would not have been if they had been treated with ranitidine.

L'Abbé plot

Figure 3 shows that all the trials are above the line of equality, but with much wider spread. The one-year remission rates with ranitidine vary from 9% to 49%, and for omeprazole between 50% and 89%. The difference - the treatment-specific effect - is much the same though, giving consistent NNTs.

NNT from ARR

For the overall results, the proportion getting benefit with omeprazole was 70%, or 0.70, and for ranitidine it was 30% or 0.30. So the NNT calculation becomes:

$$NNT = 1/ARR = 1/(0.70\text{-}0.30) = 1/0.40 = 2.5$$

which is close to the 2.8 calculated from raw data in the review.

NNT from OR

Looking at the table of odds ratios and NNTs, if we go to the column with the odds ratio nearest the overall of 4.2, and track along the control event rate nearest that in our review for ranitidine of 0.3, we obtain an NNT of 3. Again close to the overall NNT of 2.8 in the review.

NNT from RRR

The relative risk increase from the table is $(70-30)/30 = 133\%$. The PEER is 30%. The NNT from the nomogram is hard to calculate because of the high RRR, but would be about 2.

Variation in treatment and control

Using systematic reviews in this way teaches you just how variable are the effects of both treatment and control in randomised trials. It is legitimate to be surprised, but after quite a short time it seems that this is the norm.

The reasons are probably complex, but much of the variability will be just random chance. In many circumstances patients can have quite wide patterns of response to a treatment, but trial size for treatments is often relatively small, because trials are hard to do. Gathering data together in systematic review and meta-analysis gives much more power than the single trial in almost all circumstances, and especially for reviews of treatments. Seeing such variability also teaches caution when you are faced with a single trial with apparently excellent (or hopeless) results.

What we have in this example is a series of well-conducted trials, though some are small. We have compared two treatments, at standard doses, in similar patients, using an objective outcome measured in much the same way over the same treatment duration. The results of individual trials are, not surprisingly, quite similar to one another.

Overall the NNT for short-term healing was 3.3, though the range in individual trials was as low as 2.0 and as high as 8.3. Six out of 10 had NNTs between 3 and 4. A good reason for using a systematic review and meta-analysis is that it allows us to look at **all** the information, and does not mislead because of small numbers.

2 Dog bite infection

Suppose a woman presents with a dog bite. Her immune system is compromised by steroid therapy for asthma. Should we give prophylactic antibiotics to prevent infection? We know from a quantitative systematic review of RCTs [9] that there is evidence for benefit with an overall NNT of 16. That means that we would have to give prophylactic antibiotics to 16 patients to prevent an infection in one of them.

How can we apply this to our patient?

She is immunocompromised, so her risk of becoming infected is higher than the non-compromised patients in the review. We estimate her increased risk (usually called F), to be 5 times greater than the 16% average rate of infection in the review (though in individual studies risk varied between 3% and 46%). Assuming a constant relative risk, the estimated NNT corresponding to an F of 5 is then NNT/F = 16/5 = 3 [2].

So while prophylactic antibiotic treatment of dog bite to prevent infection may not be worthwhile for all patients (NNT of 16), it may well be so for our particular patient (NNT of 3).

Suppose we live in Middlesborough? We know from an RCT that infection rates there are about 50%, so we might be likely to treat all patients. For patients in Middlesborough,

the "patient expected event rate", or PEER, is 0.5 compared to the 0.16 average found in the review. The review gave us an odds ratio of 0.6 for prophylactic antibiotics.

If we look down the line of 0.6 in the Table of odds ratios and NNTs and stop at a control event rate (our PEER) of 0.5, then we find an NNT of about 8. Now if half our patients bitten by a dog are going to get an infected wound, and by using antibiotics we can stop that happening just once in every eight times, then we save six patients in every 100 from having an infected bite.

So even in Middlesborough our prophylactic antibiotics won't stop every infection, but perhaps enough to make it worthwhile. Perhaps changes in practice and knowledge will make it more likely that we want to intervene in this way [8].

3 Nursing interventions for smoking cessation

Calculating NNTs can occasionally get us into some difficulty, especially where the NNT is large, and where there may be clinical heterogeneity between trials. An example is nursing interventions for smoking cessation [10]. This Cochrane review included interventions in both hospitals and primary care where rates of smoking cessation differed considerably. It was the subject of an interesting debate about the methods we might use to calculate NNT [11, 12, 13].

The review included 10 trials with about 6,000 patients. Overall the result was of borderline statistical significance. The L'Abbé plot for all trials is shown in Figure 4, where

Figure 4: Nursing interventions for smoking cessation (filled circles in hospital, open circles in primary care)
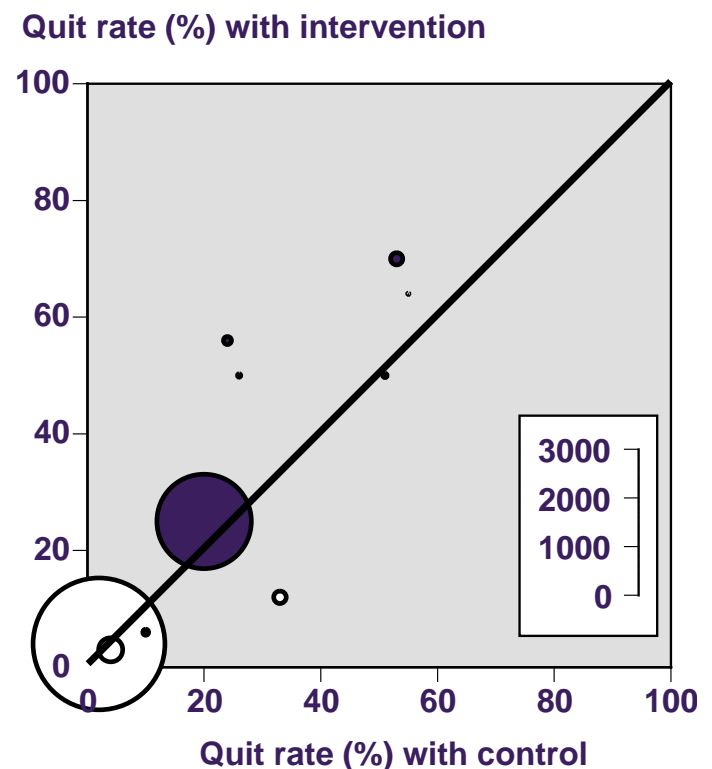
the trials in hospital are filled circles, and those in primary care are open circles. Two trials were large, and the rest quite small. Quit rates without nursing intervention ranged from 2% to 55%. Smoking was defined as at least one cigarette a day in at least one of the primary care studies, and in the hospital studies the patients included those undergoing coronary artery bypass and surgery for cancer.

Table 3 shows an analysis [12] based on all trials, and those in hospital and in primary care analysed separately. There are clearly different responses to nursing intervention to stop smoking, with not much effect in primary care and some effect in hospital. But in hospital, even without an intervention, 25% of patients stopped smoking, while in primary care without an intervention 4% did. Are the patients the same?

In hospital, a nursing intervention in 14 patients could result in one of them stopping smoking who would not have done so if there had been no intervention. In primary care an infinite number of patients would have to have a nursing intervention for one of them to stop smoking.

Logic and results would suggest that the two situations are not the same, and that perhaps combining these results is less helpful than analysing them separately. Nursing interventions appear to be useful in hospital patients, and much less so in primary care.

Calculation problems

The three papers together give some important insights into the pitfalls of calculating NNT or any other outputs from clinical trials. As well as tribulations about what trials should be combined with which, there are issues over random effects versus fixed effects models (statistical assumptions), Simpson's paradox (unbalanced groups), and whether we calculate NNT from aggregated information or from statistical outputs [11, 12, 13].

For using NNTs at the coal face, these delicate, but important, issues should not affect us. They are unlikely to be a problem when NNTs are 20 or less.

Contemplating infinity

One thing that Table 3 reminds us is that NNTs can be negative as well as positive. A negative NNT is where there are more good events with control than with treatment. Then a treatment can be doing more harm than good [14]. Because NNTs are the reciprocal of the risk difference (ARR), when

the ARR is zero (no difference between EER and CER), the NNT must be 1 divided by zero, which is infinity. The confidence interval of an NNT can therefore be positive, go through infinity, and into a negative NNT, or the other way around.

Where the difference between EER and CER is zero or small, the result of a trial or meta-analysis will probably be that there is no statistically significant difference between experimental treatment and control, as with nursing interventions for smoking cessation in primary care in Table 3. Some people stop at that point, and would not quote an NNT at all. Other authorities insist that the NNT and confidence interval be quoted, because a negative NNT in a confidence interval shows that the "true" answer could include doing more harm than good [14].

Number needed to harm

The number needed to harm (NNH) is in all ways similar to a NNT, but refers to adverse or harmful events, and can be general (withdrawal from treatment) or specifically related to a particular adverse event. NNH will be dealt with in a separate article.

Comment

There will be circumstances where systematic reviews will not yield information to generate L'Abbé plots, NNTs, relative risk, or even odds ratios. There are times when the information for quantitative systematic review is just not available. Where it is, we can use it to make choices in our practice, both as a matter of policy for treating a particular decision, or in making a decision for a particular patient.

The NNT is a tool. Like any tool, when used appropriately it will be helpful and effective. What we have to do is to establish that in any given situation what the rules are for using the tool correctly. In meta-analysis, that usually means making sure that we are combining information about similar outcomes, about similar interventions, in similar patients, in similar intensity, and over the same period of time. There will be a few occasions with large NNTs when different methods of calculation will give different results for the NNT.

What it gives us is a measure of the therapeutic effort required to obtain one beneficial event. NNTs are consequently useful in comparing the utility of treatments by their NNTs when we have comparable trials with comparable outcomes in comparable patients.

Table 3: Nursing interventions for smoking cessation

Setting	Number of studies	Number quitting/total		Percent quitting (95%CI)		Relative benefit (95%CI)	NNT (95% CI)
		Intervention	Control	Intervention	Control		
Hospital	7	435/1367	318/1295	32 (30 to 34)	25 (23 to 27)	1.3 (1.1 to 1.6)	14 (9 to 26)
Primary care	3	111/2453	41/1006	4.5 (3.7 to 5.3)	4.1 (2.9 to 5.3)	0.8 (0.3 to 2.6)	222 (52 to -98)
Combined	10	546/3820	359/2301	14 (12 to 16)	16 (14 to 18)	1.2 (0.9 to 1.6)	-76 (184 to -32)

NNTs to inform patients?

This is a difficult area. Because an NNT is treatment-specific, it will not include all the power of an intervention - a response with placebo, for instance, or no treatment, indicating what happens when we do nothing. Patients want to know their chance of getting better or being harmed, and that includes influences from all sources. The best analgesics have NNTs of 2 for at least 50% pain relief (a high hurdle), which implies that at least half the patients will achieve at least 50% pain relief because of the analgesic. Another 20% will have this outcome if we do nothing, so that reality is that 70% achieve at least 50% pain relief with the analgesic, which sounds better and reflects the reality.

But that is a simple example. Most circumstances are more complex. The LBBH (likelihood of being helped or harmed) has been suggested as one way of presenting information to patients [15], but there is a clear need for more empirical research to provide evidence on how best to do this. The simple answer is that there is, as yet, no simple answer to a question about how to use NNTs when conveying information to patients. There may be good reasons why NNTs are not the best way to do this.

NNT calculation sheet

Page 10 has an NNT calculation sheet that can be copied and used for examining trials and meta-analyses, to produce L'Abbé plots, and to calculate NNTs.

References:

1   A McColl et al. General practitioners' perceptions of the route to evidence based medicine: a questionnaire survey. BMJ 1998 316:361-5.
2   DL Sackett, WS Richardson, W Rosenberg, RB Haynes. Evidence-based Medicine: how to practice & teach EBM. Churchill Livingstone. ISBN 0-443-05686-2.
3   L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987 107:224-33.
4   Sackett DL, Deeks JJ, Altman DG. Down with odds ratios! Evidence Based Medicine 1996 1:164-6.
5   Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. J Clin Epidemiol 1994 47:881-889.
6   G Chatelier et al. The number needed to treat: a clinically useful nomogram in its proper context. British Medical Journal 1996 312: 426-9.
7   AR Jadad, RA Moore, D Carroll, C Jenkinson, DJM Reynolds, DJ Gavaghan, HJ McQuay. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Controlled Clinical Trials 1996 17: 1-12.
8   P Cummings. Antibiotics to prevent infection in patients with dog bite wounds: a meta-analysis of randomized trials. Annals of Emergency Medicine 1994 23: 535-40.
9   F Moore. "I've just been bitten by a dog". British Medical Journal 1997 314: 88-9.
10 Rice VH, Stead LF. Nursing interventions for smoking cessation (Cochrane Review). In: *The Cochrane Library, Issue 1, 2001*. Oxford: Update Software.
11 CJ Cates. Simpson's paradox and calculation of number needed to treat from meta-analysis. BMC Medical Research Methodology 2002 2:1 (http://www.biomedcentral.com/1471-2288/2/1
12 RA Moore et al. Pooling data for number needed to treat: no problem for apples. BMC Medical Research Methodology 2002 2:2 (http://www.biomedcentral.com/1471-2288/2/2
13 DG Altman, JJ Deeks. Meta-analysis, Simpson's paradox, and the number needed to treat. BMC Medical Research Methodology 2002 2:3 (http://www.biomedcentral.com/1471-2288/2/3
14 DG Altman. Confidence intervals for the number needed to treat. BMJ 1998 317: 1309-1312.
15 I Chalmers. Applying overviews in meta-analysis at the bedside. Journal of Clinical Epidemiology 1995 48: 67- 70.

Bandolier

Independent evidence-based health care

Bandolier is an independent print and Internet journal about health care, using evidence-based medicine techniques to provide advice about particular treatments or diseases for healthcare professionals and consumers. **Bandolier** has been published monthly since February 1994 and on the Internet since 1995.

Bandolier's NNT calculator

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.

1 Answer the questions

	Question/Action	Answer
A	What is the intervention (dose & frequency)?	
B	What is the intervention for?	
C	What is the successful outcome (when, over what time did it occur)?	
D	How many had the intervention?	
E	How many had successful outcome with the intervention?	
F	Express as a percentage (100 x 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 as a percentage (100 x I/H)	

2 Graph the percentages from F and J on the L'Abbé plot

This can be done for different outcomes of a trial, or individual trials in a systematic review or meta-analysis.

Percentage improved with treatment

Percentage improved with control
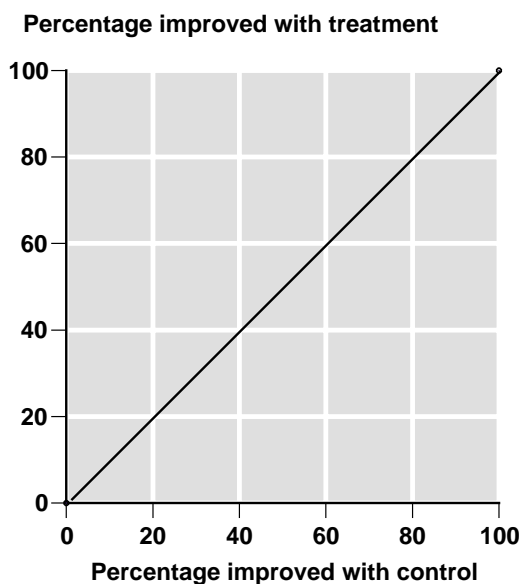
3 Now calculate the NNT using the percentages from F and J.

Again, this can be done for the different outcomes of a trial, or the overall proportions from a meta-analysis.

$$ NNT \quad = \quad \frac{100}{F - J} \quad = \quad \frac{100}{\boxed{} - \boxed{}} $$

$$ NNT \quad = \quad \frac{100}{\boxed{}} \quad = \quad \boxed{} $$

The perfect NNT is 1, where everyone gets better with treatment and no-one gets better with control. NNTs of close to 1 are seen with, for instance, antibiotic treatment of susceptible organisms. NNTs of 2-5 represent effective therapies, but NNTs of 20 or more may be useful in prophylaxis.