How to read an article about therapy in 5 minutes

Paul L P Brand, Zwolle (NL)

The number of scientific articles in print increases each year, and it is impossible to read all of these, even when limited to one's own field of interest. For example, almost 30'000 papers on childhood asthma are published each year (PubMed search). From time to time, however, even the busiest clinician will want to read a particular article carefully and completely. It would be extremely helpful if such designated article reading could be done quickly and efficiently, without jeopardizing the quality of information gleaned from the article. In this review, I propose a simple 5-step method to read an article on therapy critically and carefully, which, after a bit of exercise, can be done in 5 minutes. The five steps are presented in *table 1*.

1. be active!
2. read the methods section
3. analyze results for yourself
4. no surrogate end points, no subgroup analyses
5. draw your own conclusion

Table 1: **5-step method to read an article about therapy in 5 minutes**

Step 1. Be active!

We all know that physical activity is healthy, and promoting exercise in childhood is considered to be one of the key measures to prevent obesity. Yet, when reading a scientific paper, we usually take the «couch potato» approach: we sit on our seats, we watch the screen (or the printed paper), and we eat it all up. As a result, we usually take the results (and the authors' conclusion) at face value. Needless to say, this is not a very useful (nor a very exciting) way of collecting knowledge.

Conversely, much more information can be gleaned from a paper if the reader takes an active approach. Just by asking two questions before one starts reading a paper, the reader can quickly decide whether the paper is worth reading at all (*table 2*).

1. is the outcome measured an important one in my view?
2. what effect size do I think is relevant?

Table 2: **Step 1: be active! Questions to consider before you start reading**

Sometimes, the answer to question 1 can be easily obtained by just looking at the title. For example, in the Impact study, the title tells you that the study is about the effects of palivizumab on hospitalizations for respiratory syncytial virus (RSV) infections. (IMpact Study Group, 1998) Most commonly, however, this information must be obtained from the Methods section (*see step 2*). Perhaps even more importantly, the size of the treatment effect that one considers useful can be decided upon before one starts reading the paper. Quite commonly, effects of treatment are described to be «significant», which is useful from a statistical point of view, but utterly useless from a clinical point of view. For example, in a clinical trial comparing budesonide and fluticasone in childhood asthma, the effect on morning peak flow of fluticasone was said to be «significantly higher» than that of budesonide, but in fact, the difference in peak flow between the two groups was only 7 L/min, (Hoekx et al., 1996) which, of course, is completely irrelevant (*figure 1*).

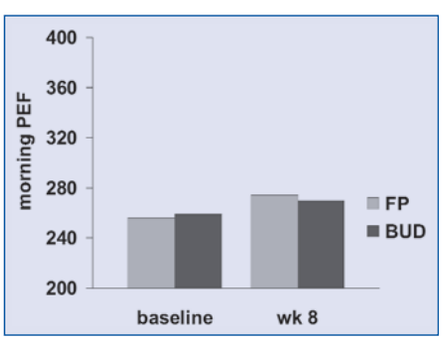

Figure 1: Results of a study in which the effects on morning peak expiratory flow (PEF) of 8-wk treatment with fluticasone proprionate (FP) and budesonide (BUD) by dry powder inhaler were compared in children with asthma. According to the abstract, morning PEF was «significantly higher» in the FP group than in the BUD group. The mean difference in morning PEF was 7 L/min (95% confidence interval 1–14 L/min) which is an irrelevant difference clinically.

Step 2. Read the methods section

Although this sounds boring and uninviting, reading the methods section is the key towards useful interpretation of data from any article. In particular, the validity of study design needs to be determined. In a study on treatment effects, the points mentioned in *table 3* are of particular importance. (Guyatt et al., 1993)

1. was the assignment of treatment to patients randomized? And was the randomization list concealed?
2. was follow-up of patients sufficiently long and complete?
3. were all patients analyzed in the groups to which they were randomized?
   Some less important points:
4. were patients and clinicians kept blind to treatment?
5. were groups treated equally, apart from the experimental therapy?
6. were the groups similar at the start of the trial?

Table 3: **Step 2: Read the methods section, and decide whether the results of this study on treatment are valid.**

The importance of randomization cannot be overemphasized. Any study that is not randomized is an observational study, which are useful for generating hypotheses, not for testing them. (Guyatt et al., 1993) For example, in a quite well-known observational study, children using inhaled steroids had far better lung function than children whose patients refused inhaled steroids because of fear of side effects, and who used cromolyn instead. (Agertoft and Pedersen, 1994) Quite correctly, the authors presented this as an observational study, and designed a proper prospective randomized controlled study to test the hypothesis that their study generated. (Pauwels et al., 2003).

Exceptionally, a study is presented as a randomized trial when, in fact, it is not. This was the case with a Finnish study comparing the effects of hypoallergenic versus cow's milk formula as supplementary feeding during the first days of life on the development of atopic disease later in life. This study was presented as a randomized study, both in the title and in the abstract. After reading the Methods section, however, it was obvious that allocation to treatment was actually not random, but «depended on the month of birth and the hospital». (Saarinen et al., 1999)

Such systematic randomization is known to introduce considerable bias. (Altman and Dore, 1990) Consequently, the study should be regarded as observational, and the results should be interpreted with caution.

Similarly, lack of concealment of allocation can undermine a study's validity. If an allocation list is open for investigators to view, purposeful (biased) allocation is a real possibility. As a rule of thought, if the methods section does not mention allocation to be concealed, it should be regarded as unblinded, and, hence, biased. For example, in a study on the preventive effect of probiotics on the development of atopic disease, concealment of allocation was not mentioned, and this issue has been raised in evidence-based journals discussing this study. (Kalliomäki et al., 2001) The prevalence of atopic eczema in the placebo group in this study is unusually high (almost 50%) when compared to similar groups of patients in other European countries, and it is possible that this was caused by purposeful open allocation.

When looking at *table 3*, it may come as a surprise that blinding of treatment is listed under «some less important points». This is because the degree of bias introduced by lack of randomization or unblinding of treatment allocation is larger than the degree of bias introduced by unblinding. This is particularly true when «hard» end points are used, such as death or results of a laboratory test. Conversely, if «soft» subjective end points such as symptom scores or diagnoses are used, unblinding may become a real problem. For example, in an unblinded study on the effects of chiropractic therapy on crying behaviour in infants with colic, the therapy was found to be highly effective as judged by the hours of crying parents recorded in a diary; (Wiberg et al., 1999) in a similar, but blinded study, both the active and the sham chiropractic treatment were associated with significant improvements in crying, indicating a strong placebo effect. (Olafsdottir et al., 2001)

Step 3. Analyze results for yourself

Reading scientific articles *really* becomes interesting if one begins to analyze results for oneself. In keeping with step 1 (Be active!), the reader is encouraged to determine *before* starting to read the paper which effect size one considers to be relevant (for a continuous end point), or which number needed to treat (NNT) one thinks to be acceptable (*table 4*).

Studies with a continuous end point: which treatment effect do I think is relevant? Calculate mean difference and 95% CI of difference *Studies with a dichotomous end point:* which number needed to treat (NNT) do I consider acceptable? Calculate NNT as 1 divided by the *absolute* risk reduction

Table 4: **step 3: Analyze results for yourself**

Studies with a continuous end point

Consider the study in *figure 1* once again. The mean difference in morning PEF after 8 wks of treatment was 7 L/min, with a 95% confidence interval (CI) from 1 to 14 L/min. This means with 95% confidence, the true difference in morning PEF between asthmatic children treated with FP and BUD would lie between 1 and 14 L/min. Is this clinically relevant? I don't think so.
A useful way of looking at this graphically is represented in *figure 2*.

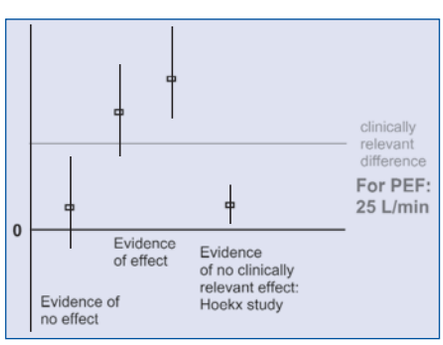

Figure 2: A graphical method to distinguish between statistical significance and clinical relevance. The squares represent the difference in treatment effect between two study groups; the vertical lines represent the 95% CIs around this mean difference. The thick horizontal line represents the zero difference line; the thin horizontal line represents a clinically relevant difference (the reader decides where this line should lie!). A CI crossing the zero (thick) line represents a study with evidence of no effect. Studies with a CI above the zero line represent studies with evidence of effect, but if the CI is below the thin line, the effect is clinically irrelevant (which, in practice, is of equal importance as a study with evidence of no effect). The Hoekx study is an example of such a study. (Hoekx et al., 1996) This figure is adapted from a textbook in Dutch on evidence based medicine. (Offringa et al., 2003)

Studies with a dichotomous end point

The usual outcome parameter in such studies is a percentage. In the Impact study mentioned previously, the percentage of hospital admissions for RSV infections was 4.8% in the palivizumab group and 10.6% in the placebo group. (IMpact Study Group, 1998) Such a difference is usually presented as a *relative risk reduction,* probably because the figure (in this case 55%!) looks impressive. It is the absolute risk reduction (ARR), however, that matters, because the *number needed to treat* can be easily computed as the inverse of this ARR. In this example, the ARR is 10.6% – 4.8% = 5.8%, and the NNT is 1/5.8% = 17.2. This means that 17 patients should be treated with palivizumab in order to prevent one RSV hospitalization. Whether or not this is relevant or acceptable, is up to the reader to decide. In this particular case, the discussion has focused around the cost effectiveness of this approach.

Step 4: no surrogate end points, no subgroup analyses

In step 1, we have decided which outcome measure we consider important. Facing difficulties in obtaining the most relevant outcomes easily and reliably, clinical trial designers often resort to measuring something that can be easily measured, but has a more or less distant relationship to the truly relevant outcome. This is called a surrogate end point. *Surrogate end points* are fraught with difficulties, which may be best explained with an example. (Greenhalgh, 1997) Obviously, in HIV disease, the relevant outcome parameter is death, but in clinical trials, surrogate end points such as CD4 counts are commonly used because they are easily measurable. It is assumed that CD4 counts predict death, but in a large and quite famous study, early treatment with zidovudine in HIV patients slowed the decline of CD4 counts… but had no effect on survival (paper quoted in Greenhalgh 1997). Thus, the reader should decide if the outcome measure in a particular therapeutic study is a relevant one, and if it is a surrogate end point, whether the evidence linking the surrogate end point to the actual outcome measure of interest is sound. It usually isn't.

The issue of subgroup analyses is even more contentious. It is common practice for a clinical trial to report results both from the study as a whole and in different subgroups. This is *only* OK if two conditions are met (*table 5*). (Yusuf et al., 1991)

1. the subgroup analysis was planned before the study (and not reported because the data looked so exciting)
2. the result in the subgroup is in the same direction as in the whole group

Table 5: **conditions that must be met before results from subgroup analyses can be accepted**

As a result of biological variation, results in different subgroups can be expected to differ from the overall study result, but this is a quantitative rather than a qualitative phenomenon. A subgroup result which is in the opposite direction or strikingly different from the overall result should, therefore, raise an eyebrow. The most common example of this is a study with an overall negative result («evidence of no effect», *see step 3*), but a significant treatment effect in one or more subgroups. For example, this was the case in a study where cetirizine was given to infants with atopic eczema in the hope of preventing the development of asthma. (Warner, 2001) The overall study result was zero, but infants with house dust mite or grass pollen allergy were found to benefit from cetirizine therapy. The paper concluded that cetirizine «truly delays or, in some cases, prevents the development of asthma in a subgroup of infants with atopic dermatitis…». From a methodological point of view, this conclusion is invalid. The study was not designed to answer the research question in a subgroup, and the subgroup analysis was not planned before the study started (in technical terms, it was a «post hoc analysis»). If the overall result was zero and one subgroup benefited from active therapy, then another subgroup must have benefited from placebo therapy. Clearly, we wouldn't draw the conclusion that these patients should be treated with placebo. The inverse conclusion, however, is just as invalid.

What *can* be done is consider subgroup analyses as hypothesis generating exercises (just like the observational studies mentioned in step 2). It is quite valid to conclude from the cetirizine study that infants with atopic eczema and house dust mite or grass pollen sensitization *might* benefit from cetirizine, but this hypothesis must be tested in a study that was designed to do so before this conclusion can be accepted. Meanwhile, patients in clinical practice should not be treated along this hypothesis (at least not if one wants to practice evidence based medicine).

Step 5: draw your own conclusion

The example of the cetirizine study brings us to step 5: draw your own conclusion. Clearly, my conclusion differs from the one drawn by the authors of the study, and I am convinced that *my* conclusion is the right one. Even though the paper was published in a major high-quality journal, I take the liberty of disagreeing – respectfully – with its authors' conclusion. This, perhaps, is the most important of my five step method. I would encourage each and every reader to draw their own conclusions after reading an article. This can be done quite easily – just follow the steps in *table 1*, and with a little practice, you will be able to do so yourself. I guarantee you: it is a rewarding exercise, and makes scientific reading much more worthwhile (and fun).
One final note: if you need a memory aid to remember the five steps – it's easy: the first letters of each of the five steps form the last name of the author of this paper.

Reference List

- IMpact Study Group. Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. Pediatrics 1998; (102): 531–537.
- Hoekx J C M, Hedlin G, Pedersen W, Sorva R, Hollingworth K, Efthimiou J. Fluticasone propionate compared with budesonide: a double-blind trial in asthmatic children using powder devices at a dosage of 400 mg.day-1. Eur Respir J 1996; (9): 2263–2272.
- Guyatt G H, Sackett D L, Cook D J, Evidence-Based Medicine Working Group. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? JAMA 1993; (270): 2598–2601.
- Agertoft L, Pedersen S. Effects of long-term treatment with an inhaled corticosteroid on growth and pulmonary function in asthmatic children. Respir Med 1994; (88): 373–381.
- Pauwels R A, Pedersen S, Busse W W, Tan W C, Chen Y-Z, Ohlsson S V, Ullman A, Lamm C J, O'Byrne P M. Early intervention with budesonide in mild persistent asthma: a randomised, double-blind trial. START investigators group. Lancet 2003; (361): 1071–1076.
- Saarinen K M, Juntunen-Backman K, Järvenpää A-L, Kuitunen P, Lope L, Renlund M, Siivola M, Savilahti E. Supplementary feeding in maternity hospitals and the risk of cow's milk allergy: a prospective study of 6209 infants. J Allergy Clin Immunol 1999; (104): 457–461.
- Altman D G, Dore C J. Randomisation and baseline comparisons in clinical trials. Lancet 1990; (335): 149–153.
- Kalliomäki M, Salminen S, Arvilommi H, Kero P, Koskinen P, Isolauri E. Probiotics in primary prevention of atopic disease: a randomised placebo-controlled trial. Lancet 2001; (351): 1076–1079.
- Wiberg J M M, Nordsteen J, Nilsson N. The short-term effect of spinal manipulation in the treatment of infantile colic: a randomized controlled clinical trial with a blinded observer. J Manipulative Physiol Ther 1999; (22): 517–522.
- Olafsdottir E, Forshei S, Fluge G, Markestad T. Randomised controlled trial of infantile colic treated with chiropractic spinal manipulation. Arch Dis Child 2001; (84): 138–141.
- Offringa M, Assendelft W J J, Scholten R J P M. Inleiding in evidence-based medicine. Klinisch handelen gebaseerd op bewijsmateriaal. Bohn Stafleu van Loghum, Houten 2003.
- Greenhalgh T. How to read a paper. Papers that report drug trials. BMJ 1997; (315): 480–483.
- Yusuf S, Wittes J, Probstfield J, Tyroler H A. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991; (266): 93–98.
- Warner J O. A double-blind, randomized, placebo-controlled trial of cetirizine in preventing the onset of asthma in children with atopic dermatitis: 18 months' treatment and 18 months' posttreatment follow-up. ETAC Study Group. J Allergy Clin Immunol 2001; (108): 929–937.

Correspondence:
Paul L P Brand
Princess Amalia Children's Clinic
Isala klinieken
Zwolle, the Netherlands
Tel. +31 38 424 5050
Fax +31 38 424 2734
p.l.p.brand@isala.nl