Literature research: Aims and design of systematic reviews*

HCW de Vet[1], AP Verhagen[2], I Logghe[3] and RWJG Ostelo[1,4]

[1]The Institute for Research in Extramural Medicine (EMGO Institute), VU University Medical Center, Amsterdam [2] Department of General Practice, Erasmus MC University Medical Center, Rotterdam [3] Avans Hogeschool, Department of Physiotherapy, Breda

[4] Department of 'Allied Health Care Research' of the Amsterdam School of Allied Health Education

In an era in which cost-effective care is high on the political agenda, considerable attention has been paid to the scientific principles of physiotherapy. Every care-provider is expected to be informed about the most effective treatment in his profession. Evidence for effectiveness should preferably be provided by randomised clinical trials. There has been a rapid increase in the number of randomised trials and it is therefore very difficult for care providers to keep up-to-date. Literature reviews, in the form of systematic reviews, or meta-analyses, make it easier for care providers to keep abreast of the knowledge in a specific field. Reviews form the basis of 'evidence-based' medicine. This research note elaborates on the importance of literature research and describes the design and execution of systematic reviews.

There is increasing pressure from politicians and insurance companies to provide scientific evidence of the effectiveness of health care. The professions have therefore been given the task of demonstrating the effectiveness of their treatments. This development is reflected in the emergence of evidence-based medicine.

In the past decade the number of published effect studies has risen dramatically. The Medline database illustrates the rapid increase: in 1964 only 16 new randomised trials were published in the field of medicine and allied health care, in 1982 this had risen to 2038, and in 2000 a further 35 000 new randomised trials were published worldwide. In the database of the Cochrane domain 'Rehabilitation and Related Therapies' there are currently over 2000 randomised trials, many of them focusing on physiotherapy. It is all too clear that practising physiotherapists can never keep up with such a wealth of literature, even if they could obtain all the journals (more than 200) in which these studies are published.

In practice, it appears that professionals are aware of only a very small portion of the published research, and this is what forms the basis of their conclusions regarding treatment. Their conclusions would, perhaps, be different if they had access to (almost) all of the available information, because the outcomes of the various studies on the same subject can vary considerably. This is why literature research plays an increasingly important role in summarising the available knowledge.

Types of literature research

The aim of literature research is to summarise the available knowledge in a specific professional field in order to obtain more precise insight into the effectiveness of certain treatments. Literature research also reveals our knowledge-gaps, and therefore often generates new research questions. Literature research was originally a comprehensive review article—a classic or *narrative review*—in which an expert in a certain professional field published his opinion about the state of affairs on the basis of (a selection of) the literature. How this selection took place, and on what the expert based the opinion, was not always clear. There came a change in the 1980s, when literature research became more systematic. Since then the term 'systematic review' has been used for literature studies in which the literature is systematically searched and assessed, including an evaluation of the methodological quality of the studies included the review. The main characteristics of a systematic review are that a clear description is given of: the method used to search the literature, which data have been extracted from the various articles, how the quality of the studies is assessed, and how the subsequent conclusions have been reached. This allows readers to follow the procedure and to form their own opinion about the quality of the review (de Vet et al 1997).

If not only the methodological quality of the studies is assessed, but a quantitative summary of the results is given, a systematic review is also often referred to as a meta-analysis. In meta-analysis the results of all the studies are summed together. This summing up is called 'pooling'. In this way small but relevant effects can be demonstrated; such effects are not always noticed in the separate studies because the size of the study population is too small.

Design of a systematic review

The design of a systematic review consists of a number of elements: specification of the research question, definition of inclusion and exclusion criteria, conduct of the literature search, final selection of articles included in the review, assessment of the methodological quality of each study, analysis, and the formulation of conclusions. This design applies not only to reviews in which the results of randomised trials are summarised, but also to reviews that summarise the results of observational studies or reviews focussing on the value of a specific diagnostic test. The discussion below applies to reviews of effect studies (randomised trials).

Research question

The aim of a systematic review is to answer a specific question. For instance, a question might be: 'How effective is laser therapy in the treatment of a recently sprained ankle?' The research question can be specified by indicating exactly which population, intervention, and outcome is of interest. An example of a more specific question is: 'Which laser dosage is the most effective in reducing swelling and improving the functional recovery of a recently sprained ankle?'

Inclusion and exclusion criteria

Then the criteria for the inclusion and exclusion of studies are determined. The criteria can be divided into four main groups. First, the type of study design must be defined. A golden rule is that if a sufficient number of randomised trials has been published, then the review should be limited to randomised trials only, because these provide the strongest evidence of the effectiveness of an intervention. If there are only a few, or no randomised trials at all—which is now seldom the case in physiotherapy—other study designs, such as quasi-experiments, can also be included. Second, the criteria regarding the *patient population* (or disorder) and the *intervention* must be defined. For instance, it is necessary to specify if the intention is to study the effect of all types of laser therapy, or only of high dose laser therapy, in all patients or only in chronic patients. Finally, the *outcome measures* that are reported must be determined. If, for instance, the study focuses on the alleviation of pain and 'quality of life' in patients with rheumatism, then studies that report only on laboratory parameters will be excluded.

Search strategy

Once the inclusion and exclusion criteria have been determined, a search is made for the relevant literature. This search must be systematic, because the aim is to actually find *all* the relevant and existing studies. The search strategy is usually quite broad, to allow for subsequent selection of those studies that meet all the requirements of the inclusion criteria. The most practical way is to begin the search strategy in computerised literature databases, such as PEDro, the Cochrane Library, Medline and Embase. The choice of an appropriate set of keywords is crucial. One way to cross-check this is to make sure that all the relevant keywords that are listed in the identified article also appear in the search strategy. Additional search strategies are essential. It is obvious that previous reviews should be sought and that the literature references should be checked. It is also recommended that the authors contact experts in the relevant field.

Selection of articles

When the search has been completed, the inclusion and exclusion criteria must be applied to determine which articles will be included in the systematic review. Sometimes this can be done on the basis of the abstract, and sometimes the whole article must be studied. Because it is often a subjective decision as to whether an article meets the inclusion or exclusion criteria this procedure is often carried out by two reviewers independently.

Blinding

A systematic review is a form of observational research, in which individual studies are the subject of the research. Observational research is prone to bias. Blinding is the best way to prevent bias. It entails blinding the reviewers regarding a number of characteristics of the article. It is often recommended that the reviewers be blinded to the papers' authors, the institutes in which the authors are employed, the journal in which the papers are published, and the sources of funding. Sometimes it is also decided to blind the reviewers to the outcomes of the study. However, this is very time-consuming because it implies that whole sections of the article must be deleted. If a reviewer recognises the authors or the journal, or sees that the outcome of the study is satisfactory, this can, for instance, influence the rating of the methodological quality.

Table 1. Delphi criteria list.

1	Randomisation
	a) Did randomisation take place?
	b) Was the randomisation code unknown to those who included the patients and allocated the treatment?
2	Were the groups comparable at the start of the study with regard to the main prognostic variables?
3	Are the inclusion and exclusion criteria described?
4	Was the effect-assessor blinded?
5	Was the care-provider blinded?
6	Was the patient blinded?
7	Are the point estimates and distribution measures presented for the primary outcome measures?
8	Was the analysis, among other things, carried out according to the 'intention-to-treat' principle*?

*Patients are analysed in the group to which they have been allocated, irrespective of which intervention they eventually received.

Methodological quality

Quality assessment is based on a number of criteria. A criteria list contains questions (items) such as: 'Did randomisation take place?' 'Were the patients in the studies blinded?' 'Were there many drop-outs?' etc. There are many types of criteria list for randomised trials, so a choice must be made. Most criteria lists cover three domains: internal validity (items on randomisation, blinding and drop-outs), external validity (items on the patient population, intervention, effect measures) and precision (items on group size, measurement variation). The Delphi criteria list (Verhagen et al 1998) (Table 1) is an example of a frequently used list. In the Delphi list all items can be answered 'Yes', 'No' or 'Don't know'.

Sometimes the items on the criteria list that are scored with a 'Yes' are added together to form a summary score, or quality score, implying all items have the same weight. Sometimes items are given different weights—a weighted sum-score. Quality scores have the advantage that they are simple and clear. One disadvantage is that a 'No' for one item can be compensated by a 'Yes' for another item. In other words, a study with a high score can still be fatally flawed. It is also possible to assess the items individually. The methodological quality of the studies is usually assessed by two reviewers independently. The results are then compared, and any discrepancies are discussed. Consensus is achieved through discussion or the (final) opinion of a third person.

Analysis

The final step is the analysis and two main questions must be answered. First: 'Will the results of the individual studies be pooled?' This depends on a number of factors. One essential prerequisite in this respect is that the individual studies present at least one point estimate (e.g. average) and a measure of distribution (e.g. confidence interval) for the outcome measure of interest. It should further be taken into consideration whether the individual studies are sufficiently

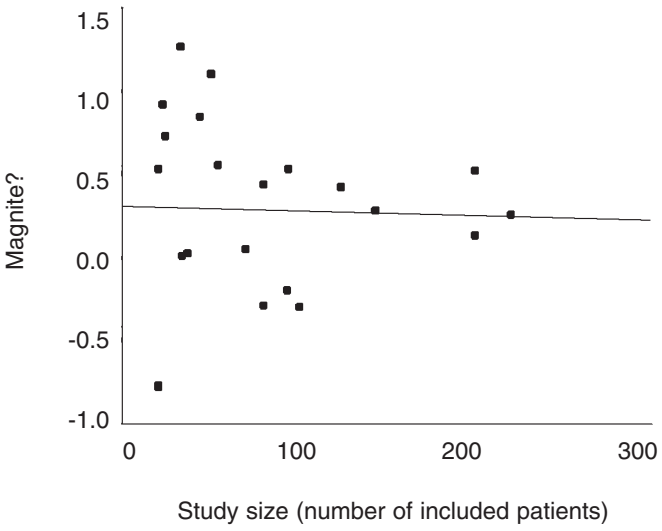

Figure 1. 'Funnel plot' for estimating the possibility of publication bias.

comparable regarding patient population, interventions, and outcome measures. This is mainly a question of comparable working mechanisms and anticipated comparable effects, and not a statistical consideration.

The second question is: 'What role does the quality of the studies play in drawing the final conclusions?' A graphic presentation of the quality scores and the effects that were demonstrated in the studies provides insight into the relationship between the two. If there appears to be no relationship between the quality and the effect size, then it could be decided not to base the final conclusions on quality (items). It is also possible to use quality scores as an inclusion criterion. For instance, it can be decided to review only studies that are randomised and in which the effect measurement was blinded. In the final analysis (pooling), a quality score can also give a certain weight to the results of a study. The final possibility is to divide the individual studies into sub-groups, based on the score for certain items.

Need for a research protocol

It is extremely important to make a protocol in advance, in which the steps and procedures described above are defined. Among other things, this protocol describes who will be responsible for the data collection (e.g. reviewers, experts on the subject), whether the articles will be blinded (e.g. anonymity of authors), how the methodological quality of the studies will be assessed, and how this will be incorporated in the final conclusions.

Problems

The two main problems in a systematic review are heterogeneity and publication bias. Heterogeneity means that the studies are not sufficiently comparable. This question can best be answered on the basis of common sense. The extent of heterogeneity of the outcomes can also be analysed statistically, but there are certain loopholes in the interpretation of this calculation. The stricter the application of the inclusion and exclusion criteria, the less chance there is of heterogeneity, but the risk of ending up with no studies at all is greater.

Publication bias means that bias occurs in the review because not all the existing studies have been published. It is generally accepted that the risk of publication bias is greatest with regard to smaller studies in which no effects or even negative effects have been found. The bias occurs when such studies are not published, or are published in less accessible journals; then the studies included in the review may provide overly optimistic estimates of the effects of intervention. In estimating the possibility of publication bias it is best to use common sense. If a graph (plot) is presented of the relationship between the effect size and the size of the study population, a sufficient number of studies will form a sort of funnel shape (Figure 1). If a large number of dots are missing in the area of the small studies with a negative outcome or no outcome, there is a possibility of publication bias. That is not the case in Figure 1.

Trend or necessity?

It is well known that it takes some time before the results of research are finally incorporated into the knowledge and daily practice of care-providers. Systematic reviews and meta-analyses play an important role in speeding up this process. In 1992 Antman et al investigated how quickly research results were incorporated in textbooks. They studied the treatment and secondary prevention of cardiovascular diseases. The beneficial effect of streptokinase was already known in 1973, but it was not mentioned in textbooks until 1985 as being an adequate method of treatment for cardiovascular diseases. From the Antman study it became apparent that, probably because people were not aware of all the available research, an unnecessary number of studies were carried out in this field of research. For instance, between 1959 and 1985, there were 33 randomised trials on the effect of streptokinase therapy for thrombosis. If a meta-analysis had been carried out after the first eight of these randomised trials (carried out on a total of 2432 randomised patients) had been published, a significant decrease in the number of deaths due to streptokinase would have been found. Meta-analysis of the 25 subsequent randomised trials (involving an additional 34 542 patients) showed no change in effect. All these 34 542 patients participated in trials for no good reason, and half of them failed to receive beneficial or effective treatment. (It must be remarked that in these meta-analyses the methodological quality of the respective studies was not assessed.)

It can be concluded that the results of literature research are important not only for care providers but also for patients, in order to ensure that non-effective treatments are not prescribed for an unnecessary period of time, and that effective treatment is not withheld from patients. Another lesson that can be learned from this story is that a systematic review must *always* be performed before a new effect study is initiated. The function of such a review is not only to investigate whether the research question has already been answered, but it may also have a positive influence on the choice of study population, the intervention and the outcome measures.

References

Antman EM, Lau J, Kupelnick B, Mosteller F and Chalmers TC (1992): A comparison of results of meta-analyses of randomised control trials and recommendations of clinical experts. *JAMA* 268: 240–248.

Verhagen AP, de Vet HCW, de Bie RA, Kessels AGH, Boers M, Bouter LM et al (1998): The Delphi list: A criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus. *Journal of Clinical Epidemiology* 51: 1235–1241.

de Vet HCW, de Bie RA, van der Heijden GJMG, Verhagen AP, Sijpkes P and Knipschild PG (1997): Systematic reviews on the basis of methodological criteria. *Physiotherapy* 83: 284–289.

*This Research Note has previously been published in: Beckerman H and Bouter LM (Eds) (1991): Effectiviteit van fysiotherapie: Een literatuuronderzoek. Maastricht: Rijksuniversiteit Limburg, Vakgroep Epidemiologie.