

Evidence-based massage therapy: a contradiction in terms?

Edzard Ernst

Introduction	11	Inclusion-exclusion criteria	18
Audit	11	Outcome measures	18
Uncontrolled data	12	Systematic reviews	19
Traditional use	12	The 'optimal' trial design	20
Case reports	12	Pragmatic problems	20
Case series	13	Why do research	21
Observational studies	13	Preconditions	21
Controlled clinical trials	14	Background reading	22
Parallel group versus		Define your research question	22
cross-over designs	15	Check the logistics	23
Placebo controlled trials	16	Recruit a 'research team'	23
Blinded versus open studies	17	Obtain funding	23
Randomized versus non-randomized		Conclusion	24
trials	17		

INTRODUCTION

Massage has a long tradition in several medical cultures. In the USA, it is presently experiencing a most remarkable boost in popularity (Eisenberg et al., 1998). Unfortunately, research has significantly fallen behind this development. This chapter is aimed at discussing issues related to research methodology as they pertain to testing the effectiveness of any form of massage therapy. In tackling some of the most common problems, I will take a pragmatic approach. This chapter is not about dry statistical formulae, it is about simple, common sense aimed at novices to medical research.

AUDIT

Practitioners often confuse audit with research and this has caused much confusion in the area of massage therapy. Clinical audit is the systematic evaluation of clinical activity in its broadest sense (Abbot & Ernst, 1997). It involves the identification of a problem and its resolution through various audit cycles. This can involve examination of the structural aspects of the delivery of care, of the processes involved in delivering care, and of the outcomes of care. The essential quality of clinical audit is that it brings about change, and this aspect is generally under-emphasized. The principal concern of clinical audit, and the outcome indicators integral to it,

should be to determine whether treatment, already shown to have a specific effect (efficacy), does so in practice (effectiveness), and whether the resources spent on it are being used to best advantage (efficiency). Thus clinical audit can be usefully applied wherever improvements are to be made in the clinical practice of massage therapy. It is, however, not strictly a research tool, and thus it is excluded from further discussion.

UNCONTROLLED DATA

Traditional use

Massage is amongst the oldest treatment known to mankind (Westhof & Ernst, 1992). Therefore, can anyone doubt that it works? The 'test of time' relies exclusively on experience. While experience is, of course, part of the basis of any clinical medicine, it can be highly deceptive. The history of medicine provides many examples for this to be true. Take blood letting for example; it represented the undisputed panacea for centuries. Its widespread practice must have killed thousands more than it ever benefited (Bauer, 1996). When it was finally discovered to be ineffective, through controlled trials, it was not the intervention but the new (and therefore suspect) method of the controlled trial that was doubted (Lilienfield, 1982). Today we know that blood letting in the form of haemodilution only helps in a few, defined conditions (Ernst et al., 1987).

Traditional use also tells us less about the safety of a therapy than we intuitively assume. But let us assume that a given traditional treatment is *not* burdened with frequent adverse events, which sooner or later make alarm bells ring. It might still be associated with rare or delayed and therefore not immediately obvious yet clinically relevant complications. The 'rule of three' tells us that the number of subjects studied must be three times as high as the frequency of an adverse drug reaction to have a 95% chance that the reaction will actually occur in a studied population (Hanley & Lippman-Hand, 1983). When an adverse drug reaction occurs with a frequency of 1 in 2000, one needs to monitor 6000 users to have a 95% chance that the adverse reaction will be observed at least once. To have a 95% chance that the reaction will occur twice or three times, one has to enroll 9600 and 13,000 patients respectively. The bottom line is that the experience of massage therapists is an unreliable tool to determine either the effectiveness or the safety of their therapy.

Case reports

A clinical research idea often starts with an interesting observation concerning the treatment of a particular patient. A therapist might report: 'I have treated condition X with massage and my patient improved dramatically'.

When put in writing, this initial observation is called a case report (Ernst, 1995). By definition, such case reports are anecdotal evidence; they are essential in clinical medicine as they generate new ideas and constitute experience, but they can never be conclusive. The patient might respond in a different manner or might even have improved without any treatment at all.

Case series

Case series are accumulated case reports evaluated either retrospectively or (more rigorous) prospectively (Ernst, 1998b). They can vary considerably in quality (have better defined inclusion/exclusion criteria, more sensitive endpoints, etc). Case series seem an attractive research tool to many therapists as they do not require informed consent, pose no problem in terms of treatment denial, and fit comfortably into clinical settings. Their most important methodological drawback is the lack of a control group. Thus they have no place in the evaluation of clinical efficacy: their results simply do not tell us whether an observed change was indubitably due to the treatment or to any of the following factors, each of which can influence the clinical results (Ernst, 1998b):

- placebo effect
- natural history of the disease
- regression to the mean
- patient's desire to please the therapist
- therapist's desire for a positive result
- concomitant therapy
- other nonspecific effects.

This, however, is not to say that case series are of no value; the opposite is the case. They are certainly useful, even essential for formulating a hypothesis. In turn, this hypothesis requires testing by other methods, e.g. randomized controlled trials.

Observational studies

Observational studies are very similar to case series. In fact, they are large and well-organized studies without a control group. Because of their size, they may allow comparisons of sub-groups and some inference as to whether or not the observed clinical effect was associated with the therapeutic intervention. For instance, one could conceive a large study of massage therapy where perhaps 1000 consecutive patients with a given condition are treated and the outcome (say pain) is determined. Sub-group analyses could then determine whether patients who were more severely affected or those who received more treatments responded better in terms

of pain relief than the rest of the group. The principal drawback does, however, remain: there is no control group that received a different (or no) therapy. Thus observational studies can hardly answer the question whether the perceived effect was caused by the therapy (specific effect) or some other factor (nonspecific effect) (Pocock & Elbourne, 2000).

CONTROLLED CLINICAL TRIALS

The need for controlled studies to evaluate the effectiveness of a treatment is often misunderstood. The 'effectiveness' observed in uncontrolled studies is really the 'perceived effectiveness', which is composed of the specific therapeutic effect plus other, nonspecific factors (see later). Whenever one wants to be certain about the relative importance of these factors and aims at defining the specific effectiveness of the therapy, one has no choice but to conduct controlled trials and compare the results of an intervention group with those of a carefully chosen control group (Fig. 1.1).

When scientifically investigating whether or not a given therapeutic intervention is effective, one essentially asks whether there is a *causal* relationship between the treatment and the outcome. Some may (rightly) argue that most if not all conditions have more than one cause and that therefore this approach is naïve and simplistic. Even though the multi-causality of disease is an indubitable fact, this argument is wrong. By definition, medical treatments are aimed at providing the cause for the clinical benefit quite regardless of multicausal etiologies — a massage therapist treating low back pain treats the patient under the assumption and with

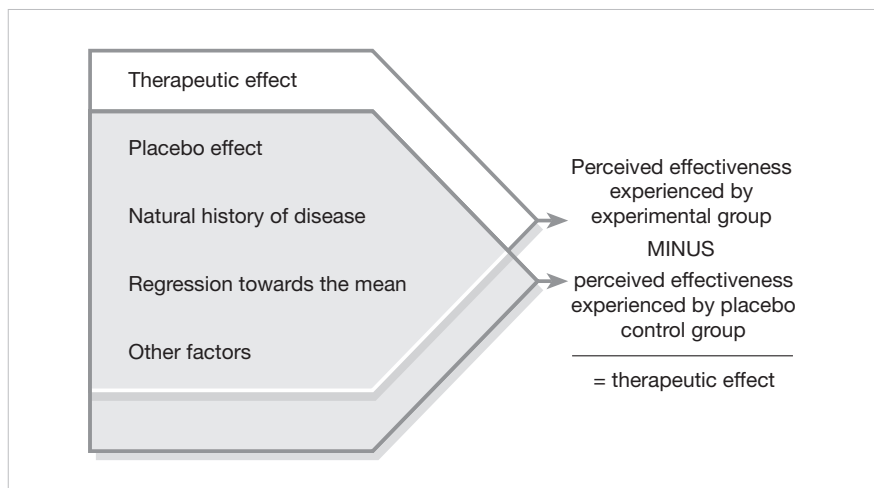


Figure 1.1 Therapeutic effect in relation to other factors determining outcome.

the hope that the massage will ease the pain (which would represent a cause–effect relationship) irrespective of the fact that back pain clearly has many causes. To *not* be interested in the cause–effect relationships in therapeutics means to disregard one of the most essential ingredients in medical therapy (Ernst & Resch, 1996).

Typically, controlled clinical trials are prospective investigations. Yet it is often easier, faster and less expensive to do research retrospectively, for instance, by looking at a number of case notes in an attempt to define which treatment helped best in a given condition. For several reasons this approach is substantially inferior to prospective investigations. There are always several factors that influence the outcome in addition to the treatment given, e.g. the natural history of the disease (Fig. 1.1). Since retrospective investigations are restricted to the data available which, of course, have not been gathered for the purpose of the study, they normally have not been produced under standardized conditions nor do they follow a rigorous predetermined protocol. Inclusion-exclusion criteria (see later) are difficult or impossible to implement on a post-hoc basis because of lack of relevant information, and because randomization (see later) cannot be achieved. Therefore, neither suitability nor validity of the data can be reliably established. Yet, to provide conclusive information on therapeutic effectiveness of a given treatment, all these factors would need accounting for. This can be done reliably *only* with prospective research designs.

Parallel group versus cross-over designs

In trials with parallel groups, participants are split into several (typically two) sub-groups. These receive two different treatments (see later) and the changes that occur in group 1 are compared with those of group 2 (Fig. 1.2). Thus different individuals are compared with each other. This creates numerous confounding factors, and the hope is that, provided both groups are large enough, these will cancel each other out, particularly if the trial was randomized (see later).

In an attempt to reduce confounding, it is tempting to compare one study participant with him/herself. This is the basic concept of cross-over studies (Fig. 1.2). In such trials all participants are treated with two different approaches (e.g. with massage therapy versus drug treatment). To minimize bias, one can randomize the sequence of the two approaches (see later). Essentially the clinical changes in one treatment phase are then compared with those that occur in the other phase.

While cross-over designs have highly attractive features, they are also burdened with numerous problems (Ernst, 1998b). Generally speaking parallel group designs are today considered to be more rigorous.

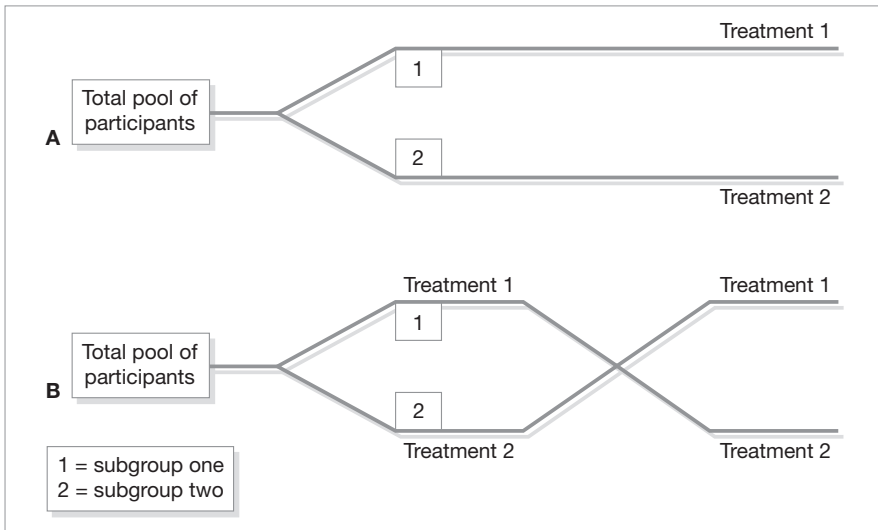


Figure 1.2 Schematic design of parallel group trials (A) and cross-over studies (B).

Placebo controlled trials

The placebo issue is also often misunderstood. No one doubts that the placebo effect can be very powerful indeed (Ernst & Resch, 1994). While in clinical practice we should do everything to make the patient benefit from nonspecific treatment (placebo) effects, we need to exclude them in research aimed at defining specific effectiveness of therapeutic interventions. This is achieved adequately by introducing a parallel group of patients who receive a treatment identical to the treatment under investigation except for the supposed specific treatment effect (i.e. a placebo group). One argument often voiced against this approach is that this neglects the importance of nonspecific treatment effects. This is, of course, not true. The fact that one eliminates a given determinant of a clinical outcome does not mean that one does not appreciate its importance — by eliminating the natural history of the disease in a controlled trial, one by no means disregards its importance. All one attempts is to create a set of circumstances where outcomes and results can be interpreted in a straightforward manner (i.e. 'causality' of the factor under investigation is confounded as little as possible by other factors or circumstances). The trial situation differs critically from the therapeutic situation in this way.

In contrast to what is often said, one can do placebo-controlled trials with *any* form of treatment, even with massage therapy — for instance, one can give sugar pills (placebo) to one group of patients and treat the experimental group with massage therapy. With several therapies (including

massage) it is, however, exceedingly difficult or even impossible to find placebos that are *indistinguishable* from the active treatment for the patient and/or therapist, and only such placebos can be used for patient-blinded studies.

In such situations one is often left with the second-best option to an ideal placebo, i.e. an intervention that mimics the active therapy as closely as possible (but not completely), e.g. superficial massage in a trial of Swedish massage of muscular pain. Admittedly these options represent compromises between the feasible and the desirable. Further features can enhance the credibility of such 'imperfect placebos' — for instance, one can make sure that only patients who have no previous experience with the type of massage under investigation are included in a trial. They are therefore less likely to tell the real thing from the imperfect placebo. The development of a credible placebo crucially depends not only on experience but also on creativity and fantasy.

There may be many situations where other controls are adequate or even superior to placebo controls. For instance, whenever a 'gold standard' (accepted form of therapy for a given condition with proven effectiveness) exists, ethical considerations demand to test a given therapy (e.g. massage) against this 'gold standard'. The research question then would be whether massage is as effective as or superior to the standard treatment.

It is also essential that any control treatment (placebo or other) is comparable in terms of factors relating to the clinical setting: identical environment, same team of caretakers, similar length of patient/therapist contact, similar therapeutic relationship, etc.

Blinded versus open studies

Blinding relates to the fact that the two, three or more parties involved in a clinical trial are masked as to the intervention (i.e. active or control). Blinding the evaluator is usually no problem: the assessor (that is, the investigator who quantifies the results, e.g. pain reduction) does not need to know what type of therapy the patient had been submitted to. Blinding patients in trials of massage therapy is probably not achievable. The same obviously applies to the therapist. In essence this means that in clinical massage research only evaluator-blinded trials are feasible.

Randomized versus non-randomized trials

Randomization is the cornerstone of an unbiased assessment of therapeutic effectiveness. A vivid example of how things can go badly wrong is the Bristol Cancer Study (Bagenal et al., 1990), where the lack of randomization was the main reason for flawed results and the confusion that followed. Randomization means that one sample of patients is divided into

two or more subgroups through pure coincidence. *Only* this method can achieve that both groups are comparable in terms of known *and unknown* potential determinants of outcome (provided the sample is big enough). Non-randomized trials are wide open to bias. This has several reasons. For instance, investigators might intuitively put the more ill patients into that treatment group for which they hope treatment is more effective, or certain other characteristics render a patient more suitable for one of the two forms of treatment tested. This and the fact that one cannot account for factors that are presently unknown, are crucial reasons why only randomization will guarantee that all treatment groups within a study are comparable and that we are prevented from comparing 'apples with pears' (Schulz et al., 1995).

Inclusion-exclusion criteria

'In view of the differing diagnostic criteria on conventional medicine and complementary therapy, it does not appear possible to define a population which can be randomized for a controlled clinical trial of one form of therapy against another...' (Watt, 1988, p. 151). This quote reflects the notorious problem of inclusion-exclusion criteria and emphasizes the different views held by orthodox and complementary therapists. Yet the problem is not insurmountable. Firstly there is no absolute need to insist on strict inclusion-exclusion criteria (i.e. 'define a population'). They are desirable in order to achieve optimally homogeneous patient samples, which in turn, reduces the 'background noise' in the experiment. Yet they are not mandatory — all we face when relaxing these criteria is the need to increase our sample size. Secondly, one can sometimes use orthodox plus unorthodox criteria in sequence. For instance, one could conceive a trial on patients with rheumatoid arthritis diagnosed by an orthodox physician where the patients are subsequently seen by a therapist who defines the suitability of each patient for the massage therapy under investigation. This 'definition' can be based on anything from reproducible variables to personal intuition. Only if a patient passes both 'filters' will he/she be included in the study. Undoubtedly, this would make any study more tedious, yet it would not render it impossible.

Outcome measures

One often gets the impression that medical research has opted to measure what is measurable instead of what is relevant. Proponents of complementary medicine frequently claim that the known criteria to evaluate success or failure of therapy are not meaningful in their field. Actually this is also true for much of mainstream medicine where surrogate endpoints

abound — for instance, blood pressure or serum cholesterol: is it relevant to lower these variables or to prevent a heart attack? The latter is not *a priori* a consequence of the former. What we really want to know is often difficult to measure.

In certain clinical situations encountered by massage therapists there may not be any hard and validated endpoints at all. Yet other meaningful, 'soft' endpoints have been and are being developed — for instance instruments to measure quality of life or well-being (Cella & Tulsky, 1990). Even simple patient preference can be quantified, for instance, in crossover trials. These can be used, depending on the research question, in conjunction with other endpoints like visual analogue scales or 'hard' physiological variables.

SYSTEMATIC REVIEWS

If we accept that the randomized clinical trial is the least biased (yet by no means perfect) method to test for therapeutic effectiveness known today, we still have to admit that one such study is rarely fully convincing. In medical research, one always wants to see independent replications. A single trial could be wrong by chance, through some undetected bias or even through fraud. Where more than one study exists, they often yield different results. For instance, it is conceivable that, for one given indication (say, depression) five studies suggest that massage is effective while five imply that it is not. In such a situation proponents of massage could publish a (apparently evidence-based) review of the positive trials. An opponent could do the same with the negative trials.

This example demonstrates the importance of systematic reviews (and meta-analyses — which are systematic reviews that include statistical pooling of data). Such research projects have to include a detailed explanation where the authors explain what they did and how. They have to demonstrate, for instance, that they included all the data (not just those they liked). This renders a review of this type reproducible and minimizes selection and random biases.

For these reasons, systematic reviews provide, according to the accepted standards of evidence-based medicine (Cook et al., 1997), the most compelling evidence for or against a given therapy (Fig. 1.3). In the realm of massage, several non-systematic (e.g. Callaghan, 1993; Tidius, 1997; Buss et al., 1997) and systematic (e.g. Ernst, 1998a; 1999a,c) reviews have been published.

Systematic reviews are perhaps the best evidence, yet they too are not flawless. Problems can arise when the primary studies are of poor quality (garbage in, garbage out) and when certain (e.g. negative) results never get published (publication bias).

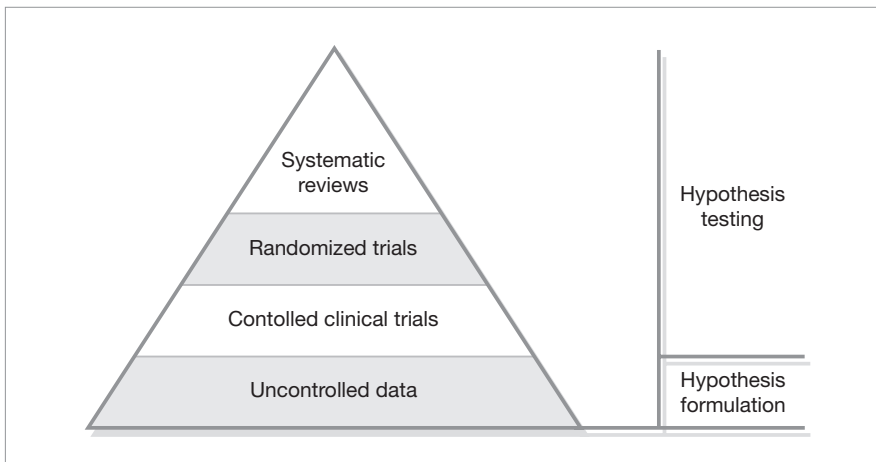


Figure 1.3 Hierarchy of evidence.

THE ‘OPTIMAL’ TRIAL DESIGN

From the discussion so far it follows that there is no such thing as an ‘optimal’ trial design. A study can only be optimal in that it answers the question it set out to answer. All types of investigation discussed above can be optimally matched to a research question. In other words, it is the match not primarily the design one should try to get right. Or, to put it bluntly, there are in principle no faulty designs only bad matches (Fig. 1.3).

If, for instance, one wants to generate or strengthen a hypothesis (which would require testing later), case reports or case series are optimal. If one wants to determine whether massage is more effective than no treatment, a randomized, evaluator-blinded study with two parallel groups — one receiving massage and the other no such therapy — is probably ideal. If one requires to know whether massage is superior to another (e.g. gold standard) treatment, the same design but with a different comparison group would be ideal.

It should be re-emphasized that the entire discussion above is directed towards testing the effectiveness of massage therapy. Obviously there are many other areas of research (Table 1.1). It is clear that for all these areas of research, different methods have to be used and the above discussion does not apply.

PRAGMATIC PROBLEMS

In this last section, I would like to give some practical guidance to those who are new to research and would like to give it a try. Many researchers

Table 1.1 Examples of research question matched with adequate research design

Research questions	Examples of possible design
How prevalent is massage therapy?	Surveys
Who uses massage therapies?	Surveys
What are the main indications?	Surveys
What are reasons for using massage therapy?	Personal interviews, postal questionnaires
Are there adverse effects?	Literature review
How frequent are these adverse effects?	Large scale observational study
Does massage offer value for money?	Cost-benefit, cost-utility studies
Which treatment will help a given patient?	Single case study
Which mechanism brings about a given clinical effect?	Investigations using physiological variables
What expectations do patients have?	Personal interviews/qualitative methods
What experiences do patients report?	Personal interviews/qualitative methods
Does massage offer value for money?	Cost evaluation studies

(including myself) have learnt research ‘the hard way’, e.g. by making all the mistakes themselves. Perhaps the following paragraphs will prevent others from making my mistakes all over again.

Why do research?

There are many reasons to do research, and some are clearly better than others. Enthusiastic novices often want to prove that their therapy works. This is probably one of the worst reasons for doing research. An investigation should *not* set out to *prove* a point but rather to *test* a hypothesis. An investigator with an ‘axe to grind’ is hardly an objective researcher. Clinical research, in particular, must be patient-centered. Unquestionably, the best reason for doing research is the hope of coming one step closer to the truth and to help (future) patients.

Preconditions

Certain items are essential because, without them, there is no use in even attempting research. It is worth remembering that bad research can be unethical (Emanuel et al., 2000). It can mean not only a waste of resources but also the needless suffering of patients.

An adequate knowledge of research methodology and of the subject area under investigation — for example of treatment modality (e.g. the form of massage therapy to be tested) and disease — are absolute prerequisites. To some degree expertise can be ‘bought in’ (see later), but the project leader must have at least a minimal understanding of all the issues involved. If you do not have this expertise, acquire it — or do not embark on research.

It almost amounts to a platitude to state that certain infrastructures are also essential. By this I mean things like the time to carry out the work, access to a library, electronic databases and computers as well as the (prospect of) funds to finance all the work and equipment involved. Before you even start planning a research project it might be a good idea to draw up a simple checklist of all the preconditions required in your particular case and go through it one by one.

Background reading

You may want to embark on a subject, say a study of Swedish massage to treat back pain, and not be fully aware of what has been published on this subject already. Yet it is mandatory that you are! Thus it is highly advisable to conduct an in-depth search for all published articles, read all of them thoroughly and make sure you understand all aspects (if you do not, seek help). Failure to do this background research properly might seriously embarrass you and your colleagues later on. You (or someone else) might, for instance, find out that the study you have just done has already been conducted in a more definitive way by someone else. This would obviously render your work redundant and a waste of time, energy and money.

Define your research question

Using the above example, you may have started out with the idea of studying massage for back pain. Now that you have read the published articles on the subject, you will almost certainly have found that the question you are asking is much more complex than originally anticipated. Do you want to formulate or test a hypothesis with your research? What type of patients do you want to study? What type of back pain? What type of massage do you want to test? How do you want to recruit your patients? Do you need to conduct a controlled trial or an observational study? What should the control treatment be (if any) — a ‘placebo’ or a standard treatment? Can you randomize the treatment groups? Is the treatment under investigation representative for its class? Do you need one therapist or more? What should their qualifications be? Are all conditions optimal for the treatment to work? And so on. Only when you have answered such questions (they will invariably come up when you do your background reading and they will differ according to the nature of your project) will you be able to define the research question. Doing this is essential for deciding which methodology is the best for what you have in mind. It is also a decisive step towards developing a protocol (see ‘Recruit a research team’).

Check the logistics

This preparatory work will have led you to a more concise idea of what may be coming up. Certain things will have become clear to you and you might, at this stage, what to (re)check whether the logistic preconditions for your research project are fulfilled. For instance, do you have access to the type (and adequate numbers) of patients you need to study? How large should your patient sample be? Is it realistic for you to obtain sufficient funding? Is it likely that you can obtain patient consent for what you plan to do? Is the evolving proposal ethical? Do you have the necessary rooms, help (secretarial back up, research nurses), etc? There will almost certainly be other questions to ask. My advice is, again, to draw up a checklist and tackle one problem at a time.

Recruit a ‘research team’

You will probably find that your general research knowledge and experience are not enough to cover all aspects of your project competently. It is therefore usually mandatory to assemble a team for developing a sound protocol of your study and guide you through its experimental phase. Depending on the type of your investigation this team will vary in size and composition. In the example of massage for back pain, it might include a statistician (almost invariably advisable), a clinical expert in back pain (for example, a rheumatologist) and an experienced massage therapist. Make the team as small as possible but as large as necessary.

Within this team you should now organize a series of discussions to evolve a protocol. Subsequently, you might take the lead and draft an outline and circulate it within the team until every team member is satisfied. The team should supervise the entire investigation. Once the protocol is finalized, the planning phase is (almost) finished. All that is needed now is to submit it to the appropriate ethics committee, and secure funding. During this process several (hopefully small) revisions of your protocol may prove necessary.

Obtain funding

Funding is, of course, very often the real obstacle (Ernst, 1999b). Research funds are invariably limited and rejection rates are often high, particularly if you have to compete with applications from mainstream research. Rejections can be extremely disappointing, but you must not be deterred. To succeed you have to try over and over again and learn from the criticisms of those who review your application. Here, too, you should seek expert advice. Establish contact with patient organizations, try all the charities you can think of, use your imagination and leave no stone

unturned. If research in massage therapy is ever to get anywhere, I strongly feel that some dramatic changes to the all too miserable present funding situation have to be brought about.

At present there are few funds especially dedicated to such research. Thus we find ourselves competing with mainstream scientists for a more and more limited amount of money. This means that our applications are judged by panels who usually have little understanding of (or sympathy for) complementary medicine. This in turn results in the undeniable fact that very little money is spent on such research (Ernst, 1999c).

I have said and written it before, and I will carry on doing so: in view of the high popularity of complementary medicine (Eisenberg et al., 1998), it is quite simply unethical not to research the subject systematically — and this, of course, requires adequate research budgets.

CONCLUSION

Massage therapy remains grossly under-researched. In particular, clinical trials need to test the effectiveness of defined types of massage for defined conditions. The methodology for doing this is similar to clinical research in other areas. Existing trials of massage therapy are often burdened with significant limitations (Cawley, 1997). Lack of research expertise and research funds are probably the two main reasons for the paucity of reliable evidence in this area. We should find ways of overcoming these obstacles.

REFERENCES

- Abbot, N. C., & Ernst, E. (1997). Clinical audit, outcomes and complementary medicine. *Forschende Komplementärmedizin*, 4, 229–234.
- Bagenal, F. A., Easton, D. F., Harris, E., & Chilvers, C. E. D. (1990). Survival of patients with breast cancer attending Bristol Cancer Health Care Centre. *Lancet*, 336, 1185–1188.
- Bauer, J. (1996). *Die geschichte der aderlässe*. München: Fritsch.
- Buss, I. C., Halfens, R. J. G., & Abu-Saad, H. H. (1997). The effectiveness of massage in preventing pressure sores, a literature review. *Rehabilitation Nursing*, 22, 229–234.
- Callaghan, M. J. (1993). The role of massage in the management of the athlete, a review. *British Journal of Sports Medicine*, 27, 28–33.
- Cawley, N. (1997). A critique of the methodology of research studies evaluating massage. *European Journal of Cancer Care*, 6, 23–31.
- Cella, D. F., & Tulsky, D. S. (1990). Measuring quality of life today, methodological aspects. *Oncology*, 4, 29–38.
- Cook, D. J., Mulrow, C., & Hayes, R. B. (1997). Systematic reviews, synthesis of best evidence for clinical decisions. *Annals of Internal Medicine*, 126, 376–380.
- Eisenberg, D. M., David, R. B., Ettner, S. L., et al. (1998). Trends in alternative medicine use in the United States, 1990–1997. *Journal of the American Medical Association*, 280, 1569–1575.
- Emanuel, E. J., Wendler, D., & Grady, C. (2000). What makes clinical research ethical? *Journal of the American Medical Association*, 283, 2701–2711.

- Ernst, E. (1995). What is wrong with anecdotal evidence? *European Journal of Physical Medicine and Rehabilitation*, 5, 145–146.
- Ernst, E. (1998a). Does post-exercise massage treatment reduce delayed onset muscle soreness? A systematic review. *British Journal of Sports Medicine*, 32, 212–214.
- Ernst, E. (1998b). Establishing efficacy in chronic stable conditions, are $n = 1$ study designs or case series useful? *Forschende Komplementärmedizin*, 5(suppl 1), 128–138.
- Ernst, E. (1999a). Abdominal massage therapy for chronic constipation, a systematic review of controlled clinical trials. *Forschende Komplementärmedizin*, 6, 149–151.
- Ernst, E. (1999b). Funding research into complementary medicine, the situation in Britain. *Complementary Therapies in Medicine*, 7, 250–253.
- Ernst, E. (1999c). Massage therapy for low back pain, a systematic review. *Journal of Pain and Symptom Management*, 17, 65–69.
- Ernst, E., Matrai, A., & Kollar, L. (1987). Placebo controlled, double-blind study of haemodilution in peripheral arterial disease. *Lancet*, 2, 1449–1451.
- Ernst, E., & Resch, K. L. (1996). Evaluating specific effectiveness of complementary therapies — a position paper — Part one, methodological aspects. *Forschende Komplementärmedizin*, 3, 35–38.
- Ernst, E., & Resch, K. L. (1994). The science and the art of the placebo. *Update*, 619–622.
- Hanley, J. A., & Lippman-Hand, A. (1983). If nothing goes wrong, is everything alright? *Journal of the American Medical Association*, 259, 1743–1745.
- Lilienfeld, A. M. (1982). The evolution of the clinical trial. *Bulletin of Historical Medicine*, 56, 1–18.
- Pocock, S. J., & Elbourne, D. R. (2000). Randomized trials or observational tribulations? *New England Journal of Medicine*, 342, 1907–1909.
- Schulz, K. F., Chalmers, J., Hyes, R. J., & Altman, D. G. (1995). Empirical evidence of bias. *Journal of the American Medical Association*, 273, 408–412.
- Tidius, P. M. (1997). Manual massage and recovery of muscle function following exercise, a literature review. *Journal of Orthopaedic and Sports Physical Therapy*, 25, 107–112.
- Watt, J. (1988). *Talking health* (p. 151). London: Royal Society of Medicine.
- Westhof, S., & Ernst, E. (1992). Geschichte der massage. *Deutsch Medizinische Wochenschrift*, 117, 150–153.