RANDOMISED, CONTROLLED EFFICACY-TRIALS IN LOW-BACK PAIN PATIENTS

A protocol format

FIMM SCIENTIFIC COMMITTEE

Editor: Lars Remvig, MD, DrMedSci

Preface to the 2nd Efficacy Protocol

Based on the continuing debate within the International Federation for Manual/Musculoskeletal Medicine (FIMM) Scientific Committee, it became clear that the first protocol showed certain shortcomings.

This second protocol has been developed for efficacy studies within the area of low-back pain, but it can easily be adopted to other areas of the body.

The FIMM Scientific Committee is aware that developing this kind of protocol is a continuous process. By publishing the 2nd protocol on the website of FIMM, it is our hope that those scientists who use this protocol will send their comments to the Chairman of the Scientific Committee. In this way, we hope to improve the present protocol.

Furthermore the Scientific Committee asks those scientists who receive this protocol to distribute the protocol to their fellow scientists. In this way, the protocol becomes accessible for all practitioners in the field of M/M Medicine.
This protocol is the end product of all the energy of the members of the Scientific Committee.

Dr. Jacob Patijn, MD, PhD, Neurologist, Physician for Manual/Musculoskeletal Medicine
Chairman of the FIMM Scientific Committee

Dr. Lars Remvig, MD, DrMedSci, Rheumatologist
Responsible member for the Efficacy Group of this Committee

International Federation for Manual/Musculoskeletal Medicine

FIMM SCIENTIFIC COMMITTEE

Address for reprints and comments:
Dr. Lars Remvig, MD, DrMedSci, Rheumatologist
Rigshospitalet, University Hospital of Copenhagen, Clinic for Orthopaedic Medicine and Rehabilitation, Blegdamsvej 9,
DK-2100 Copenhagen Ø, Fax: +45 35 45 20 08,
E-mail : remvig@rh.dk

I. INTRODUCTION		**5**
I.1	Background	5
I.2	Argumentation for the planned efficacy trial: “the formulation of the problem”	5
I.3	Specifying the problem formulation	6
II. MATERIALS AND METHODS		**7**
II.1	Inclusion/exclusion criteria	7
II.2	Baseline measurements	7
II.3	Randomisation and stratification	7
II.4	Blinding	7
II.5	Intervention	8
III. DATA SAMPLING		**8**
III.1	Qualitative versus quantitative parameters	8
III.2	Intention to treat principle	9
III.3	Follow-up studies	9
IV. STATISTICS		**9**
IV.1	The “nil-hypothesis”	9
IV.2	Level of significance	9
IV.3	Estimation of sample size	9
IV.4	Confidence limits	10
IV.5	Testing of your nil-hypothesis	11
V. ETHICS		**11**
V.1	Ethical considerations	11
V.2	Informed consent	11
VI. PUBLICATION		**12**
APPENDIX A		**12**

I. Introduction

I.1 Background

Previously published Randomised Controlled Trials (RCT) on Low back pain (LBP) have showed many methodological flaws, which make a definite conclusion of a positive therapeutic effect uncertain. Most RCTs with respect to M/M Medicine are only dealing with the effect of a single therapeutic modality: the manipulation. M/M Medicine in the daily practice of LBP comprises mostly a combination of different therapeutic approaches.

The FIMM Scientific Committee has come to the conclusion that, for the time being, the treatment of LBP in our profession in principle has a multi-modal therapeutic approach. One of the reasons is the fact that the LBP population is a heterogeneous group of patients with respect to pain aetiology, medical history, clinical findings, etc. In consequence, diagnosing and treating patients with LBP is a continuous process, in which – based on changing diagnostic findings – different therapeutic modalities are indicated.

The lack of specific clinical syndromes in M/M Medicine for LBP makes it impossible to perform "fastidious trials", i.e. RCTs in which a single therapeutic modality is tested on a homogeneous population with a specific diagnosis. In the future however, when reproducible and evaluated diagnostic procedures in M/M Medicine are at our disposal, we expect that fastidious trials can be performed.

For the time being, with heterogeneous LBP populations, pragmatic trials are the best format in LBP. The present standardised protocol is designed to make future trials in M/M Medicine for LBP comparable.

I.2 Argumentation for the planned efficacy trial: "the formulation of the problem"

The idea to perform a scientific trial is mostly based on an observation in daily practice. This phase of a protocol development is called "the formulation of the problem". This is a very important phase because it influences all aspects of the design of the trial. For clarification we can take an example:

You have observed in your daily practice that LBP patients have better therapeutic results when you treat them with a combination of manipulation and injection therapy. The question is whether your

observation is true. Before developing a protocol you have to study the literature in order to see whether someone else has already noticed this problem and whether RCTs have already been published. If not, "the formulation of the problem" can take place.

I.3 Specifying the problem formulation

Having formulated the problem, one firstly has to decide which questions with respect to the problem the trial is supposed to answer.

As an example the formulation could be: "Is a combination of manipulation and injection therapy more effective in the treatment of LBP than manipulation alone?"

Besides the main research question one or more secondary questions may be formulated, for example:
1. Is there a difference in treatment effect between male and female?
2. Is there a difference in treatment effect between acute, sub-acute and chronic LBP patients?
In order to answer these supplementary questions, you must however increase the size of the study population and consequently worsen the logistic feasibility of the trial.

Secondly the scientist must define the different components of the problem formulation. In our example, phrases such as "manipulation", "injection therapy", "LBP", "combination" etc. have to be well defined.

Does LBP include pain radiation to the extremities? With or without objective neurological signs of radicular affection? Acute, subacute and/or chronic?

Is manipulation with or without impulse? Does it include Muscle Energy Technique and/or Myofascial Release? Is every modality acceptable?

Does injections therapy include local anaesthetics and/or corticoid steroids? Does it include prolotherapy? Are the injections given intramuscularly and/or in ligaments? How large is the volume? How often are injections given?

What is meant by combination? Is it a must to combine the treatment modalities or is it optional? Must it be given in a specific sequence?

Finally what do you mean by the phrase "effect"? Is it pain relief? Increase in mobility? General well-being increased? Disability? Cheaper? This is extremely important with respect to the choice of outcome measurements.

In conclusion: It is essential that commonly used phrases of our profession are well defined in the protocol and in the final publication.

II. Materials and methods

II.1 Inclusion/exclusion criteria

According to the above mentioned expression of the problem the test population, as well as the ones excluded, can easily be described.

How the patients are recruited – after public announcement, referred by general practitioners, found in the outpatient clinic etc. – must be described, and it must be emphasised that the patients are enrolled consecutively after informed consent (see below).

II.2 Baseline measurements

After the selection and inclusion of patients in the trial and before the randomisation procedure takes place, the baseline measurement is scheduled. This baseline measurement usually includes a set of demographic variables (age, gender, etc.) and a set of clinical variables (symptom duration, clinical or para-clinical diagnostic tests results). Also the baseline status of the primary and secondary outcomes are measured (pain severity, functional status, quality of life, etc.).

II.3 Randomisation and stratification

Having fulfilled the inclusion/exclusion criteria the test person is enrolled, and not until then can the randomisation procedure take place.

If the population is well defined and very homogeneous, a simple randomisation by chance – even/odd numbers, drawing of envelopes etc. – can be used.

In case of multicenter trials a block randomisation is preferable. (In block randomisation you split the population in blocks, say 6 persons per block. Having picked treatment A three times, further allocation to treatment A within this block of 6 is cancelled, etc. For further information consult statistical textbooks.)

However, in case of a heterogeneous population with respect to various characteristics – say sex, age, diagnosis, duration of disease etc. –, and where these characteristics are supposed to influence the treatment effect, a stratification ("minimisation procedure") should be used. With small populations (i.e. approximately 40 participants) this is particularly valid.

Hereby one can increase the possibility to obtain comparable test populations, with respect to the characteristics, in the treatment and in the control groups.

II.4 Blinding

Double blinding or double blinding with open therapeutical guidance (triple blinding) should be used whenever possible. (Example:

If evaluating the effect of traction versus sham traction on patients with lumbar herniation, the patient and the doctor are blinded with respect to traction force but not to the effect of the traction, while the physiotherapist knows the traction force to be given but not the effect.)

However, treatment modalities for musculoskeletal diseases cannot always be blinded neither to the patient nor to the therapist (massage, manipulation etc.). Thus a third person, a blinded ob-server, is necessary to register the effect in order to avoid bias.

II.5 Intervention

According to the above-mentioned expression of the problem, the trial could be 100 % pragmatic or there might be some restrictions in number or type of intervention modalities, in the sequence of the interventions or total number of interventions.

In any case, a precise description of the various treatments used must be given, as well as a description of how often and for how long a period they are – or could be – used. If not, other scientists will not be able to repeat the trial or to copy the treatments in their daily practice.

Which kind of interventions are accepted for controls must be stated, and if there are any restrictions in normal physical activities for both the controls and the actively treated patients, it must be mentioned.

Co-intervention should be avoided in the design of the trial, or at least co-intervention should be equally distributed in the experimental and the control groups.

III. Data sampling

III.1 Qualitative versus quantitative parameters

As mentioned above, it is a must to specify what is meant by effect and furthermore to specify the parameter, qualitative or quantitative, which should be used as the "core parameter" for measuring the result of the intervention(s).

The same parameter should of course be used for estimation of sample size.

Generally quantitative parameters are preferable. Thus, qualitative parameters should be "translated" into quantitative parameters, i.e. pain by a visual analogue scale, impairment by a Health Assessment Questionnaire, etc.

However, also the doctor's as well as the patients' general assessment should be included in a data sampling.

Subjective and objective data could be united in a low-back pain rating scale, which should be validated and internationally accepted (e.g. Manniche Low-Back Pain Rating Scale, validated in "Pain", 1994; 57(3): 317–26).

An international group of back pain researchers published a proposal for a standardised use of outcome measures which can be considered when choosing the appropriate outcome measure and instruments (Deyo RA et al., Spine 1998; 23: 2003–2013).

III.2 Intention to treat principle

It is mandatory to use the "Intension to treat principle".
Consequently, when a patient fulfills the inclusion criteria, a precise registration of all patient data and all trial-parameters must be performed, also if the patient suddenly is withdrawn – no matter the reason.

III.3 Follow-up studies

Besides the determination of the short term effects (during and just after the invention period), insight in the long-term effects is usually also desirable. Depending on the expected effect 6, 12 and/or 24 months follow-up studies may be chosen, registering the same "core parameters" as in the primary study.

IV. Statistics

IV.1 The "nil-hypothesis"

One should be aware that the different results of +/– intervention could be due to coincidence within one or within several other populations or between populations. Experimentally it would be extremely difficult to decide this, and therefore a remedy, the nil-hypothesis, is introduced.
This hypothesis says that there is really no effect of the tested intervention, i.e. that the two test spots (+/– intervention respectively) are subpopulations of the same population and that the difference registered is due to stochastic variation.
Then it is possible to estimate the possibility that a difference at least as large as the registered one is due to stochastic variation.

Based on the previously mentioned problem formulation – "Is a combination of manipulation and injection therapy more effective in the treatment of LBP than manipulation alone?" – the nil-hypothesis would be: "The effect of a combination of manipulation and injection therapy, in the treatment of low back pain, does not differ from the effect of manipulation alone".

One can never prove that the nil-hypothesis is correct, neither that it is wrong. One can only demonstrate that coincidences are in such and such a probable explanation for the registered difference, in case the nil-hypothesis is correct.
If one rejects a *de facto* correct nil-hypothesis, one commits a type-1 error.
If one accepts a *de facto* incorrect nil-hypothesis, one commits a type-2 error.

IV.2 Level of significance

In general, incl. musculoskeletal diseases, it is accepted to use a type-1 error (2 alpha) = 0.05 and a type-2 error (beta) = 0.10.

IV.3 Estimation of sample size

In clinical trials one is not only interested in significant differences, but one is in particular interested in significant *relevant* differences

between two test populations, i.e. two populations with separate treatments.
It is up to the scientist (but preferably based on previous studies and/or broad consensus within the scientific and clinical community) to judge which difference he/she finds relevant with respect to the severity of the disease, the treatment possibilities and their side effects, and finally the resources available and to make an estimation of the expected treatment effects.
If he/she has also decided the risks of type-1 and type-2 errors, estimated the Standard Deviation (SD) of the results and chosen the statistical test, he/she can "count backwards" to the number of test persons.

The Minimal Relevant Difference (MIREDIF) and the SD should be estimated by use of the results from previous trials or by a "qualified" estimate. In case of quantitative data and supposing that $SD_1 = SD_2$ and $n_1 = n_2$ the total number of participants (N) can be estimated by use of a Student's unpaired t-test, simplified to the following equation:

$$N = (t_{2a} + t_b) \times 4 \times SD^2/MIREDIF^2$$

In order to make a quick estimation, a diagram has been constructed (Appendix A).

Even though this calculation of sample size is based on an estimation of SD and MIREDIF (and on a parametric test even when using a non-parametric test for the analysis of the results), it is much better than a simple guess.
The reason is that if the trial shows a significant difference in effect between +/– intervention, one knows that the level of type-1 error is below the acceptable limits.
And, on the contrary, if there is no significance, the nil-hypothesis cannot be rejected, and MIREDIF tells you how large the difference cannot be.

IV.4 Confidence limits

The result of the intervention is of course only a result representing the small population studied in the actual trial.
For the clinician it is interesting to know the limits within which he/she can expect to find the "true" result with a certain possibility, the confidence limits.
If the observations from the study are not normally distributed and the distribution unknown, it would be misleading to use Mean +/–2 SEM to illustrate the confidence limits.
Instead the Median and confidence limits of the Median should be used, when dealing with biological materials.
Conventionally one uses the 95% confidence limits.

IV.5 Testing of your nil-hypothesis

The preferable methods for analysing the effect(s) of an intervention depend particularly on the test material and the problem presentation.

As mentioned above, the variable is preferably presented on a rating scale.

When comparing the results of the intervention group with the control group, one deals with an unpaired analysis. However, paired analyses are also needed when comparing results within the group (pre- versus post-intervention).

In small trials analysing effect(s) on biological materials, non-parametric analyses are preferable, and they are also useful when materials are larger.

Consequently, the following non-parametric methods are in general advisable in order to test the level of significance:

Comparison of two groups:

1. Unpaired analysis, variables registered on a rating scale: *Mann-Whitney's test (Wilcoxon test for two samples).*
2. Paired analysis, variables registered on a rating scale: *Wilcoxon test (Wilcoxon test for pair differences).*

Comparison of more than two groups:

1. Unpaired analysis, variables registered on a rating scale: *Kruskal-Wallis test.*
2. Paired analysis, variables registered on a rating scale: *Friedman test.*
3. Paired and unpaired analysis, variables registered on a ratio/interval scale: Analysis of variance *(ANOVAR).*

Correlation analysis:

1. *Spearman's rho.*

V. Ethics

V.1 Ethical considerations

One always has to consider whether the planned trial is ethically acceptable. In particular the risks confined to treatment or no treatment must be considered and considered with respect to the severity of the disease.

In most countries Ethical Committees shall accept the Trial Protocol, including the ethical considerations and the written information to the patients, before starting the trial.

The demands of the Helsinki Declaration can almost be united in two words: "Informed consent".

V.2 Informed consent

It is essential that the patients, before entering a trial, are informed of the purpose of the trial, the distribution by chance to the two (or more) experimental groups and the potential risks of treatment versus no treatment. Information must be given in a non-professional language understandable for the patients.

The treatment that is offered in case the patient does not want to participate or the patient dropout of the trial must be described.

The information should be written as well as oral.

VI. Publication

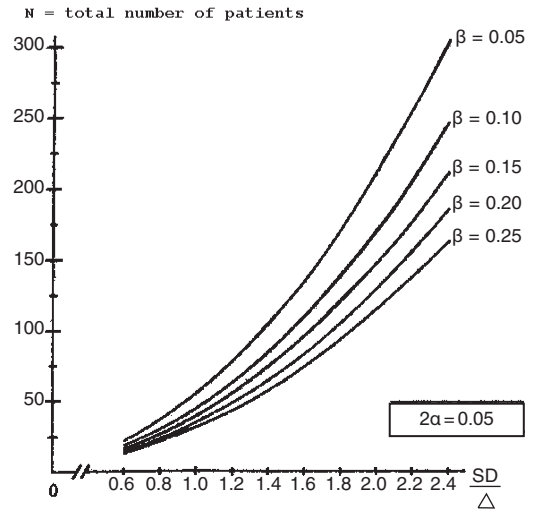
When writing an article, one must of course follow the instructions given by the journal. However a disposition as the following is generally accepted:

Title, possible Subtitle, Authors, Abstract or Résumé, Index words, Introduction with formulation of the problem, Materials and methods, Statistics, Ethical considerations, Results, Discussion, References written as the journal demands and finally Acknowledg-ments.

In the protocol one should mention the two most relevant journals for publication of the results, and it is advisable to determine the order of authors or at least rules for the order.

No paper should be forwarded to an Editor without acceptance from all the participating scientists, and the head of an institution should also be given a possibility to review the paper before it is forwarded.

Appendix A



Diagam for estimation of sample size in unpaired group comparison trials.

The figure illustrates the relation between sample size (N), minimal relevant difference (Δ), standard deviation (SD), type-2 error (β), and two-sided level of significance = 5% (2α).