Prevalence of methodologic errors in rehabilitation research

Lex M. Bouter

Introduction

This paper will consist of two related parts. The first deals with a project in which we reviewed the available literature on the efficacy of physiotherapy systematically. After explaining what we did, I will focus on the most prevalent methodological errors we identified. The reason for this is that these errors, and especially options to prevent them, also seem to be relevant for research in rehabilitation medicine. Recently, this opinion has been confirmed for the field of stroke rehabilitation. The second part of this paper will deal with the question whether single-case studies might be an alternative for large randomized clinical trials which are often so difficult to realize. My answer will be negative, but I will list purposes for which single-case studies are very suitable, and identify ways in which this attractive study design can contribute to research in rehabilitation medicine.

Meta-analyses

In order to gain insight into the efficacy of the most prevalent forms of physiotherapy for the most common indications, we summarized the literature in 10 meta-analyses. I will focus on the strategy we used in reviewing the literature and the common methodological flaws we identified in the studies. From the literature we identified both explanatory and pragmatic randomized clinical trials dealing with indications and interventions relevant to physiotherapy. Outcome measures should include pain, mobility, functional capacity or activities of daily living. Up to this moment about 200 RCTs have been included in one of the 10 meta-analyses.

Starting from an explicit question regarding efficacy, we initially search bibliographical databases, such as Medline and Embase. In our experience, this typically identifies 40-45% of the eligible studies. Subsequently, this is supplemented by screening of non-indexed journals or proceedings, and by citation-tracking. Finally, we ask experts and authors in the field whether they consider our list to be complete. Although randomized clinical trials offer the best chance for a valid study of efficacy, even RCTs can be seriously biased. Therefore, in reviewing we used a predefined set of methodological criteria and corresponding weights. These criteria were operationalized explicitly for every meta-analysis and were applied independently by 2 or 3 reviewers. These were blinded for the authors, the journal and the outcomes of the study. Typically, there was 70 to 80% initial agreement on item level. Differences were resolved by discussion, and for every trial a methodological score was calculated on a scale of 0 to 100. In the resulting article only the outcomes of the best studies were discussed. Sometimes for each study the difference in success rate between the groups and the corresponding confidence interval was calculated. Due to the fact that populations, interventions and outcomes always differed substantially over the studies, we never decided to pool the data statistically.

Abstract

The methodological problems arising in randomized clinical trials (RCTs) on the efficacy of interventions in rehabilitation medicine are probably very similar to those encountered in physiotherapy research. Therefore, the methodological errors prevalent in physiotherapy trials will be discussed and options for prevention will be listed. Special attention is given to the potential role of single-case designs in treatment evaluation. From ten criterion-based meta-analyses on the efficacy of physiotherapy published recently by our group, the most prevalent methodological errors were identified. The methodological quality of about 200 RCTs in total was typically low, and rated somewhere between 30 and 40 points on a scale from 0 to 100. This finding forms a clear contra-indication to drawing strong conclusions concerning the efficacy of physiotherapy. Prevalent flaws were (1) the use of heterogeneous groups of patients, (2) incomplete description of the interventions used, (3) lack of suitable outcome measures, and (4) much too small sample sizes. The methodological errors identified seem, indeed, also to hold for RCTs in rehabilitation medicine and can be avoided by applying general methodological solutions. Although their popularity is still increasing, single-case designs can offer no alternative to RCTs.

Keywords: Rehabilitation Medicine; Research; Reproducibility of Results.
Table 1. Meta-analyses of randomized clinical trials on the efficacy of physical therapy for musculoskeletal disorders.

<table>
<thead>
<tr>
<th>Indication</th>
<th>Intervention</th>
<th>Number of RCTs</th>
<th>Median score (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Back pain</td>
<td>Spinal manipulation (6)</td>
<td>30</td>
<td>35 (20-56)</td>
</tr>
<tr>
<td></td>
<td>Exercise therapy (7)</td>
<td>16</td>
<td>40 (24-61)</td>
</tr>
<tr>
<td></td>
<td>Traction (8)</td>
<td>17</td>
<td>36 (23-68)</td>
</tr>
<tr>
<td></td>
<td>Back schools (9)</td>
<td>16</td>
<td>36 (16-70)</td>
</tr>
<tr>
<td>Neck pain</td>
<td>Spinal manipulation (6)</td>
<td>5</td>
<td>39 (26-50)</td>
</tr>
<tr>
<td></td>
<td>Traction (8)</td>
<td>3</td>
<td>39 (36-51)</td>
</tr>
<tr>
<td>Shoulders complaints</td>
<td>Physiotherapy (10)</td>
<td>18</td>
<td>49 (22-76)</td>
</tr>
<tr>
<td>Knee disorders</td>
<td>Physiotherapy (5)</td>
<td>63</td>
<td>29 (6-52)</td>
</tr>
<tr>
<td>Musculoskeletal disorders</td>
<td>Ultrasound therapy (11)</td>
<td>16</td>
<td>41 (17-70)</td>
</tr>
<tr>
<td></td>
<td>Laser therapy (12)</td>
<td>33</td>
<td>40 (4-72)</td>
</tr>
</tbody>
</table>

*Reference added between brackets

In Table 1 an overview of the 10 meta-analyses is presented. Although we were surprised by the large number of RCTs available in the literature, we were generally disappointed by the methodological quality of the studies. While the best studies included in a meta-analysis usually scored something between 60 and 70 points, the median scores were typically low and in the range between 30 and 40 points. Consequently, we hesitate to draw strong conclusions regarding efficacy in the resulting meta-analyses and tend to concentrate on the most prevalent methodological flaws.

Methodological flaws

A common problem is inclusion of a group of patients which is too heterogeneous, often on the basis of a vague diagnosis like lumbago or frozen shoulder. In such a population the susceptibility for the intervention at issue may vary substantially. Dilution of efficacy will be the consequence of this. More attention should be paid to diagnosis, and especially prognosis, to enable inclusion of homogeneous groups in a trial.

A related problem is posed by the often incomplete description of the interventions used. This makes implementation of positive results unfeasible. Furthermore, we got the impression that the interventions were often clearly sub-optimal. For instance, it turned out that the dosages used in many laser trials were so low that a specific effect became very improbable. Of course, the solution to this problem would be the use of more explicit and optimal treatment protocols.

The clinical relevance of the outcome parameters is often doubtful, for example sophisticated measures concerning range of motion or muscle strength in low back pain. Also the validity and precision of outcome measurement is typically unknown and probably not very impressive. We fear that a lot of the outcome parameters may be insensitive to a clinically meaningful change over time. Consequently, real treatment effects may be missed. It seems urgent to give more attention to the design of outcome measures in physiotherapy trials.

This also seems to hold for rehabilitation medicine. In my opinion, the popularity of outcome measures dealing with impairments has two reasons. The first is a general reluctance to rely on 'soft' subjective data, which often leads to a 'hard' objective and precise measurement of the wrong phenomenon. The second reason seems to be confusion between the questions whether an intervention is effective and how it works. Once clinical efficacy is established on the disability level, data on the corresponding changes in impairments, of course, can offer an insight in the mechanism involved, but can never be a substitute for it.

Before concluding from a negative finding that there is no difference in effect, one should always have a look at the statistical power of the trial. Our meta-analyses show clearly that sample sizes in physiotherapy trials are often very small. The groups included in the trial typically consist of 20 patients or less. Consequently, the chance of making a Type II error is substantial. In other words: a fairly large proportion of the negative findings may be 'false negatives'. The solution to this is obvious: enlarging sample sizes to, say, more than 50 patients per group.

But it is not only a matter of sample size determination, but also of recruitment. The formulas can be found in any textbook on biostatistics and the main decision to be made is about the magnitude of the difference in efficacy you would like to be able to detect. Once you have decided upon the minimal difference in effect you consider to be clinically relevant, there is no point in aiming at making smaller differences statistically significant by increasing the sample size further. The real problem is getting the number of patients you need, which will often make a multi-centre design obligatory.

In summary, it can be said that our meta-analyses have not convincingly shown efficacy for most indications and interventions. On the other hand, because of the high prevalence of methodological flaws, it cannot be concluded that physiotherapy has no effect either. This conclusion is very similar to the one in the review on stroke rehabilitation I mentioned in my introduction.

Single-case design

Although not impossible, it is certainly not easy to overcome the methodological problems identified. This brings me to the question about how to prevent choosing
the wrong study design. As a partial answer, I will try to explain why a single-case design offers, in my view, no solution.* Single-case designs focus on the efficacy of interventions for an individual patient. This is very attractive, because it offers a straightforward answer to the most important question for clinical practice. In contrast, RCTs are often not available, and if they are, they only provide data on mean efficacy among rather heterogeneous groups of patients. From the wide range of single-case designs available, I will focus on the ‘N-of-1 trial’ which is, in fact, a randomized multiple cross-over study in a single subject. This design offers an optimal chance to separate specific treatment effects from natural history, placebo effects and systematic measurement errors. Randomly allocated episodes of index, reference or placebo treatment are separated by wash-out periods.

For a ‘N-of-1 trial’ to be feasible, a number of conditions ought to be met. Firstly, there must be real doubts about treatment efficacy and the problem must be serious enough to warrant all the extra work. Secondly, the status of the patient must be reasonably stable, and the expected effects of treatment must appear relatively quickly and also disappear again quickly when treatment is stopped. The latter is necessary for the avoidance of so-called carry-over effects between different treatment episodes. This obviously will often constitute a problem which can sometimes be avoided by choosing a multiple baseline design in which, for instance, treatment of different joints is started at randomly chosen different moments in time. The third condition for a ‘N-of-1 trial’ is obtaining informed consent from the patient at issue and reaching agreement on the primary outcome phenomenon, number of pairs of episodes to be randomized and rules for stopping the treatment early.

The potential validity problems in ‘N-of-1 trials’ are similar to those in RCTs, and deal with comparability of prognostic factors at the start of treatment episodes, comparability of the non-specific elements of the interventions compared, and comparability of outcome measurements between treatment episodes. Visual inspection of the results summarized in graphs is probably enough to enable a decision to be taken, although several elegant methods for statistical analysis are available.

So far so good. Real problems arise, in my view, when the results of ‘N-of-1 trials’ are extrapolated to conclusions about treatment efficacy in general. In RCTs you know which patients were eligible, and thus for which group of patients they provide an insight in average efficacy. In ‘N-of-1 trials’ it is often unclear how the patient was selected and it is not known which individual characteristics are relevant for extrapolation of the result to other patients. When a number of ‘N-of-1 trials’ with an identical design are available, of course a secondary analysis of the pooled data might be feasible. In fact, the evidence is then treated as a regular multi-centre cross-over trial. Similarly, different ‘N-of-1 trials’ dealing with the same research question can be reviewed and statistically pooled in a meta-analysis. Of course, in reviewing publications of ‘N-of-1 trials’ publication bias will always be a major problem.

While isolated ‘N-of-1 trials’ cannot substitute RCTs involving groups of patients, they can play an important role in the preparation of the necessary multi-centre studies. ‘N-of-1 trials’ can be very helpful in identifying productive and specific hypotheses, the optimization of treatment parameters, and the choice of relevant moments for follow-up. Even more importantly, ‘N-of-1 trials’ can help enormously in identifying outcome measures which are responsive to treatment effects and categories of patients for whom treatment might be really effective.

**Conclusion**

In conclusion, I have tried to show what might be the most prevalent errors in research in the field of rehabilitation medicine. These turned out not to be unique for this specific field, and I have suggested that they can, at least in principle, be avoided by applying general methodological solutions. I have also tried to show that ‘N-of-1 trials’, although they are increasingly popular, do not offer any solution for the problems of RCTs. Instead, ‘N-of-1 trials’ may offer a clear answer for dilemmas in the choice of treatment for individual patients and can also be very helpful in designing valid and informative RCTs.

**References**


