Validity and comparability of studies on the effects of back schools

Jolanda F.E.M. Keijsers, Lex M. Bouter and Ree M. Meertens

In this paper, studies on the efficacy of back schools are judged on their methodological quality (internal validity) and mutual comparability (external validity). All available randomised trials \((n=8)\) were selected for this purpose. The studies were judged on the following criteria: duration and content of the back school programme, selection criteria and number of patients involved, interventions and measures of effect used. With regard to the internal validity of the studies, a number of gaps were found. The numbers of patients involved in the studies were relatively low, no placebo-attention group was used and the principal target of the back schools was not evaluated directly. Because of these gaps, it is as yet not possible to formulate a strong and valid judgement on the efficacy of the back schools studied in the trials. In addition to this, the comparability of the studies is rather poor. Taken together, the available studies do not suggest that back schools are particularly effective.

INTRODUCTION

Low back pain (LBP) is a frequent health problem. In most cases, the complaints are self-limiting, and recovery occurs within 2–3 months in 90% of cases (Bergquist-Ullman and Larsson, 1977). The recurrence of LBP is frequent (40–60%) (Haanen, 1984). The complaints are in most cases non-specific, which means that no underlying pathology has been established. In the Netherlands, 8% of all consultations with general practitioners involve such non-specific back complaints (Hoekstra, 1985). It is commonly recognised that, in addition to the possible physical factors, psychosocial factors are also involved in both the aetiology and prognosis of LBP (Meilman, 1984; Nachemson, 1979; Turk and Flor, 1984). With regard to the duration of the complaint, LBP is usually subdivided into acute (0–6 weeks), sub-acute (6–12 weeks) and chronic (>12 weeks) stages (Frymoyer, 1988). The majority of patients with acute symptoms are commonly managed with bed rest, analgesic medications, exercise, traction and education. However, little evidence was found to support the efficacy of these techniques when treating LBP (Spitzer et al, 1987). In 10% of patients, LBP persists for more than 6 weeks. In this sub-acute phase, 5–10% of the patients require a disc hernia operation. With the exception of epidural steroids and braces, non-surgical treatments are similar to treatments given in the acute phase, although the analgesic medication is more specified. Only 5% of patients presenting with LBP still have symptoms after 3 months (chronic
Most patients with chronic LBP benefit to some extent from anti-inflammatory medications and anti-depressants (Frymoyer, 1988). Exercise programmes and educational programmes are popular additional treatments during this phase of LBP (Frymoyer, 1988).

A promising educational programme for back complaints is the ‘back school,’ which is of Swedish origin (Zachrisson Forsell, 1981). Back schools offer an educational and skills programme in a group setting that consists of a number of lessons (1-16) (Linton and Kamwendo, 1987). Back schools are aimed primarily at the management of pain: information is given about effective ways of dealing with pain, so that the patient will be able to control the pain better. Ultimately, this is supposed to lead to decreases in work absenteeism and medical consumption. The purpose of Linton and Kamwendo’s (1987) paper was to review and evaluate the empirical literature concerning back schools. They show that most studies that have been conducted lacked control conditions. Their conclusion is that the efficacy of back schools is controversial (Fisk et al, 1983; Linton and Kamwendo, 1987).

In this article, we suggest that this is due to the mediocre methodological quality (internal validity) and the limited comparability (external validity) of the available studies. For this purpose, those published studies into the efficacy of back schools that involved randomly assigned control groups (randomised trials) will be discussed here (Bergquist-Ullman and Larsson, 1977; Berwick et al, 1989; Hurri, 1989; Keijser et al, 1989; Klaber Moffett et al, 1986; Lankhorst et al, 1983; Postacchini et al, 1988; Stankovic and Johnell, 1990). It is generally accepted that the randomised trial is the most appropriate way in which to obtain a valid assessment of treatment efficacy (Pocock, 1983). Furthermore, this review deals with non-clinical back schools in a group setting only. Individualised back schools, such as reported by Lindequist et al (1984), and clinical rehabilitation programmes, such as discussed by Aberg (1984), are left out for the sake of homogeneity.

The first section of this paper presents the methodological criteria by which we judged the studies. Then, the randomised trials are discussed critically with regard to these criteria (internal validity), and it is argued to which extent studies on the efficacy of back schools are comparable (external validity). The paper ends with a discussion of the claims on the effectiveness of back schools. Furthermore, some recommendations for future research on the efficacy of back schools are formulated.

**CRITERIA**

There is a reasonable degree of consensus on the methodological requirements of randomised trials (Pocock, 1983). For a complete study of the efficacy of back schools to be made, it is not only the internal validity of the individual studies which is important, but also the comparability of such studies (external validity). Questions relating to the latter include whether the interventions, patients and effect parameters of the individual studies are sufficiently similar. This does not mean that the studies need to be completely identical. If differences between the studies exist, it should be argued to what extent they will have an important prognostic influence on the effect of the intervention. In the present study, research on the efficacy of back schools was assessed using the following criteria:

1. **Duration and content of the back school programme:** How many lessons does the back school programme consist of, and what information and/or training is given?

2. **Selection criteria and numbers of patients involved:** Which patients are included and which patients excluded, and how many patients are recruited for the study?

3. **Interventions and execution of the study:** Which interventions are compared and is there any long-term follow-up?

4. **Effect parameters:** Which effect parameters are used to assess the efficacy of the back school?

**DURATION AND CONTENT OF BACK SCHOOL PROGRAMMES**

The back school programme developed by Zach-
Zachrisson Forssell (1981) consists of four mainly audio-visual lessons. In these lessons, the physical factors influencing back complaints are made clear. Little time is available for exercises (Zachrisson Forssell, 1980). Table 1 summarises a number of randomised trials based upon similarities in content and duration. It shows that the programmes of the back schools in several studies differ substantially.

The articles by Bergquist-Ullman and Larsson (1977), Lankhorst et al (1983) and Hurri (1989) deal with different versions of the Swedish back school. With respect to the duration of these back schools, it appears that that presented by Hurri (1989) differed substantially from the original Swedish back school designed by Zachrisson Forssell (1981). The back school reported on by Klaber Moffett et al (1986) was characterised by a mainly physical approach to the pain problem (as has the original); however, compared to the original, it paid more attention to the practical application of the information given. Berwick et al's (1989) back school consisted of a single 4-hour session during which the anatomy, pathophysiology, ergonomics and psychology of back pain complaints were discussed. The mini back school presented by Stankovic and Johnell (1990) consisted of a single 45-minute session in which the anatomy and function of the back were explained. Patients at this back school were advised to refrain from exercise but lie supine in the semi-Fowler position several times during the day, but to keep on the move during the day in order to avoid inactivity syndrome. The back school presented by Postacchini et al (1988) was based on the model of the Canadian Back Education Unit (Hall, 1980). In the Maastricht back school developed by Keijser et al (1989), a much wider range of the possible factors influencing low back pain was discussed. The Maastricht back school consisted of seven lessons and approached pain in a multidimensional way: in addition to physical factors, attention was given to the psychosocial factors influencing LBP. Furthermore, exercises that strengthen the back and abdominal muscles and relaxation exercises were performed during the lessons.

### SELECTION CRITERIA AND NUMBERS OF PATIENTS INVOLVED

Three studies included employees with LBP (Bergquist-Ullman and Larsson, 1977; Hurri, 1989; Stankovic and Johnell, 1990). Contrary to the studies of Hurri and of Bergquist-Ullman and Larsson, Stankovic and Johnell's (1990) randomised trial involved employees from more than one company. All of the studies included patients suffering from non-specific LBP. However, what is defined as non-specific LBP in one study is sometimes defined as specific LBP in another. In the study presented by Lankhorst et al (1983), for example, a scoliosis of more than 15 degrees was an exclusion criterion; this was not the case in the other seven studies. Four of the eight studies were restricted to a pain duration of more than 6 months; three were restricted to a pain duration of less than 6 months. In the study presented by Postacchini et al (1988), patients with (sub)acute as well as chronic LBP were admitted. Klaber Moffett et al (1986), Stankovic and Johnell (1990), Hurri (1989) and Berwick et al (1989) excluded patients with a prior history of back surgery. In the studies presented by Stankovic and Johnell (1990) and Postacchini et al (1988), pregnant women were also excluded. Klaber Moffett et al (1986) also excluded patients who were concurrently receiving physiotherapy treatment. Hurri (1989) included patients with LBP symptoms on at least 1 day each week during the month preceding the initial examination. Lankhorst et al (1983) used the inclusion criterion of not responding to conventional physiotherapy; Bergquist-Ullman and Larsson (1977) and Berwick et al (1989) required a pain-free year before onset of the current episode of LBP. In the studies reported by Postacchini et al (1988) and Keijser et al (1989), patients were excluded when there was evidence of some psychological or psychiatric morbidity. Postacchini et al (1988) also excluded patients with serious general diseases or when medico-legal litigation was pending.

Because of the various selection criteria used, it is not easy to interpret, generalise and compare the results of the eight available randomised
Table 1  
Characteristics of randomised trials on the efficacy of back schools

<table>
<thead>
<tr>
<th>Duration</th>
<th>Content</th>
<th>Selection criteria</th>
<th>Number of patients</th>
<th>Interventions</th>
<th>Effect parameters</th>
<th>Follow-up</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergquist-Ullman and Larsson (1977)</td>
<td>4 x 45 min in 2 weeks</td>
<td>LBP &lt; 3 months</td>
<td>217</td>
<td>Back school</td>
<td>Pain</td>
<td>12 months</td>
<td>Effective</td>
</tr>
<tr>
<td>Lankhorst et al (1983)</td>
<td>4 x 45 min in 2 weeks</td>
<td>Non-specific LBP</td>
<td>48</td>
<td>Combined physiotherapy</td>
<td>Functional restrictions</td>
<td>12 months</td>
<td>Effective</td>
</tr>
<tr>
<td>Hurri (1989)</td>
<td>6 x 60 min in 3 weeks</td>
<td>LBP &gt; 6 months</td>
<td>188</td>
<td>Back school placebo shortwave</td>
<td>Spinal mobility</td>
<td>12 months</td>
<td>Effective</td>
</tr>
<tr>
<td>Stankovic and Johnell (1990)</td>
<td>1 x 45 min</td>
<td>LBP &gt; 12 months</td>
<td>95</td>
<td>Back school written instructions</td>
<td>Medical consumption</td>
<td>12 months</td>
<td>Not effective</td>
</tr>
<tr>
<td>Berwick et al (1989)</td>
<td>1 x 240 min</td>
<td>LBP &lt; 4 weeks</td>
<td>222</td>
<td>Back school McKenzie treatment</td>
<td>Work absenteeism</td>
<td>12 months</td>
<td>Not effective</td>
</tr>
<tr>
<td>Klaber Moffett et al (1986)</td>
<td>3 x 60 min in 1 week</td>
<td>Non-specific LBP</td>
<td>92</td>
<td>Back school exercise-only regimen</td>
<td>Activity test</td>
<td>18 months</td>
<td>Not effective</td>
</tr>
<tr>
<td>Postacchini et al (1988)</td>
<td>4 x 60 min in 1 week</td>
<td>No history of back surgery</td>
<td>398</td>
<td>Back school (in chronic LBP) waiting list control group</td>
<td>Pain</td>
<td>16 weeks</td>
<td>Effective</td>
</tr>
<tr>
<td>Kejser et al (1989)</td>
<td>7 x 90 min in 7 weeks</td>
<td>LBP &gt; 6 months</td>
<td>40</td>
<td>Back school (in acute LBP) Bed rest</td>
<td>Psychological factors</td>
<td>6 months</td>
<td>Neither effective nor ineffective</td>
</tr>
</tbody>
</table>
trials. Even when identical back school programmes were offered, there are still pitfalls. For example, the selection criteria used by Bergquist-Ullman and Larsson (1977) and Lankhorst et al (1983) were completely different. In principle, any of the differences in the selection criteria could be responsible for the differences in the effects found.

In several of the studies, only a small number of patients was used. Hence, only relatively large differences between the experimental and control groups on the effect parameters could be statistically significant. In his review, Grahame (1980) pointed out that a minimum of 100 patients per group is necessary to achieve statistical confirmation of trends showing differences between the groups; only Bergquist-Ullman and Larsson (1977), Berwick et al (1989) and Hurri (1989) approached this.

INTERVENTIONS AND EXECUTION OF THE STUDY

In all of the studies, patients were randomly assigned to either a back school group or a control group. Whether the control group received any intervention and, if so, what kind of intervention, is shown in Table 1. One study used a waiting list control group, i.e. after a period of waiting, the control group was admitted to the back school course. In two studies on the efficacy of the Swedish back school, a placebo group received detuned shortwave applications to the back (Bergquist-Ullman and Larsson, 1977; Lankhorst et al, 1983). Besides a placebo group, Bergquist-Ullman and Larsson (1977) included a group receiving physiotherapy. In the study presented by Postacchini et al (1988), their patients were randomly assigned to either manipulation, drug therapy, physiotherapy, bed rest (only in acute LBP), back school (only in chronic LBP) or placebo drug treatment. In the study presented by Hurri (1989), the patients in the control group were presented with information in a written form. Klaber Moffett et al (1986) compared their back school group with a group of LBP patients receiving an exercise-only regimen. The mini back school group of Stankovic and Johnell (1990) was compared to a group receiving the McKenzie method of treatment. In the study by Berwick et al (1989), the back school group was compared with a group which, in addition to back school treatment, received a "compliance package" programme, designed to encourage appropriate self-management for back pain, and with another group receiving the usual care for their back pain.

Research on the efficacy of back schools has to take into account the role of non-specific effects of the treatment. Participation in a back school programme, for example, could lead to an improvement unrelated to the specific content of the programme. Including placebo treatment in the study design provides insight into these non-specific effects. A placebo treatment is a treatment which cannot be distinguished externally from the experimental treatment, but in which the supposed active component is left out. For back schools, this placebo treatment would consist of a number of group sessions in which the supposed active components "education" and "training" are left out. This is called a placebo-attention group and can be realised by organising a self-help group. In research on the efficacy of back school programmes, detuned shortwave applications and drug therapy have been used as a placebo (Bergquist-Ullman and Larsson, 1977; Lankhorst et al, 1983; Postacchini et al, 1988). The value of these placebo treatments is questionable. Research involving no placebo compares different treatment strategies. The question is which treatment has the greatest effect, and not what are the differences between the specific and the non-specific effects of the treatment. Such studies are called pragmatic trials. The so-called explanatory trial (including a placebo group) allows the detection of possible differences between the specific and non-specific effects of a treatment. The study presented by Bergquist-Ullman and Larsson (1977) used a combination of both types of design. On the one hand, a comparison was made between back school treatment and physiotherapy and, on the other, a placebo treatment was used to study the effects of the supposed active component of back school treatment and of physiotherapy treatment.
It is an inherent feature of randomised trials that patients are randomly assigned to a back school treatment group or to a comparison group. The randomisation, however, is not always performed adequately. In the study presented by Lankhorst et al. (1983), the first six consecutive patients were assigned to the experimental group, the next six to the control group, and so on. The results of this study could be said to be biased, due to the fact that the first six patients were perhaps not comparable to the next six patients. The reasons for dropping out were seldom mentioned in the studies. Such an effect could, however, be biased because of selective drop out. In the studies of Bergquist-Ullman and Larsson (1977), Lankhorst et al. (1983), Hurri (1989), Berwick et al. (1989) and Stankovic and Johnell (1990), measurements were made to determine the long-term effects of the back school programme 1 year or more after the intervention.

**EFFECT PARAMETERS**

Pain is difficult to measure. It encompasses subjective, physiological and behavioural aspects. This diversity is not always sufficiently reflected in the effect parameters used. What is measured in order to determine the efficacy of back schools is summarised in Table 1. Although a reduction in pain intensity is rarely mentioned as a main objective of back school programmes, it plays an important role in its assessment (Guck, 1984). Klaber Moffett et al. (1986), Lankhorst et al. (1983), Postacchini et al. (1988) and Stankovic and Johnell (1990) only measured pain intensity. In the study of Postacchini et al. (1988), pain intensity was not only assessed by the patients themselves but also by their physicians. Linton and Kamwendo (1987) concluded in their review that reductions in pain intensity are not to be expected from back school programmes. The operationalisation of pain in the studies by Bergquist-Ullman and Larsson (1977), Keijser et al. (1989), Hurri (1989) and Berwick et al. (1989) consisted of measuring the multidimensional aspects of pain.

The degree of disability as a result of back pain (functional restrictions) was measured by Klaber Moffett et al. (1986), Lankhorst et al. (1983), Berwick et al. (1989), Hurri (1989) and Postacchini et al. (1988). Klaber Moffett et al. (1986) extended this to include observations of patient activities. Lankhorst et al. (1983), Hurri (1989), Postacchini et al. (1988) and Stankovic and Johnell (1990) measured spinal mobility. Keijser et al. (1989) used a so-called pain diary, which enabled them to make statements regarding the degree of disability of patients: a low level of activity was equivalent to a high level of disability. In summary, it appears that in six of the eight studies the degree of disability was used as an effect parameter.

The psychosocial factors influencing pain were measured by Keijser et al. (1989) and Berwick et al. (1989), while subjects' knowledge regarding back pain was measured by Klaber Moffett et al. (1986) and Hurri (1989) only. Bergquist-Ullman and Larsson (1977), Berwick et al. (1989), Hurri (1989) and Stankovic and Johnell (1990) recorded absenteeism from work, which is important with regard to a cost–benefit analysis of back school programmes. In the study reported on by Stankovic and Johnell (1990), the mean duration before returning to work as well as the number of recurrences of LBP were measured.

It is remarkable that the main aim of back school programmes (pain management) was not measured directly. In the study presented by Stankovic and Johnell (1990), patients' self-help was measured. From their article it is not clear whether this was supposed to be similar to pain management. In most of the studies, a number of factors which have to do with pain (e.g. level of activity) formed the variables by which the efficacy of the back school was assessed. The implicit assumption at the basis of this practice is that an improvement in, for example, an activity list, indicates that a patient is better able to manage pain. However, it is possible that a patient is able to manage pain without showing a corresponding improvement in their level of activity. It is therefore preferable to measure pain manage-
ment directly. The concept of self-efficacy offers a theoretical basis for this (Bandura, 1977). Perceived self-efficacy refers to people's judgements of their capabilities to execute a certain behaviour. The role of self-efficacy in the management of pain has been demonstrated in several studies (Council et al, 1988; O'Leary, 1985). Self-efficacy in managing pain can be seen as a common factor underlying both cognitive and affective influences on pain experience. Research has shown that self-efficacious patients are more inclined to manage their pain with the help of non-medical strategies (O'Leary, 1985). Based on the theory of self-efficacy, pain management can be measured in the following way. The patient is asked whether he or she is capable of managing his or her back pain in a number of pain-related situations. In this way, information regarding the target of pain management is provided directly by the patient him or herself. It is also remarkable that most of the questionnaires used in these studies were not standardised and that they were developed solely for a single study. This was especially true for those questionnaires enquiring about medical consumption and work absenteeism. This lack of standardisation means that the validity and reliability of the questionnaires are, in most instances, not clear. Consequently, the comparability of the data collected by means of these questionnaires becomes doubtful.

DISCUSSION

Pain was operationalised in all of the studies. Bergquist-Ullman and Larsson (1977) suggest that the Swedish back school is an effective means of pain management for acute and sub-acute LBP patients. Lankhorst et al (1983) were unable to demonstrate any effect of their back school on pain in chronic LBP patients. In the studies reported on by Postacchini et al (1988), Klaber Moffett et al (1986) and Hurri (1989), it was concluded that for chronic LBP patients the back school is effective for the effect parameter pain. Hurri's conclusion, however, was only supported by the data at 6 months follow-up. At 12 months follow-up, the effect of the back school on pain appeared to have receded. The effect on pain of those back schools presented by Berwick et al (1989) and Stankovic and Johnell (1990), both consisting of a single session, seems to have been limited for (sub)acute LBP. The results of a pilot study on pain presented by Keijzers et al (1989) could not confirm nor refute the supposed positive effect of the back school.

Both internal and external validity are important for an overall judgement on the efficacy of back schools. With regard to internal validity, the following gaps were found. The numbers of patients involved in the studies were relatively small and randomisation was not always properly executed. None of the studies made use of a placebo-attention group and the main target of the back school was never evaluated directly. Because of these methodological problems, it is as yet not possible to formulate a valid judgement on the efficacy of the back schools studied in the available trials. The studies of Berwick et al (1989), Bergquist-Ullman and Larsson (1977) and Hurri (1989) seem to be the most informative. These studies involved a relatively large number of patients, they determined the long-term effect of the back school and they recorded work absenteeism and medical consumption, so that a cost–benefit analysis could be made. Of the back schools assessed in these three studies, at 12 months follow-up only that of Bergquist-Ullman and Larsson (1977) was found to be effective. A back school consisting of a single 4-hour session of instruction was clearly ineffective (Berwick et al, 1989).

The problems concerning the internal validity of the studies impair the comparability of the studies (external validity) as well. In addition to this, the selection criteria appeared to vary a great deal. This would not be a problem if the differences found in the criteria did not have an influence on the effect, but they often do. Two studies are hardly comparable if, for example, selection is based on psychological variables (Keijzers et al, 1989) in one study and not in the other. It is often assumed that the presence or absence of certain psychological characteristics has a prognostic influence on the effect of the intervention. In addition, criteria which sound similar also need some critical attention: non-
specific LBP, for example, turned out to be multi-interpretable. The same is true for the different contents of back schools. Back schools in which no attention is paid to the execution of exercises during lessons are hardly comparable to back schools in which such exercises are practised. The studies presented by Bergquist-Ullman and Larsson (1977) and Hurri (1989) investigated the efficacy of Swedish back schools. Nevertheless, these studies are hardly comparable, because of the different selection criteria used. Patients with a pain duration of 12 months or more are probably not comparable to patients with a pain duration of less than 3 months.

Future research on the efficacy of back schools will have to guarantee both internal and external validity. With regard to internal validity, for example, randomised trials need to include sufficient numbers of patients and pain management needs to be measured directly. External validity can be improved by reducing the differences in interventions and patients between various studies. This means that questions such as which patients are to be admitted should be taken very seriously. A certain degree of consensus about the central issues seems desirable. Such a consensus will increase the comparability of studies on the efficacy of back schools. The standardisation of issues with regard to both internal and external validity would improve the justification of conclusions within trials as well as between trials.

Acknowledgement

This study was supported by a grant from the Dutch Prevention Foundation.

References

Haanen HCM 1984 Een epidemiologisch onderzoek naar lage rugpijn. Dissertation, Erasmus University, Rotterdam
Hall H 1980 The Canadian back education units. Physiotherapy 66: 115–117
Spitzer WO, LeBlanc FE, Dupuis M 1987 Scientific approach to the assessment and management of activity-related spinal disorders. Spine 7: 1–19 (suppl 1)