Development of a reliable method to assess footwear comfort
during running

Anne Mündermann a,*, Benno M. Nigg a, Darren J. Stefanyshyn a, R. Neil Humble b

a Human Performance Laboratory, Faculty of Kinesiology, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada, T2N 1N4
b Division of Podiatry, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada

Received 3 May 2001; accepted 5 November 2001

Abstract

The purposes of this study were: (a) to determine whether subjects are able to distinguish between differences in footwear with respect to footwear comfort; and (b) to determine how reliably footwear comfort can be assessed using a visual analogue scale (VAS) and a protocol including a control condition during running. Intraclass correlation coefficients (ICCs) between comfort ratings for repeated conditions were high (ICC = 0.799). Differences in comfort ratings between the insert conditions were significant. A paired t-test revealed a significant difference in overall comfort ratings for the control insert when tested after the soft insert compared to when tested after the hard insert (P = 0.008). The results of this study showed that VASs provide a reliable measure to assess footwear comfort during running under the conditions that: (a) a control condition is included; and (b) the average comfort rating of sessions 4–6 is used. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Comfort; Visual analogue scale; Footwear; Running

1. Introduction

Comfort is an important factor for footwear in recreational physical activities. Most people can quickly identify comfortable or non-comfortable footwear situations. Increasing interest in footwear comfort resulted in several investigations that associated comfort with plantar pressure distribution [1,2], vertical impact force, rearfoot motion [3], foot and leg shape and alignment [4], and foot sensitivity [5]. It has been speculated that comfort is related to muscle activation and, thus, to fatigue and performance [6,7]. The specific design, physical properties and construction of footwear have been shown to affect these variables [8,9] and, thus, seem to be important factors for footwear comfort.

However, comfort is difficult to quantify and it has been suggested that comfort cannot be measured directly [10,11]. Despite the reported measurements of comfort [1–5], to date a reliable method to assess footwear comfort has not been developed. All studies investigating comfort in the past had one common shortcoming: The reliability of the comfort measures used was not reported or, if reported, poor. In fact, only two studies reported on the repeatability of the comfort measures used. Miller et al. [4] and Mündermann et al. [5] found rather low intraclass correlation coefficients (ICCs) of comfort ratings for repeated conditions. These results can be in part explained by the following considerations.

Individuals have some feeling regarding the absolute liking of an object, but such sentiments are influenced by the range of objects available. In other words, when expressing sentiments, subjects tend to anchor their responses in terms of: (1) similar stimuli that they have experienced in the past; and (2) the range of stimuli in the set presented [12]. Thus, if someone is required to rate the comfort of a shoe they cannot make such response unless they subjectively compare the features of the shoe within those of the many shoes they have worn before. Therefore, it is speculated that the reliability of a comfort measure can be considerably improved by adding a control condition to the testing protocol.
such that all conditions to be assessed are tested after and, thus, compared to the same footwear condition.

Since the repeatability of the measures in earlier studies was poor, it must be questioned whether the results of these studies provide sufficient evidence for the existence of a relationship between footwear comfort and biomechanical variable and subject characteristics. A reliable measure for footwear comfort is required to quantify comfort and to determine valid relationships between comfort and shoe constructions, subject characteristics, and biomechanical variables.

Most investigators used ordinal scales to assess footwear comfort [1–4]. In ordinal scales or ranking scales, objects are ordered from ‘least’ to ‘most’ with respect to an attribute, e.g. comfort. However, in such scales, there is no indication of how much, in an absolute sense, any of the objects possesses the attribute [12]. Furthermore, no information can be obtained as to how different the conditions are with respect to comfort.

One particular kind of ordinal scale is where subjects rate various objects on a 7- or 15-step scale ranging from ‘least’ to ‘most’. For example, the Borg-scale was initially developed to assess perceived exertion [13] and has been used to assess footwear comfort [2–4]. In general, this scale allows the determination of relative differences between conditions. However, due to the discrete spacing of ratings, very small differences between conditions cannot be detected. Furthermore, non-parametric statistical methods must be used to evaluate ordinal data. Because the sets of possible answers consist of discrete numbers, measuring comfort using an ordinal scale will introduce errors to correlations between comfort and other variables such as biomechanical variables that are measured on continuous scales.

It is proposed that footwear comfort should be measured using a continuous scale. Visual analogue scales (VASs), for example, have been proven to be a reliable measure to assess subjective pain [14,15]. VAS responses are very easy to obtain from patients and normal volunteers and require little instruction. However, not all pain VASs are bias-free [16].

Studies comparing the use of different types of VASs have shown that the sensitivity and reliability of VASs are somewhat influenced by the words used to anchor the endpoints, by the length of the VAS, and by other factors [17,18]. Those VASs that most clearly delineate extremes (e.g. the best condition imaginable, the worst condition) and are 100–150 mm in length have been shown to have the greatest sensitivity and are the least vulnerable to distortions or biases in ratings [18]. It has been suggested that giving specific instructions to subjects will increase the reliability of VASs. For example, Price et al. [14] found an inter-test correlation coefficient of 0.970 for a 150 mm VAS and specific instructions.

Table 1
Shoe insert material tested in this study

<table>
<thead>
<tr>
<th>Shoe</th>
<th>Description</th>
<th>Bottom layer</th>
<th>Top layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Control insert</td>
<td>3 mm EVA a</td>
<td>3 mm Spenco</td>
</tr>
<tr>
<td>I</td>
<td>Soft insert</td>
<td>3 mm Poron</td>
<td>3 mm Spenco</td>
</tr>
<tr>
<td>I</td>
<td>Hard insert</td>
<td>3 mm EVA a</td>
<td>3 mm Korex</td>
</tr>
<tr>
<td>I</td>
<td>Medial wedge</td>
<td>6 mm wedge EVA a</td>
<td>3 mm Spenco</td>
</tr>
</tbody>
</table>

a Ethylene vinyl acetate.

Therefore, the purposes of this study were: (a) to determine whether subjects are able to distinguish between differences in footwear with respect to footwear comfort; and (b) to determine how reliable footwear comfort can be assessed using a VAS and a protocol including a control condition during running.

2. Methods

2.1. Procedures

Nine volunteers (4 females, 5 males; age 26.7±4.9 years; height 171.2±7.1 cm; weight 71.7±8.5 kg) gave informed consent and participated in this study. All subjects were physically active and free of lower extremity injury or pain. A standard running shoe was used (SuperNova, Adidas Inc.) with four different shoe inserts. The four footwear conditions are described in Table 1. The top layer of all shoe inserts had the same color and, thus, subjects were not able to visually notice any difference between conditions. Shore C values of the materials used are listed in Table 2. All chosen materials are materials commonly used for foot orthotics.

Data for each subject were collected in ten 45 min sessions on ten consecutive workdays. Prior to each session, subjects were instructed with regards to the testing protocol and the comfort assessment. In each session, subjects ran two laps using the control insert on a 450 m indoor running track to warm up. Subjects were instructed that the first, third, fifth and seventh condition were the control condition but were unaware of the order of the test inserts. They then ran one lap using the insert conditions C–I 1–C–I 2–C–I k–C–I R, where C corresponded to the control condition, I to an

Table 2
Hardness of the materials used

<table>
<thead>
<tr>
<th>Material</th>
<th>Shoe C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spenco</td>
<td>40</td>
</tr>
<tr>
<td>EVA</td>
<td>73</td>
</tr>
<tr>
<td>Poron</td>
<td>28</td>
</tr>
<tr>
<td>Korex</td>
<td>76</td>
</tr>
</tbody>
</table>
insert condition and IR to an arbitrarily chosen repeat insert condition. The sequence of the insert conditions was randomly varied. After each lap, subjects assessed footwear comfort for the tested shoe insert condition.

2.2. Comfort measure

The comfort measure consisted of a 150 mm visual analogue scale with the left end labeled ‘not comfortable at all’ (0 comfort points) and the right end ‘most comfortable condition imaginable’ (15 comfort points). Since many aspects of footwear may influence comfort, specific comfort ratings were included such as forefoot cushioning, heel cushioning, arch height, heel cup fit, shoe heel width, shoe forefoot width, and shoe length (Fig. 1). Written instructions were given to each subject to eliminate differences in assessments between subjects and sessions resulting from inconsistent instructions (Fig. 2). The first comfort rating of the control condition in each session was used for further analysis as this rating was taken after wearing the control condition for the warm-up and other preceding conditions could have affected this comfort rating.

2.3. Statistical analysis

Intraclass correlation coefficients for repeated comfort ratings were determined for all subject and sessions and individually for subjects and sessions. Differences in ratings between conditions were detected using univariate ANOVA and student t-tests for repeated conditions. The level of significance was set at $\alpha = 0.05$ for all tests.
3. Results

3.1. Overall comfort

ANOVA and Student’s t-tests revealed significant differences in the overall ratings for the four shoe inserts. The control and the soft inserts were on average rated more than four comfort points higher than the hard insert and the insert with the medial wedge ($P < 0.001$). Although rather small (0.56 comfort points), the difference in average overall ratings between the control insert and the soft insert was significant ($P = 0.047$) (Fig. 3).

Overall comfort ratings for all subjects, sessions, and conditions are shown in Fig. 4. Several different patterns of comfort ratings were observed.
- The control condition was rated more consistently than the remaining insert conditions.
- Some subjects rated conditions in a very consistent way (e.g. subjects 4 and 5) whereas others showed large fluctuations in ratings (e.g. subject 6).
- Some individuals made use of the whole length of the visual analogue scale (vertical axis) (e.g. subject 3) compared to others who only used a fraction of the scale (e.g. subject 5).
- Some subjects showed a converging pattern in their comfort ratings (e.g. subject 7) and the comfort

![Graph showing comfort ratings for different subjects and conditions](image-url)

Fig. 4. Overall comfort ratings for all subjects, sessions, and conditions. Comfort ratings are shown on a scale from 1 (not comfortable at all) to 15 (most comfortable condition imaginable).
Fig. 5 shows averages of four specific comfort ratings. The differences in heel cushioning comfort ratings between conditions corresponded to the differences in the overall comfort rating. Differences in forefoot cushioning and medio-lateral control comfort ratings between the control insert and the soft insert were not significant. The largest deviation from the control insert in both heel and forefoot cushioning comfort ratings occurred for the hard insert whereas the largest deviation in average medio-lateral comfort rating was observed for the medial wedge insert (Fig. 6). No significant differences in the length comfort ratings for the four insert conditions were observed.

3.3. Intra-test repeatability

Intraclass correlation coefficient for intra-test repeatability was 0.799 for all subjects and all conditions and ranged between 0.108 and 0.952 for individuals (Table 3). Correlations were significant for six of the nine subjects. Fig. 7 shows a scatter plot of repeated comfort ratings for all subjects. If subjects with $r$-values smaller than 0.500 were excluded, the ICC for repeated conditions for all sessions increased to 0.911. The $r$-values for both groups of subjects (ICC > 0.500, ICC <
Table 3
Intraclass correlation coefficients of comfort ratings for repeated conditions

<table>
<thead>
<tr>
<th>Subject</th>
<th>ICC</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.821</td>
<td>0.004</td>
</tr>
<tr>
<td>2</td>
<td>0.108</td>
<td>0.766</td>
</tr>
<tr>
<td>3</td>
<td>0.952</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.968</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5</td>
<td>0.896</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>6</td>
<td>0.185</td>
<td>0.606</td>
</tr>
<tr>
<td>7</td>
<td>0.935</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>8</td>
<td>0.709</td>
<td>0.022</td>
</tr>
<tr>
<td>9</td>
<td>0.379</td>
<td>0.280</td>
</tr>
</tbody>
</table>

All subjects/all sessions 0.799 <0.001
All subjects/sessions 1–3 0.796 <0.001
All subjects/sessions 4–6 0.796 <0.001
All subjects/sessions 7–9 0.796 <0.001
Subjects >0.500/all sessions 0.834 <0.001
Subjects <0.500/all sessions 0.373 0.042
Subjects <0.500/sessions 1–3 0.111 0.776
Subjects <0.500/sessions 4–6 −0.083 0.833
Subjects <0.500/sessions 7–9 0.705 0.034

0.500) increased from sessions 1–3 to sessions 4–6 and was the greatest for sessions 7–9.

3.4. Inter-test repeatability

In general, the variability in comfort ratings between sessions decreased with increasing session number. Specifically, comfort ratings of sessions one to three varied more than comfort ratings of sessions 4–10 (Fig. 4). In most cases, the comfort ratings leveled out after session 3. The comfort rating of session one differed in average by more than two comfort points from the average rating of sessions 4–10 (Fig. 8). The average comfort rating for sessions 4–6 differed by less than 0.5 comfort points from the average rating of sessions 4–10. The average comfort rating for sessions 7–9 did not provide any improvement.

3.5. Preceding insert

For all subjects, a paired t-test revealed a significant difference in overall comfort ratings for the control insert when tested after the soft insert compared to when tested after the hard insert (P = 0.008). The average difference in the repeat assessments of the control insert following different preceding conditions was 0.5 comfort points for all subjects with a difference of up to 1.6 comfort points for one subject. It is important to note that the ICC of this particular subject was greater than 0.500.

4. Discussion

4.1. Overall comfort

Although differences between subjects in comfort ratings for the four shoe inserts existed, a general statement can be made based on the average overall comfort ratings. In general, subjects seem to prefer soft over hard materials. This finding is in agreement with an earlier study [5]. The medial wedge insert was rated significantly lower in comfort than the flat control insert. Thus, overall comfort ratings can be affected by varying shoe inserts worn with identical running shoes.

In average, the control condition was rated highest. This may be due to the fact that the materials and shape of the control condition was most similar to sockliners commercially available with running shoes. However, some subjects did rate the soft insert (subjects 1, 3, 4 and 9) or the medial wedge insert (subject 7) higher than the control insert. This result indicates that different individuals have different preferences with respect to shoe insert material and shape.
4.2. Specific comfort

Average heel cushioning, forefoot cushioning, and medio-lateral control comfort ratings corresponded to the varying shoe insert components used. The small difference in heel and forefoot cushioning comfort ratings between the control insert and the soft insert may be due to the fact that the top layers of these shoe inserts were made from the same material and, thus, the difference in material of the bottom layer may have been masked.

Length comfort ratings were almost identical for all insert conditions which was expected since the length of the footwear was identical for all insert conditions. The magnitudes of the average differences in specific comfort ratings are summarized in Fig. 6. This result demonstrates the ability of subjects to detect actual differences in footwear and differentiate between them with respect to specific comfort ratings.

In this study, average comfort ratings of four shoe insert conditions were compared using parametric statistics based on the assumption that the VAS used provided parametric data. However, to date footwear comfort is a subjective variable that yet needs to be related to physical characteristics of footwear. Comfort ratings obtained using a VAS can then be adjusted according to the nature of these relationships.

4.3. Intra-test repeatability

As hypothesized, the reliability of a VAS and a protocol including a control condition to assess footwear comfort was very good. The results of this study showed that some individuals are more repeatable in their comfort assessment than others. When excluding individuals with low repeatability, the reliability of the used protocol was excellent.

Several reasons may be responsible for the low repeatability in comfort assessment of some individuals. First, these subjects may have low foot sensitivity. Each individual has a given sensory threshold to pressure and vibration stimuli at the plantar surface of the foot. The contact of the foot with the ground causes an input signal into the body. The input signal into the system can be modified by differences in footwear materials [19]. The modified input signal may be above the given threshold for some footwear condition and below for others and, thus, lead to differences in comfort perception. Foot sensitivity measurements are required to determine the relevance of this parameter.

Second, it is possible that some individuals are initially not able to use a VAS in a reliable way and may need some time to get used to this method. This speculation is supported by the fact that subjects with ICC < 0.500 were able to reliably assess footwear comfort in sessions 7–9 (Table 3).

4.4. Inter-test repeatability

Although the variability of comfort ratings within session one to three was greater than within sessions four to ten, the within-session repeatability of these ratings was still very good. Thus, the greater variability within the first three sessions is an indication for an adjustment to the shoe insert conditions and that footwear comfort changes over time. Although earlier studies assessed footwear comfort just once [1–5], the relationships between footwear comfort and other variables may be valid in terms of short-term assessment. However, it must be questioned whether these relationships are valid as a long-term evaluation of footwear.

4.5. Preceding insert

Footwear worn prior to a comfort assessment has been shown to affect footwear comfort. For example, if a control insert is tested after harder inserts it may be perceived as softer compared to being tested after a softer condition. The effect of preceding footwear on following comfort assessments seems to be subject specific. Thus, in a short-term perspective, some subjects get quickly used to a condition that then becomes their standard used for following assessments. In comparison, other subjects seem to be able to recall the perception of the control condition. However, the results of this study did not allow the identification of characteristics that distinguish between these subjects. Therefore, a control condition should always be included for comfort assessment.

The results of this study showed that VASs provide a reliable measure to assess footwear comfort during running. The following recommendations should be taken into consideration when assessing footwear comfort:

- a control condition should be used before each test condition;
- subject specific repeatability should be determined; and
- for long-term comfort, assessments of sessions four to six should be used.

In this study, subjects were tested on 10 days with one session per day. Future research should be conducted to determine the optimal combination of number of conditions per session, running distance per condition, total running distance per day, and number of test days. Furthermore, the appropriate selection of the control condition requires further investigation.

Especially for customized footwear such as foot orthotics, comfort may play an increasingly important role. To date, clinicians allow patients an adjustment period of two to three weeks during or after which orthotics may be modified. However, the patient feedback is commonly based on subjective comments. A
standardized method to assess comfort may improve the outcome of evaluations during this period.

Acknowledgements

The authors thank the Canadian Department of Foreign Affairs and International Trade, the International Society of Biomechanics and the Natural Science and Engineering Council of Canada for their financial support.

References