OCRA inter- and intra-ergonomist reliability in ten video recorded work tasks

Ida-Märta Rhénab,c, Natalja Balliuc, Mikael Forsmanab,c,d

aProduct and production development dep., Chalmers University of technology, Gothenburg, SWEDEN; bFraunhofer Chalmers research centre, Gothenburg, SWEDEN; cCentre for Occupational and Environmental Medicine, Stockholm County Council, SWEDEN; dIMM Institute of Environmental Medicine, Karolinska Institutet, Stockholm, SWEDEN

Keywords Agreement, Observation, Risk assessment, Biomechanical exposures

Introduction

In occupational health services (OHS) observation-based risk assessment methods are generally accepted as an objective mean for evaluation of exposure. The assessments are normally performed by well-trained ergonomists and based on real time observations or observing video recordings.

According to a review of 30 assessment methods (Takala et al. 2010), several methods were insufficient tested for validity and reliability. Furthermore, OHS have called for systematic and scientifically based methods that are user friendly and cost effective (Andersson et al. 2006). The OCRA (Occupational Repetitive Actions) method is included as a reference method in ISO (ISO 11228-3) and CEN (1005-5) standards regarding risk assessment of upper limb repetitive actions. It includes time-based exposure variables such as recovery and frequency, and is more comprehensive than most other methods. Moreover, the final risk score, that predict the risk of developing musculoskeletal disorders, is based on epidemiological research (Colombini 1998, Occhipinti 1998).

In the present paper, the OCRA checklist (Occhipinti and Colombini 2006) is used, which is a simplified version of the more comprehensive OCRA index. The method may be a seen as a suitable alternative for OHS-ergonomists. However, the method is not commonly used in Sweden and it has not until now been translated to Swedish. Moreover, validity and reliability studies are limited.

The purpose of the present study was to examine the reliability of the OCRA checklist assessed by experienced ergonomist in a broad range of occupational situations. The objectives were to, with respect to exposure variables and the calculated risk level, estimate the consistency of 1) assessments performed by different ergonomists (inter-rater reliability) and 2) repeated assessments performed by one and the same ergonomist (intra-rater reliability) of the OCRA checklist.

Method

Eleven female OHS-ergonomists, with more than five years of experience of ergonomic risk assessments, were educated in the OCRA checklist method, translated to Swedish by the researches of the present study with support from prof. Daniela Colombini. Each ergonomist made assessments of 10 video-recorded work tasks (supermarket work, meat cutting and packing, engine assembly, cleaning, post sorting and hairdressing) chosen for assessment because of their varying levels of upper limb repetitive, load and complexity.

Video-recordings (2-6 minutes) from each task from two or three camera-angles, were synchronised and shown simultaneously. Individually, the ergonomists could pause or repeat the video as they needed. Information such as e.g. work task length, breaks, subjective perceived work effort and information concerning organisational work structure was given separately.

Rating and risk level evaluation of the ten work task were performed according to the OCRA general procedure (Occhipinti and Colombini 2006), which includes assessment of six primary exposure variables: i) frequency of technical actions, ii) forceful exertions, iii) awkward postures and motions, iv) work duration, v) lack of recovery time and vi) additional risk factors. The OCRA checklist score is produced by summation of frequency, force, posture and additional factors scores, and weighted by the scores for the lack of sufficient recovery and work duration variables. Risk of developing musculoskeletal disorders was evaluated by OCRA
score classification criteria’s consisting of five risk levels; acceptable, very low, low, medium and high (Occhipinti and Colombini 2006). All ergonomists but one repeated assessments after 6 weeks.

Three parameters were computed as measures of the inter- and intra-rater reliability: Intraclass correlation, ICC (2, 1), as suggested by Shrout and Fleiss (1979); and Lights kappa (Light 1971) and linearly weighted kappa (Fleiss and Cohen 1973), which were both computed as Light suggested by averaging pairwise kappas (inter-rater reliability), and averaging individual test-retest, Cohen’s and linearly weighted kappas (intra-rater reliability). ICC and kappa values were computed for each exposure variable as well as for the over-all risk level.

Landis and Koch’s (1977) recommendations (0.21-0.4 fair, 0.41-0.6 moderate 0.61-0.8 substantial, 0.81-1 almost perfect and 1 perfect) was used for interpretation of kappa values. Cicchetti (1994) provide similar cut-offs in interpretation of ICC results.

Results

The results show that OCRA risk-level rating differed among the eleven ergonomists and the ten work tasks. The percentage of the over-all risk-level ratings were, in the lowest to highest levels of risk: 27%, 16%, 16%, 33%, 8%. Relative agreement in these over-all risk-level ratings was 39%.

In six of the ten video-clips, four of five possible OCRA risk-levels were represented. One video clip was rated equally (with the lowest level of risk) by all ergonomists while in another video-clip, the ratings were spread from minimum to maximum level of risk.

The parameters of reliability are shown in Table 1 (inter-rater reliability) and Table 2 (intra-rater reliability).

Table 1. Parameters of inter-rater reliability for exposure variables and overall risk level.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Frequency</th>
<th>Force</th>
<th>Posture</th>
<th>Additional factors</th>
<th>Recovery</th>
<th>Duration</th>
<th>Risk level</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>0.63</td>
<td>0.51</td>
<td>0.33</td>
<td>0.40</td>
<td>0.50</td>
<td>0.48</td>
<td>0.58</td>
</tr>
<tr>
<td>Linearly weighted kappa</td>
<td>0.46</td>
<td>0.39</td>
<td>0.23</td>
<td>0.31</td>
<td>0.46</td>
<td>0.48</td>
<td>0.41</td>
</tr>
<tr>
<td>Lights kappa</td>
<td>0.14</td>
<td>0.23</td>
<td>0.10</td>
<td>0.21</td>
<td>0.32</td>
<td>0.36</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Table 2. Parameters of intra-rater reliability for exposure variables and overall risk level.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Frequency</th>
<th>Force</th>
<th>Posture</th>
<th>Additional factors</th>
<th>Recovery</th>
<th>Duration</th>
<th>Risk level</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC</td>
<td>0.78</td>
<td>0.68</td>
<td>0.62</td>
<td>0.58</td>
<td>0.52</td>
<td>0.69</td>
<td>0.72</td>
</tr>
<tr>
<td>Linearly weighted kappa</td>
<td>0.59</td>
<td>0.51</td>
<td>0.43</td>
<td>0.47</td>
<td>0.40</td>
<td>0.62</td>
<td>0.53</td>
</tr>
<tr>
<td>Lights kappa</td>
<td>0.24</td>
<td>0.30</td>
<td>0.25</td>
<td>0.32</td>
<td>0.30</td>
<td>0.50</td>
<td>0.31</td>
</tr>
</tbody>
</table>
Discussion

Inter-rater reliability of the OCRA overall risk-level, showed an ICC of 0.58 with a corresponding average linearly weighted kappa of 0.41 indicating moderate to fair reliability. Lichts kappa generally showed lower kappa-values which is likely due to Lichts tougher punishment in disagreement. However, according to Warren (2013) the linearly weighted kappa is to be preferred for ordinal scales which put our focus on that parameter. Linearly weighted kappa value together with the relative agreement, which for these 5 categories was 39%, are similar to those found in other studies of inter-rater reliability for other observational methods.

The intra-rater agreement was as expected higher than the inter-rater ditto, with an average linearly weighted kappa of 0.53, hence in the middle of Landis and Koch’s moderate range. The corresponding ICC was 0.72, which also is similar to findings in studies of repeated assessments in other observational studies, e.g. ICC 0.56-0.82 for over-all scores (Stephens et al. 2006).

The results differ between the exposure variables. For example, “posture”, revealed poorer inter-rater as well as intra-rater agreement than frequency and force. This is in line with previous studies of other methods (e.g. Ebersole and Armstrong 2006, Stephens et al. 2006), indicating that posture is difficult to assess. A possible contribution in OCRA checklist is that the variable “posture” includes ratings of six sub components, all of which seems difficult to assess.

The reliability of a method determines the level of confidence in the results obtained, and both inter-rater and intra-rater repeatability are considered to be the most important in reliability testing. Since this study is part of a larger project, comparison of the reliability of different methods, as well as studies of the ergonomist’s assessments to expert assessments and technical measurements are planned.

Acknowledgements

This work was supported by the Swedish Research Council for Health, Working Life and Welfare (FORTE; project no. 1212-1202).

References


