

Inter-ergonomist reliability in rating risk level - without any specific method - in ten video recorded work tasks

Mikael FORSMAN^{1,2}, Kristina ELIASSON^{1,3}, Ida-Märta RHEN², Teresia NYMAN^{1,2,3}, Per LINDBERG⁴, Katarina KJELLBERG^{1,2}, Natalja BALLIU² and Peter PALM⁵

¹*IMM Institute of Environmental Medicine, Karolinska Institutet, Stockholm,*

²*Centre for Occupational and Environmental Medicine, Stockholm County Council, ³School of Technology and health, KTH Royal Institute of Technology, Huddinge,*

⁴*Centre for Musculoskeletal Research, University of Gävle,*

⁵*Department of Medical Sciences Occupational and Environmental Medicine, Uppsala University, and Uppsala University Hospital, Uppsala, Sweden*

Keywords. Reliability, Observation, Risk assessment, Biomechanical exposures

1. Introduction

Recently, 30 observational methods for assessment of biomechanical exposures at work were evaluated through a literature review (Takala et al., 2010). It was found that several methods are insufficiently tested in terms of validity and reliability. Only in a few cases estimates of the components of the methods have been validated against technical measurements. Also comparisons between methods' resulting risk levels are rare. The Swedish Work Environment Authority has recently increased the demands on ergonomic risk assessments. These assessments are usually made by ergonomists in occupational health services (OHS).

This study is included in a larger on-going project, with the overall purpose to evaluate six observational methods for assessment of biomechanical exposures of repetitive work in respect of validity, reliability and usability, as well as provide information on which of the methods are best suited for practitioners in risk assessment of repetitive work. The methods' resulting risk levels are compared not only to each other, but also to ergonomists' "own" assessments of the risk of developing musculoskeletal disorders (i.e., assessed without any specific method). The specific aim of this sub-study was to investigate the inter-observer reliability of ergonomists' own risk assessments.

2. Methods

Nine female OHS-ergonomists with more than 5 years of experience of general ergonomic risk assessments, made assessments of 10 video-recorded (2-6 minutes) work tasks (supermarket work, meat cutting and packing, engine assembly, cleaning, post sorting and hairdressing). Video sequences of two or three camera angles were syn-chronized and showed together. For each work task, the ergonomists were given data of the work task length (see Table 1), pause- and rests-schedules, weights of handled goods, physical factors, and the employees' ratings of discomfort, work demands and control.

The ergonomist could pause or repeat the playback as she needed, but the maximum allocated time per work task assessment was 20 minutes. The risk of developing musculoskeletal disorders and need for improvements was rated into green (no risk), yellow

(investigate further), and red (immediate risk) categories. Ratings were done for 8 specific body regions: neck, lower back, right and left shoulders, -arms/elbows, and - wrists/hands), and for the over-all risk.

The agreement of the ratings (in percent), and Light's multi-observer kappa (i.e. Cohen's pairwise kappa averaged over all pairs; Light, 1971; Cohen, 1960) were calculated per body region and for the over-all risk assessment.

3. Results

Of the 720 risk assessments (9 ergonomists, 8 body regions, 10 work tasks) of the separate body regions, 37% were green, 44% yellow and 19% red. Of the over-all risk assessments (Table 1), 14% were green, 50% yellow and 36% red. The consistency among observers differed markedly between the tasks. For three work tasks all three risk categories were represented, whereas in one task all ergonomists rated the over-all risk equally. The average agreement of the ratings were 48% for the body regions, and 57% for the over-all risk assessments, Light's kappa was 0.18 and 0.30, respectively.

Table 1. Work tasks, hours per work task, and the ergonomists ratings of over risk

4. Discussion

The results showed poor to fair inter-observer reliability according to Altman's table for interpretation of kappa (kappa below .2, poor; between 0.21 and 0.40, fair; Altman, 1991). These kappa values will, in the major project, be compared to those of six systematic observation methods.

References

- Altman DG (1991) Practical statistics for medical research. London: Chapman and Hall.
- Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 1960;20(1):37-46.
- Light RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychological Bulletin. 1971;76(5):365-377.
- Takala EP et al. 2010. Systematic evaluation of observational methods assessing biomechanical exposures at work. Scand J Work Environ Health. 36(1):3-24.