



## The inter-rater reliability of Strain Index and OCRA Checklist task assessments in cheese processing



Robert Paulsen<sup>a</sup>, Tommaso Gallu<sup>b</sup>, David Gilkey<sup>a</sup>, Raoul Reiser II<sup>c</sup>, Lelia Murgia<sup>b</sup>, John Rosecrance<sup>a,\*</sup>

<sup>a</sup> Colorado State University, Environmental and Radiological Health Sciences Department, 1681 Campus Delivery, Fort Collins, CO 80523, USA

<sup>b</sup> Università degli Studi di Sassari, Dipartimento di Agraria, Viale Italia, 39, 07100 Sassari, Italy

<sup>c</sup> Colorado State University, Department of Health and Exercise Science, 220 Moby B Complex, Fort Collins, CO 80523, USA

### ARTICLE INFO

#### Article history:

Received 21 May 2014

Accepted 27 April 2015

Available online 5 June 2015

#### Keywords:

Exposure assessment

Inter-rater reliability

Upper extremity musculoskeletal disorders

### ABSTRACT

The purpose of this study was to characterize the inter-rater reliability of two physical exposure assessment methods of the upper extremity, the Strain Index (SI) and Occupational Repetitive Actions (OCRA) Checklist. These methods are commonly used in occupational health studies and by occupational health practitioners. Seven raters used the SI and OCRA Checklist to assess task-level physical exposures to the upper extremity of workers performing 21 cheese manufacturing tasks. Inter-rater reliability was characterized using a single-measure, agreement-based intraclass correlation coefficient (ICC). Inter-rater reliability of SI assessments was moderate to good (ICC = 0.59, 95% CI: 0.45–0.73), a similar finding to prior studies. Inter-rater reliability of OCRA Checklist assessments was excellent (ICC = 0.80, 95% CI: 0.70–0.89). Task complexity had a small, but non-significant, effect on inter-rater reliability SI and OCRA Checklist scores. Both the SI and OCRA Checklist assessments possess adequate inter-rater reliability for the purposes of occupational health research and practice. The OCRA Checklist inter-rater reliability scores were among the highest reported in the literature for semi-quantitative physical exposure assessment tools of the upper extremity. The OCRA Checklist however, required more training time and time to conduct the risk assessments compared to the SI.

© 2015 Elsevier Ltd and The Ergonomics Society. All rights reserved.

### 1. Introduction

There are many semi-quantitative and observational physical risk assessment tools available to occupational health researchers and practitioners (Takala et al., 2010). The analyst should select a method based on the practicality, validity, reliability, and purpose of the risk assessment (David, 2005; Kilbom, 1994; Li and Buckle, 1999; Takala et al., 2010). Most semi-quantitative and observational tools are practical alternatives to more intensive (in terms of cost, labor, and time) biomechanical instrumentation (*i.e.*, direct measures) (David, 2005; Kilbom, 1994). The most comprehensive semi-quantitative assessments aim to precisely quantify worker exposure to physical musculoskeletal disorder (MSD) risk factors,

such as the Strain Index (SI), the Occupational Repetitive Actions (OCRA) methods, or the National Institute for Occupational Safety and Health (NIOSH) Lifting Equation. However, a common limitation with most of these tools is that their underlying models assume that the job exposures under evaluation result from simple task completion (Garg and Kapellusch, 2011).

Simple tasks are typified by repetitive, single-exertion work activity performed with limited postural variability. Yet, most repetitive industrial tasks are complex, *i.e.*, they are comprised of subtasks (or work elements) which vary in their force, frequency and postural demands (Bao et al., 2009b). Several authors have reported the validity and reliability of risk assessments completed using semi-quantitative measures (Bonfiglioli et al., 2013; Garg et al., 2012; Occhipinti and Colombini, 2007; Spielholz et al., 2008; Stephens et al., 2006; Stevens et al., 2004; Waters et al., 1998), but these studies have not characterized how task-complexity affects assessment results. Additionally, the inter-rater reliability of OCRA Checklist assessments has not been characterized using standard statistical techniques. The purpose of the

\* Corresponding author. Tel.: +1 970 491 1405; fax: +1 970 491 2940.

E-mail addresses: [rob.paulsen@rams.colostate.edu](mailto:rob.paulsen@rams.colostate.edu) (R. Paulsen), [thomasgul@tiscali.it](mailto:thomasgul@tiscali.it) (T. Gallu), [david.gilkey@colostate.edu](mailto:david.gilkey@colostate.edu) (D. Gilkey), [raoul.reiser@colostate.edu](mailto:raoul.reiser@colostate.edu) (R. Reiser), [diti\\_mecc@uniss.it](mailto:diti_mecc@uniss.it) (L. Murgia), [john.rosecrance@colostate.edu](mailto:john.rosecrance@colostate.edu) (J. Rosecrance).

present study was to characterize the inter-rater reliability of SI and OCRA Checklist assessments of simple and complex cheese production tasks. The present study included these two exposure assessment tools because they comprehensively model physical exposure to upper extremity (UE) MSD risks differently, and occupational health professionals across the globe advocate for their use (Kapellusch et al., 2013; Occhipinti and Colombini, 2012). Additionally, the choice of the SI and OCRA Checklist for use in the present study was based on the type of work assessed, the research questions addressed and the author's experience with these assessment tools.

## 2. Material and methods

### 2.1. Study tasks

Physical exposure data for 21 cyclic work tasks were collected at a cheese production facility in Sardinia, Italy. These tasks represented all of the major stages of Pecorino Romano production, which yields a hard cheese in 25 and 35 kg wheels (Fig. 1). Researchers video-recorded the 21 tasks using handheld digital camcorders focused on the UE in the sagittal and frontal planes of the worker. Video recordings captured a minimum of five work cycles during normally scheduled work periods while workers were paid their usual wage. Task duration and break/recovery times were collected by direct observation and interviews with the facility management. The management and owners of the processing facility agreed to the study procedures and all participating employees provided informed consent. No personal (other than hand dominance) or identifying information was collected and no individual participant (employee) data was provided or available to the employer.

Tasks were predominantly externally-paced (by conveyor or demand from the next process) and represented a spectrum of repetitive UE activity, with work cycle times ranging from 6 to 106 s (mean = 41.5, SD = 31.2). The complexity of the cheese production tasks varied; six work cycles were comprised of a single subtask (mono-element), nine were comprised of two subtasks (dual-element), and six were comprised of three subtasks (tri-element).

### 2.2. Raters

Seven members of occupational health research groups from the University of Sassari in Italy and Colorado State University in the



Fig. 1. A worker tightens a mold around a 25 kg Pecorino Romano cheese wheel.

United States were recruited to assess the 21 cheese production tasks using the SI and OCRA Checklist separately. These ergonomics analysts (hereafter referred to as raters) included three university faculty performing occupational health research and four graduate students specializing in occupational ergonomics. Two of the raters were experienced SI users (using the tool for more than one year in manufacturing settings) and another researcher was an experienced OCRA Checklist rater. Two of the three participating faculty members were Board Certified Professional Ergonomists.

The majority of raters were novice users of the methods and all raters underwent appropriate training prior to task assessment. Strain Index training was administered separately from OCRA Checklist training. Training sessions included didactic instruction on the principles and procedures of each method, practice applying the methods to video segments of manufacturing tasks, and feedback from an experienced rater regarding method application. Training sessions continued until trainees achieved competency. Competency for each method was reached when trainees consistently (80% of time) assigned exposure ratings that were similar (within 20%) to the most experienced rater. The OCRA Checklist training required 10 h while the SI training required 4 h.

Once trained, raters were provided digital copies of the 21 video-recorded tasks and electronic SI and OCRA Checklist worksheets. The SI worksheet was based on Moore and Garg's (1995) original procedures. The OCRA Checklist worksheet was based on Colombini's et al. (2011) update of the method.

### 2.3. Exposure assessment

Raters assessed task exposures for the worker's left and right UEs, and tasks were assessed in alphabetic order according to task name. Task names had no inherent relationship to any of the physical exposure parameters, work cycle time or task complexity. Three raters completed the SI first and the other four raters completed the OCRA Checklist first. Each rater performed SI and OCRA Checklist assessments separately, and they did not communicate the results with one another. Additionally, after completing the initial 21 task assessments, raters did not have access to that data while reevaluating tasks with the second method. After completing each job assessment, raters were instructed to report time spent evaluating that job. This data was summarized for both the OCRA and SI after the study was finished.

Raters assigned scores to physical exposure parameters for the individual SI task variables *intensity of exertion*, *duration of exertion*, *efforts per minute*, *hand/wrist posture*, and *speed of work*; and for the individual OCRA Checklist task variables *force*, *frequency*, *awkward postures/movements*, and *additional factors*. To ensure consistency between force/exertion intensity estimates with each method, raters applied the Borg CR-10 scale (Borg, 1982) themselves rather than requesting (as the OCRA methods instruct) self-reported force estimates from workers. All estimates of forceful exertions were based on video observation of tasks and user motions/facial expressions as no direct measures were included in this study. Previous research does not indicate that expert-estimated force exposures are less accurate than workers' self-reported estimates (Bao et al., 2006; Spielholz et al., 2001). Data for SI *task duration per day* variable and the OCRA Checklist *lack of sufficient recovery* and *task duration* variables were provided to raters and all were scored the same.

The SI and OCRA Checklist risk classification cut points are depicted in Table 1. The SI risk index is normally classified along a three-level scale (Garg and Kapellusch, 2011), and the OCRA Checklist risk index is normally classified along a five-level scale. To facilitate comparison between the two tools, the cut points for the OCRA Checklist were condensed: level one (OCRA Checklist scores

**Table 1**  
Strain Index (SI) and Occupational Repetitive Actions (OCRA) Checklist risk classification/cut-off criteria.

Risk index classification criteria		
SI	OCRA Checklist	
<3	Safe	<7.6
3–6.9	Caution	7.6–14
≥7	Hazardous	≥14.1

≤7.5) formed the “safe” classification; levels two and three (scores 7.6–11.0 and 11.1–14.0) were combined to form the “Caution” classification; and levels four and five (15.1–22.5 and ≥ 22.6) were combined to form the “Hazardous” classification. These cut-point combinations were structured similarly to previous SI and OCRA method studies (Apostoli et al., 2004; Chiasson et al., 2012; Jones and Kumar, 2010), but the present study used the most recent versions of both assessment tools (Colombini et al., 2011; Moore and Garg, 1995) whereas previous researchers referred to the original OCRA Checklist or OCRA Index classification criteria (Colombini et al., 2002).

The present study referred to Bao's et al. (2009b) definitions of job, task, and work elements or subtasks to best describe the complexity of tasks assessed. According to that definition, an individual performs a job that is comprised of one or more tasks, and these tasks are comprised of one or more work elements or subtasks (Bao et al. 2009b). However, the consensus method of performing SI assessments of complex tasks (those tasks comprised of more than one subtask) required that raters assign a single task-variable score at the task level (Moore and Garg, 1995; Bao et al. 2009b; Garg and Kapellusch, 2011). For example, the rater selects the “most appropriate” SI score for the *intensity of exertion* task-variable based on the worker's “overall” exposure to forceful exertions during the complex task (Garg and Kapellusch, 2011). To minimize the between-rater variation in task-variable scoring, it is important that raters have familiarity with the biomechanical stresses to the body during work task performance. In the present study, raters were instructed to score each task variable for both the SI and OCRA Checklist based on the worker's overall risk factor exposure as described by Garg and Kapellusch (2011).

#### 2.4. Analysis

In ten of the 21 cheese processing tasks, the left and right UE activity was nearly identical during the complete task cycle (i.e., the task exposures were symmetric). For these ten tasks, only the exposure assessments of the dominant limb were included in the inter-rater reliability analyses. For the other 11 tasks (i.e., the asymmetric task exposures), both left and right limb exposures were included. This analytical approach increased the total task exposure sample size from 21 to 32 without violating the assumption that UE exposure assessments were independent from one another.

Inter-rater reliability was characterized using a single-measure, agreement-based intraclass correlation coefficient (ICC), termed the ICC (2,1) by Shrout and Fleiss (1979). This is the most common ICC measure for rater reliability studies of semi-quantitative and observational ergonomic methods (Bao et al., 2009a; Dockrell et al., 2012; Ebersole and Armstrong, 2002; Stephens et al., 2006; Stevens et al., 2004; Xu et al., 2011). A weighted kappa statistic could have been used to characterize inter-rater agreement, but the ICC (2,1) analyses produce coefficients equivalent to weighted Kappa using Fleiss-Cohen weights (Shrout and Fleiss, 1979). The ICC interpretations were made according to the verbal criteria suggested by Stevens et al. (2004):  $\rho < 0.40$ , poor reliability;  $0.40 \leq \rho \leq 0.75$ ,

moderate to good reliability; and  $\rho > 0.75$ , excellent reliability. The ICCs were calculated using SPSS software (IBM Corp., Armonk, NY) version 20.0 (2011). All confidence intervals were calculated with 95% certainty.

### 3. Results

Seven raters assessed 32 UE exposures using both the SI and OCRA Checklist, resulting in a total of 224 pairs of assessments. Of the 224 task exposures, 49.1% were classified as hazardous according to the OCRA Checklist and 60.2% were classified as hazardous according to the SI. Raters spent a mean 29.6 min (SD = 11.4) conducting each OCRA Checklist evaluation compared to mean 7.1 (SD = 2.9) minutes conducting each SI evaluation. Intraclass correlation coefficient measures for risk indexes, risk classifications, and individual task-variable scores are depicted in Table 2. The ICCs for the SI and OCRA Checklist risk indexes were 0.59 (95% CI: 0.45–0.73) and 0.80 (95% CI: 0.70–0.89), respectively. The three-level classification of risk indexes reduced the ICC-values associated with both measures, but the difference was insignificant. Of the individual task variables contributing to the total SI or OCRA Checklist exposure scores, the lowest inter-rater reliability was associated with the SI *hand/wrist posture* ( $\rho = 0.16$ ) and with the OCRA Checklist *additional factors* ( $\rho = 0.21$ ). The individual variables with the highest inter-rater reliability were *efforts per minute* (frequency) for the SI ( $\rho = 0.60$ ) and *frequency of technical action* for the OCRA Checklist ( $\rho = 0.68$ ).

The inter-rater reliability for the overall SI and OCRA Checklist risk indexes was similar across task complexity groupings (Table 3). The SI task variable *duration of exertion* was higher for mono-element tasks ( $\rho = 0.74$ ) compared to dual-element ( $\rho = 0.24$ ) and tri-element tasks ( $\rho = 0.31$ ). In addition, the OCRA Checklist task variable *force exertion* was associated with lower reliability for mono-element tasks ( $\rho = 0.15$ ) than for dual-element ( $\rho = 0.45$ ) and tri-element tasks ( $\rho = 0.61$ ). None of these differences in ICC values were statistically significant at the 0.05-level given the confidence interval overlap. The lack of statistical difference was likely related to the small sample size of task exposures within each complexity grouping ( $n = 7$  for mono-element,  $n = 14$  for dual-element, and  $n = 11$  for tri-element) and therefore limited statistical power.

### 4. Discussion

Inter-rater reliability was characterized for two physical exposure assessments, the SI and the OCRA Checklist. The inter-rater reliability of SI assessments determined in the present study was similar to prior studies, and reliability of OCRA Checklist assessments was higher than SI assessments. Additionally, task complexity did not appear to significantly affect the inter-rater reliability of either assessment, although the power to compare results between task complexity groupings was low.

Across all task exposures, the inter-rater reliability of the risk index was moderate to good for the SI ( $\rho = 0.59$ ; 95% CI: 0.45–0.73) and excellent for the OCRA Checklist ( $\rho = 0.80$ ; 95% CI: 0.70–0.89). The difference in the assessment comprehensiveness and scoring scale complexity of each method may explain the difference in reliability coefficients. With regard to scale complexity, the SI task variables have a maximum of five possible scores whereas the OCRA Checklist task variable scales allow for dozens of possible scores. For example, there are five possible scores for the SI *exertions per minute* variable compared to 21 possible scores for *frequency of technical action* for the OCRA Checklist.

As for comprehensiveness, both methods do assess awkward posture exposures with the use of biomechanically-based posture criteria. Force intensity is assessed using similar verbal descriptors

**Table 2**  
Distributions and intraclass correlation coefficient (ICC) calculations for Strain Index (SI) and Occupational Repetitive Actions (OCRA) Checklist risk indexes, risk classifications, and task-variable scores.

Task variables and risk ratings	Mean	Standard deviation	ICC, $\rho$ (95% CI <sup>a</sup> )
SI risk classification	2.5	0.7	0.54 (0.40–0.70)
SI risk index	25.6	30.7	0.59 (0.45–0.73)
Intensity of exertion	4.0	2.1	0.39 (0.24–0.56)
Duration of exertion	1.6	0.7	0.40 (0.25–0.57)
Efforts per minute	2.2	0.8	0.60 (0.46–0.74)
Hand/wrist posture	1.6	0.5	0.16 (0.06–0.31)
Speed of work	1.1	0.2	0.30 (0.18–0.47)
OCRA Checklist risk classification	2.3	0.8	0.80 (0.70–0.89)
OCRA Checklist risk index	16.8	10.6	0.68 (0.56–0.80)
Frequency of technical actions	4.2	3.7	0.68 (0.54–0.81)
Force exertion	6.7	7.7	0.42 (0.28–0.59)
Awkward posture/movement	4.9	3.4	0.54 (0.39–0.69)
Additional factors	1.5	1.2	0.21 (0.10–0.37)

<sup>a</sup> CI, Confidence Interval.

and a Borg CR-10 scale analogue, and the assessment of force exposure duration is based on a duty-cycle percentage. However, the SI and OCRA Checklist utilize slightly different constructs for quantifying repetitive UE activity, UE posture, and work–rest cycles. In addition, unlike the SI, the OCRA Checklist was designed to assess exposures at the shoulder as well as additional MSD risk factors, such as hand–arm vibration, contact/impact stress, and cold work environments. Consequently, the comprehensive nature of the OCRA Checklist and the greater complexity of its scoring scale increase the range of possible assessment ratings, thereby influencing measures of methodological reliability. That is, a more comprehensive scale increases the probability that the variance in assessment results is due to task variance rather than rater variance (Streiner and Norman, 2008). In effect, even if the same percentage of raters agreed in their assessment scores, the OCRA Checklist results would be characterized by higher reliability than would the SI results. As can be expected with any measurement tool, the more comprehensive and complex, the more time is required for training (10 h compared to 4 h for the SI) and evaluating each job task (nearly 30 min per task for the OCRA compared to a mean 7 min per task for the SI).

Overall, the OCRA Checklist risk index, risk classification, and individual task variables were characterized by higher inter-rater reliability than compared to the SI variables and risk scores. At

least moderate inter-rater reliability ( $\rho \geq 0.40$ ) was associated with the SI task variables *efforts per minute* and *speed of work* and the OCRA Checklist task variables *frequency of technical actions*, *force exertion*, and *awkward postures/movements*. However, the majority of individual task variables were characterized by poor reliability lower confidence limits ( $0.06 < \rho \text{ LCL} < 0.39$ ). The lower confidence limits were not surprising given the small number of independent UE exposures ( $n = 32$ ) and the fact that the majority of tasks had high UE risk factor exposures.

Additionally, task-variable scores would be characterized by low ICCs if there were low variability in scores (regardless of rater agreement). The ICC statistic is biased toward zero when the variability in the scale (*i.e.*, the possible scores) is low (Streiner and Norman, 2008). For example, the frequency of scores equal to 1.0 for the SI task-variable *wrist/hand posture* was over 80%. This high proportion of agreement is in part due to the design of the SI posture scale; three out of the five possible posture exposure levels yield a score of 1.0. Had the present analyses used the five-level posture categories rather than the posture scores for ICC calculations, higher reliability scores might have been observed. Reliability could be improved by increasing the distribution of the scoring scales for those variables with weaker reliability coefficients (*e.g.* OCRA *additional factors* and SI *speed of work and posture additional factors*). This is worth considering for any future revisions to these tools.

Alternatively, the variability in scores would likely have increased (along with the inter-rater reliability) had the tasks selected for evaluation exposed workers to a greater range of hand and wrist postures. For example, Stevens et al. (2004) used the ICC(2,1) to characterize the inter-rater reliability of SI risk assessments. In their study, fifteen raters evaluated 73 mono-element task exposures that were purposefully chosen to exhibit a full and balanced range of physical work exposures. The SI task-variable ICCs ranged from 0.77 to 0.81 ( $\rho \text{ LCLs} \geq 0.63$ ) for all but the *hand/wrist posture* variable, which was still associated with moderate to good reliability ( $\rho = 0.66$ ,  $\text{LCL} = 0.45$ ). The task-variable ICCs from the present study would likely improve if tasks were purposefully chosen for assessment as opposed to being a random sample of exposures in the food manufacturing industry. Nonetheless, the SI risk index ICC determined in the present study ( $\rho = 0.59$ ; 95% CI: 0.45–0.73) was similar to what Stevens et al. reported ( $\rho = 0.43$ , 95% CI: 0.25–0.70).

The inter-rater reliability for SI and OCRA Checklist assessments were similar across simple and complex tasks. Some task-variables

**Table 3**  
Intraclass correlation coefficient (ICC) calculations, stratified by task complexity, for Strain Index (SI) and Occupational Repetitive Actions (OCRA) Checklist risk indexes, risk classifications, and task-variable scores.

Task variables and risk scores	Mono-element task exposures ( $n^a = 7$ )	Dual-element task exposures ( $n = 14$ )	Tri-element task exposures ( $n = 11$ )
	ICC, $\rho$ (95% CI <sup>b</sup> )	ICC, $\rho$ (95% CI)	ICC, $\rho$ (95% CI)
SI risk classification	0.59 (0.30–0.89)	0.42 (0.22–0.69)	0.39 (0.17–0.70)
SI risk index	0.69 (0.43–0.92)	0.58 (0.37–0.80)	0.55 (0.32–0.81)
Intensity of exertion	0.43 (0.18–0.79)	0.41 (0.21–0.67)	0.40 (0.18–0.70)
Duration of exertion	0.74 (0.49–0.94)	0.24 (0.09–0.51)	0.31 (0.12–0.63)
Efforts per minute	0.63 (0.35–0.90)	0.43 (0.23–0.69)	0.62 (0.38–0.84)
Hand/wrist posture	0.18 (0.00–0.63)	0.09 (0.00–0.30)	0.23 (0.07–0.54)
Speed of work	0.09 (0.00–0.51)	0.29 (0.12–0.57)	0.27 (0.08–0.59)
OCRA risk classification	0.50 (0.22–0.85)	0.71 (0.53–0.87)	0.72 (0.51–0.89)
OCRA risk index	0.76 (0.52–0.94)	0.84 (0.71–0.93)	0.75 (0.55–0.91)
Frequency of technical actions	0.74 (0.48–0.94)	0.64 (0.44–0.84)	0.69 (0.46–0.89)
Force exertion	0.15 (0.00–0.58)	0.45 (0.25–0.71)	0.61 (0.38–0.84)
Awkward posture/movement	0.76 (0.52–0.94)	0.39 (0.20–0.66)	0.38 (0.17–0.69)
Additional factors	0.37 (0.13–0.77)	0.15 (0.02–0.39)	0.20 (0.05–0.51)

<sup>a</sup> n, number of upper extremity task exposures per job-complexity grouping.

<sup>b</sup> CI, confidence interval.

(e.g., SI duration of exertion and OCRA Checklist force exertion) appeared to differ between mono-element and dual- or tri-element tasks. These differences were not significant given the small number of each task complexity grouping. However, the inter-rater reliability of semi-quantitative assessments may be higher when ratings are determined using video analysis rather than onsite in real time. For instance, the Spielholz's et al. (2008) characterized the inter-rater reliability of 125 task assessments (most of which were not mono-element tasks) analyzed by three expert raters and one novice rater. Tasks were rated in real time onsite without video. The inter-rater reliability the SI risk index and individual task-variable scores was characterized as poor to moderate using kappa coefficients ( $0.22 \leq \kappa \leq 0.44$ ). Semi-quantitative exposure assessment would likely be less reliable between raters when observations are made in real time and without the aid of video analysis, which can be replayed numerous times and slowed down.

The present study was the first to investigate inter-rater reliability of the OCRA Checklist and the first to compare the inter-rater reliability between the OCRA Checklist and SI. The 21 tasks (32 independent UE exposures) that raters assessed were similar to the type of motions, levels of forcefulness and repetition common to light to moderately intensive hand/wrist work often seen in manufacturing. Generalizability of these results, however, may be limited to the moderately hand intensive food processing type tasks as described in this study.

The use of risk assessment tools for applied research or for workplace practice should be based on the purpose of the evaluation, time required to conduct risk assessment, level of assessment detail required, the validity and reliability of the measure as well as the evaluator's experience with the measures. For example, if the purpose is to conduct a risk assessment involving the shoulder, elbow and wrist, the OCRA Checklist would be more appropriate than the Strain Index. However, if the specific question was focused on the risk of carpal tunnel syndrome, an assessment of the wrist/hand with the SI would be appropriate. With regards to time resources, training sessions for the OCRA Checklist were more than twice as long as the SI and the time to conduct a risk assessment with the OCRA Checklist was four times that of the SI.

## 5. Conclusion

Strain Index and OCRA Checklist risk indexes and risk classifications were characterized by moderate to excellent inter-rater reliability, and the reliability of individual task-variable scores was poor to moderate. In general, the OCRA Checklist was characterized by higher inter-rater reliability than the SI. If the inter-rater reliability of exposure assessment is a concern, researchers and occupational health practitioners should choose to use average or consensus-based rating strategies. The inter-rater reliability of the SI and OCRA Checklist was similar across simple and complex tasks, but the power to detect statistical differences at the work-element or subtask level was low. The inter-rater reliability for both measures was at least moderately good. The choice between the OCRA Checklist and SI for risk assessment should be based on several issues but primarily on the researchers or practitioner's specific question of interest.

## Acknowledgments

The National Institute for Occupational Safety & Health (NIOSH) provided partial support for this study through the Mountain and Plains Education and Research Center (grant number: T42OH009229-09). Its contents are solely the responsibility of the

authors and do not necessarily represent the official views of the NIOSH.

## References

- Apostoli, P., Sala, E., Gullino, A., Romano, C., 2004. Comparative analysis of the use of 4 methods in the evaluation of the biomechanical risk to the upper limb. *G. Ital. Med. del Lav. Ergon.* 26, 223–241.
- Bao, S., Howard, N., Spielholz, P., Silverstein, B., 2006. Quantifying repetitive hand activity for epidemiological research on musculoskeletal disorders – part II: comparison of different methods of measuring force level and repetitiveness. *Ergonomics* 49, 381–392.
- Bao, S., Howard, N., Spielholz, P., Silverstein, B., Polissar, N., 2009a. Interrater reliability of posture observations. *Hum. Factors* 51, 292–309.
- Bao, S., Spielholz, P., Howard, N., Silverstein, B., 2009b. Application of the Strain Index in multiple task jobs. *Appl. Ergon.* 40, 56–68.
- Bonfiglioli, R., Mattioli, S., Armstrong, T.J., Graziosi, F., Marinelli, F., Farioli, A., Violante, F.S., 2013. Validation of the ACGIH TLV for hand activity level in the OCTOPUS cohort: a two-year longitudinal study of carpal tunnel syndrome. *Scand. J. Work Env. Health* 39, 155–163.
- Borg, G.A., 1982. Psychophysical bases of perceived exertion. *Med. Sci. Sp. Exerc.* 14, 377–381.
- Chiasson, M.E., Imbeau, D., Aubry, K., Delisle, A., 2012. Comparing the results of eight methods used to evaluate risk factors associated with musculoskeletal disorders. *Int. J. Ind. Ergon.* 42, 478–488.
- Colombini, D., Occhipinti, E., Grieco, A., 2002. Risk Assessment and Management of Repetitive Movements and Exertions of Upper Limbs Job Analysis, OCRA Risk Indices, Prevention Strategies and Design Principles. Elsevier, Oxford.
- Colombini, D., Occhipinti, E., Cerbai, M., Battevi, N., Placci, M., 2011. Updating of application procedures and criteria for OCRA Checklist. *Med. Del Lav.* 102, 1–39.
- David, G.C., 2005. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occup. Med.* 55, 190–199.
- Dockrell, S., O'Grady, E., Bennett, K., Mullarkey, C., Mc Connell, R., Ruddy, R., Twomey, S., Flannery, C., 2012. An investigation of the reliability of rapid upper limb assessment (RULA) as a method of assessment of children's computing posture. *Appl. Ergon.* 43, 632–636.
- Ebersole, M.L., Armstrong, T.J., 2002. Inter-rater reliability for hand activity level (HAL) and force metrics. *Human Factors and Ergonomics Society 46th Annual Meeting, Baltimore, MD*, pp. 1037–1040.
- Garg, A., Kapellusch, J., 2011. Job analysis techniques for distal upper extremity disorders. *Rev. Hum. Factors Ergon.* 7, 149–196.
- Garg, A., Kapellusch, J., Hegmann, K., Wertsch, J., Merryweather, A., Deckow-Schaefer, G., Malloy, E.J., Team, W.H.S.R., 2012. The Strain Index (SI) and threshold limit value (TLV) for hand activity level (HAL): risk of carpal tunnel syndrome (CTS) in a prospective cohort. *Ergonomics* 55, 396–414.
- Jones, T., Kumar, S., 2010. Comparison of ergonomic risk assessment output in four sawmill jobs. *Int. J. Occup. Saf. Ergon.* 16, 105–111.
- Kapellusch, J.M., Garg, A., Bao, S.S., Silverstein, B.A., Burt, S.E., Dale, A.M., Evanoff, B.A., Gerr, F.E., Harris-Adamson, C., Hegmann, K.T., Merlino, L.A., Rempel, D.M., 2013. Pooling job physical exposure data from multiple independent studies in a consortium study of carpal tunnel syndrome. *Ergonomics* 56, 1021–1037.
- Kilbom, A., 1994. Assessment of physical exposure in relation to work-related musculoskeletal disorders – what information can be obtained from systematic observations. *Scand. J. Work Env. Health* 20, 30–45.
- Li, G., Buckle, P., 1999. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics* 42, 674–695.
- Moore, J.S., Garg, A., 1995. The Strain Index – a proposed method to analyze jobs for risk of distal upper extremity disorders. *Am. Ind. Hyg. Assoc. J.* 56, 443–458.
- Occhipinti, E., Colombini, D., 2007. Updating reference values and predictive models of the OCRA method in the risk assessment of work-related musculoskeletal disorders of the upper limbs. *Ergonomics* 50, 1727–1739.
- Occhipinti, E., Colombini, D., 2012. IEA/WHO toolkit for WMSDs prevention: criteria and practical tools for a step by step approach. *Work – J. Prev. Assess. Rehabil.* 41, 3937–3944.
- Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations – uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428.
- Spielholz, P., Bao, S., Howard, N., Silverstein, B., Fan, J., Smith, C., Salazar, C., 2008. Reliability and validity assessment of the hand activity level threshold limit value and Strain Index using expert ratings of mono-task jobs. *J. Occup. Environ. Hyg.* 5, 250–257.
- Spielholz, P., Silverstein, B., Morgan, M., Checkoway, H., Kaufman, J., 2001. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics* 44, 588–613.
- Stephens, J.P., Vos, G.A., Stevens, E.M., Moore, J.S., 2006. Test-retest repeatability of the Strain Index. *Appl. Ergon.* 37, 275–281.
- Stevens, E.M., Vos, G.A., Stephens, J.P., Moore, J.S., 2004. Inter-rater reliability of the Strain Index. *J. Occup. Environ. Hyg.* 1, 745–751.
- Streiner, D.L., Norman, G.R., 2008. *Health Measurement Scales: a Practical Guide to Their Development and Use*. Oxford University Press, Oxford.

- Takala, E.P., Pehkonen, I., Forsman, M., Hansson, G.A., Mathiassen, S.E., Neumann, W.P., Sjogaard, G., Veiersted, K.B., Westgaard, R.H., Winkel, J., 2010. Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scand. J. Work, Env. Health* 36, 3–24.
- Waters, T.R., Baron, S.L., Kemmlert, K., 1998. Accuracy of measurements for the revised NIOSH lifting equation. *Appl. Ergon.* 29 (6), 433–438.
- Xu, X., Chang, C.C., Faber, G.S., Kingma, I., Dennerlein, J.T., 2011. The validity and interrater reliability of video-based posture observation during asymmetric lifting tasks. *Hum. Factors* 53, 371–382.