Investigating Intra-Examiner and Inter-Examiner Reliability of Three Upper-Limb Risk Assessment Methods

Majid Motamedzade[1], Mahdi Mohammadiyan*[2], Javad Faradmal[3]

1) Center of Excellence for Occupational Health, Research Center for Health Science, School of Public Health, Hamadan University of Medical Sciences, Hamadan, Iran
2) Department of Occupational Health Engineering, School of Health, Semnan University of Medical Sciences, Semnan, Iran
3) Department of Biostatistics and Epidemiology, School of Public Health and Modeling of Non-communicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran

* *Author for Correspondence*: mohammadiyan.mahdi@yahoo.com

Received: 17 Sept. 2017, Revised:14 Apr. 2018, Accepted: 17 Jun. 2018

ABSTRACT
There are several methods for assessing the ergonomics risk of upper-limb disorders in occupational health studies. The main objective of the present study is to determine the intra and inter-examiner reliability of three risk assessment methods of SI, OCRA Checklist and ART in repetitive tasks. Early, nine examiners used the three methods to assess the physical exposure level of the upper limb in workers 30 tasks with different levels of the profession for two assessment periods. In the results obtained using ICC for inter-examiner reliability, the risk categorizations were ρ: 0.54 to 0.77 for SI, OCRA Checklist and ART, Also the intra-examiner reliability for the final score of the three methods was κ: 0.59 to 0.67. The reliability for risk categorizations and the final score of this method using ICC obtained as moderate to good. However, the overall results showed that the reliability of the ART method was higher than the two other methods. The findings demonstrated that all three methods are appropriate for assessment in workplaces. These methods should be used based on assessment objectives, examiner skill and required duration time for assessing the tasks.

Keyword: Musculoskeletal Disorders, OCRA Checklist, Strain Index, ART

List of Abbreviations
MSDs: Musculoskeletal Disorders
ULMSDs: Upper-Limb Musculoskeletal Disorders
OCRA: Occupational Repetitive Action
SI: Strain Index
ART: Assessment of Repetitive Tasks
ICC: intra-class correlation coefficient

INTRODUCTION
Nowadays, upper-limb musculoskeletal disorders (ULMSDs) are prevalent among workers in different industries, especially for repetitive and assembling occupations [1]. ULMSDs are also known as major factors which are resulted from working and reduced production in developed countries. In previous studies, the repetitive activities and risk factors associated with ULMSDs (such as excessive force, improper postures, vibration, and psychosocial-organizational factors) [2] are reported as the cause of carpal tunnel syndrome, hand and forearm muscle cramps etc.[3]. Therefore, ergonomics main challenge is to design the work environment with the purpose to prevent the ULMSDs from it with no negative effects on products production and quality [4. According to the health procedures and suggestions, the risk assessment should be performed in a way to prevent the work environment harmful ergonomic factors [5].
For assessing the risk of musculoskeletal disorders, approximately 30 observational assessment methods have been developed to identify and quantify the physical occupational exposures [6]. The pen and paper observational methods are often used to evaluate posture and based on their simplicity and lack of need for special equipment, these methods are the proper replacement for direct measurement methods [7]. They can also be used in a wide range of work environments [8]. For the observational methods, different physical exposures such as manual loading and repetitive tasks in the upper limb are formally applicable. These methods should be selected based on the need for assessment and type of exposure. Strain Index (SI), Assessment of Repetitive Tasks (ART), and Occupational Repetitive Action (OCRA) is among the most important assessment methods that are

widely used for physical exposure of upper limbs [6-9]. In the earlier studies, the reliability of the ART method is not specified by using the standard statistical methods.

The overall study objective was to determine the inter-examiner reliability of ART, OCRA Checklist, and SI methods by dissimilar examiners as well as the intra-examiner reliability of these methods by the same examiners. The SI, OCRA Checklist, and ART methods are based on the type of evaluated occupations and research questions.

MATERIALS AND METHODS

In this study, data were collected from 30 occupational tasks in industries such as poultry slaughtering, assembling and manufacturing aluminium containers. The researchers recorded 30 videos of tasks by a digital camera and with focus on the upper limb on the sagittal and frontal planes of the workers. The videos contained at least 5 working cycles during the worker's tasks. Data related to the duration of tasks, breaks and work cessations were directly collected via observation and interviews with the factory manager. Most of the tasks were repetitive upper-limb activities with working cycle with 4-32 sec time intervals.

In sum, 9 raters with the background of at least 2 years of ergonomic risk assessment participated in this study. Three of them were men and 6 were women who were graduated students of occupational health and ergonomics. Because of different expertise levels of the raters participating in this study, the training courses for three methods of Strain Index, OCRA Checklist and ART were held separately for them. The training sections were developed based on the principles and procedures of each method with practical exercises. The training continued until the raters reached the desired competency level. Next, 30 Microsoft Excel files were prepared and distributed between them. The SI worksheet was developed based on Moor and Gray's [10] main procedure, OCRA Checklist worksheet was developed based on the updated method by Colombini *et al.* [11] and ART worksheet was based on Ferreira *et al.* [12] procedure. The assessments were separately performed for the activity of the right and left upper-limb was asymmetric in the 30 tasks of the workers. The results were assessed separately for each hand. In total, 60 assessments were carried out by the examiner for the three methods of OCRA Checklist, SI and ART individually and without communication or access to the results of others. After finishing the assessments, the results were sent to the research team in the form of digital files. The examiner performed their assessments in two periods and four weeks apart. Both evaluation processes performed on the selection of similar videos. The assessed task variables belonging to the SI method were as follows: intensity of exertion, efforts per minute, hand/wrist postures, duration of exertion and speed of work; the task variables of OCRA Checklist were: awkward posture/movement status frequency, force exertion, lack of recovery time, work duration and additional factors; and task variables of the ART method were: force, frequency/repetition movement, additional factors and awkward posture. The Borg-CR-10 scale [13] was used for the homogeneity and coordination in approximating the intensity of efforts in the methods. Thus, all the estimates of exertion intensity were performed based on observing the videos of tasks and changes in the facial expression of the workers without any direct measurement. Previous studies showed that the estimation of the applied force by experts is more accurate than the self-reporting estimation by the workers [14]. The data of task duration variable in the daytime and lack of retrieval time were given to the examiners and all the points considered for these variables were similar. The risk categorizations criteria have three levels for SI [15] and ART [12] methods. In order to compare these instruments with the OCRA Checklist method, the cut-off of this method was condensed and moderated from 5 to 3 levels. In Table 1, the risk categorizations criteria (risk-level cut-off) in this study are mentioned, which is similar to the studies on other methods (SI, OCRA Checklist, and ART) [16].

Several statistical methods have been conducted to calculate reliability in other studies. Intra-examiner reliability was considered based on the outcomes of 60 first assessments. Inter-examiner reliability is determined based on the first and second scores of 9 examiners. A statistical test which is often used for reliability is Cohen's Kappa coefficient [17]. One of the problems with this method is its application for class data; Kappa only considers the overall agreement. In order to solve this problem, weighted Kappa, which is used for ordinal data, was used in this study. Other conventional methods are intra-class correlation coefficient (ICC) and mutual absolute agreement method (1, 2), studied by Shroutard Fleiy [18]. In order to interpret the Kappa coefficient, Landis and Kochs' [19] suggestion of 0.21-0.41 as poor, 0.41-0.6 as a medium, 0.6-0.8 as considerable and 0.8-1.0 as almost complete was employed. For ICC, the verbal change was used based on Steven's (2005) recommendation: (ρ <0.4) poor reliability, (0.40 < ρ <0.75) medium-good reliability and (ρ> 0.75) excellent reliability. The statistical tests were implemented in R (v.3.02) and SPSS (v.20) software.

Table 1: Parameters of SI, OCRA checklist and ART risk categorizations criteria

Risk level	Risk index		
	SI	OCRA	ART
1	<3	<7.6	<11.9
2	3-6.9	7.6-14	12-21.9
3	≥7	≥14.1	≥22

RESULTS

The results of this investigation showed that the inter-examiner reliability with ICC categorizations obtained for SI and OCRA Checklist methods equal to (ρ =0.54, 95%CI: 0.49-0.61) and (ρ =0.72, 95%CI: 0.69-0.79), as well as (ρ =0.77, 95%CI: 0.70-0.81) for ART. In Table 2, the calculation of ICC and Kappa coefficient for OCRA Checklist, SI and ART are presented for risk categorizations, final score and task variables. Among the variables of three methods, the minimum inter-examiner reliability with ICC statistical factor was related to the hand/wrist posture (ρ =0.42) for SI, awkward posture/ movement (ρ =0.56) for OCRA Checklist and additional factors (ρ =0.58) for ART.

In Table 3, the intra-examiner reliability parameter with the ICC, intra-examiner Kappa coefficient for risk categorization, final score, and exposure variables are presented for the three methods. The reliability for the final score of OCRA Checklist, SI and ART was equal to (κ = 0.72), (κ = 0.68) and (κ = 0.76), respectively. The maximum agreement and reliability in the exposure variables were frequency/repetitive movement (κ=0.85) for the ART method and working speed (κ = 0.82) and frequency of technical movement (κ = 0.84) for the SI and OCRA Checklist.

Table 2: Parameters of inter-examiner reliability for SI, OCRA Checklist and ART.

Risk score and task variable	$Kappa_{(kw)}$	ICC_c(95% CI)
SI risk classification	0.52	0.54 (0.49-0.61)
Strain Index Score	0.44	0.46 (0.33-0.59)
Intensity of exertion	0.39	0.44 (0.34-0.48)
Duration of exertion	0.50	0.53 (0.47-0.60)
Efforts per minute	0.51	0.55 (0.46-0.62)
Hand/wrist posture	0.37	0.42 (0.37-0.45)
Speed of work	0.46	0.50 (0.45-0.54)
OCRA risk classification	0.68	0.72 (0.69-0.79)
OCRA checklist score	0.62	0.66 (0.50-0.71)
Frequency of technical actions	0.70	0.73 (0.58-0.80)
Force exertion	0.60	0.63 (0.53-0.69)
Awkward posture/movement	0.52	0.56 (0.47-0.61)
Additional factors	0.65	0.68 (0.63-0.74)
ART risk classification	0.72	0.77 (0.70-0.82)
ART score	0.67	0.73 (0.65-0.78)
Frequency/Repetition movement	0.79	0.84 (0.76-0.88)
Force	0.71	0.76 (0.71-0.83)
Awkward postures	0.64	0.70 (0.65-0.74)
Additional factor	0.47	0.58 (0.51-0.65)

Table 3: Parameters of intra-examiner reliability for SI, OCRA Checklist and ART.

Risk score and task variable	$Kappa_{(kw)}$	ICC_c(95% CI)
SI risk classification	0.76	0.82 (0.77-0.87)
Strain Index Score	0.72	0.65 (0.58-0.71)
Intensity of exertion	0.71	0.83 (0.76-0.88)
Duration of exertion	0.80	0.86 (0.81-0.90)
Efforts per minute	0.78	0.85 (0.80-0.89)
Hand/wrist posture	0.77	0.72 (0.67-0.78)
Speed of work	0.82	0.86 (0.78 -0.90)
OCRA risk classification	0.79	0.85 (0.79-0.89)
OCRA checklist score	0.68	0.76 (0.67-0.83)
Frequency of technical actions	0.84	0.90 (0.86-0.93)
Force exertion	0.70	0.74 (0.69-0.78)
Awkward posture/movement	0.74	0.80 (0.75-0.82)
Additional factors	0.82	0.88 (0.82-0.91)
ART risk classification	0.82	0.90 (0.85-0.94)
ART score	0.76	0.81 (0.77-0.85)
Frequency/Repetition movement	0.85	0.92 (0.86-0.95)
Force	0.81	0.86 (0.81-0.89)
Awkward postures	0.77	0.80 (0.75-0.86)
Additional factor	0.75	0.78 (0.71-0.85)

DISCUSSION

There are several conventional methods for quantifying reliability rate of measurement results of observational methods; agreement ratio of weighted Kappa and ICC (1, 2) were used in this study. The overall results showed higher ART reliability than the two other methods. Inter-examiner reliability for risk categorizations and a final score of ART method using ICC was moderate to good and, using Kappa was moderate to considerable. The results of reliability for both OCRA and SI methods using ICC were moderate to good and, using Kappa, were moderate to considerable for OCRA Checklist and moderate for SI. The difference in evaluation methods and scoring scales in three methods can explain a difference in the finding of the reliability coefficient. Comprehensively, in all three methods, assessment of exposure to awkward postures of biomechanics is used based on the posture criteria. The intensity of force is evaluated using a similar verbal interpretation (CR-Borg scale). Exposure duration is also evaluated based on the percentage of the task cycle. Nevertheless, ART, SI, and OCRA Checklist methods have different structures for measuring repetitive activities, posture, and work and rest cycle. Since ART is the developed method of OCRA, the physical exposure variables of both methods are similar in structure, but the number of scoring levels is different. For example, the scoring of the retrieval time variable has 5 levels in OCRA Checklist and 4 levels in ART. The ART method, in contrast to OCRA Checklist, considers awkward postures of the head, neck and back in the assessment. In the structure of SI method, compared to ART and OCRA Checklist, the shoulder posture and additional

risk factors (MSDs) such as hand and arm vibration, contact stress, and cold workplace are not considered. General nature of the OCRA Checklist and ART structures leads to further complexity of scoring scales and an increase in the range of scores. As a result, the measurement of reliability is also affected. So, it is expected that tools and assessment methods of the complexity risk require more time for training and, also, the required time for assessing each occupational task is longer. In the present study, this issue is observed.

The finding of inter-examiner reliability of task variables in SI method, compared to OCRA Checklist and ART methods, are lower at the moderate level ($\rho > 0.42$) and ($\kappa>0.31$). This is one of the causes of low-reliability results is the low number of assessment tasks. In addition, the low scoring scale can reduce ICC. An increase in scoring scale distribution for the variables can improve the reliability coefficients.

Stevens *et al.* [20], ICC (1, 2) used to determine the inter-examiner reliability of the SI method. In the mentioned study, 15 examiners were selected and 73 tasks purposefully chosen among the occupations within homogenous physical exposure. The task variables of SI were obtained in the range of 0.77 to 0.81, except for hand/wrist posture with medium to good reliability ($\rho = 0.60$). In the present study, if the samples were selected purposefully for the assessment, the ICC of task variables could be improved, but the samples were randomly selected from several industries. However, in the present research, the ICC for the final score of SI ($\rho = 0.45$, CI= 0.33-0.59) was similar to the studies by *Stevens et al.* ($\rho =0.43$) and Paulsen *et al.* ($\rho =0.59$) [21]. The inter-examiner reliability in semi-quantitative assessments might be higher when scoring is done based on videos, compared to the assessment on the site and workplace. For example, Spielholz [14] assessed the intra-examiner reliability of SI method in 125 tasks by 3 experts and 1 novice. The assessment of tasks was performed on the site without any videos. The SI scores of task variables were poor to moderate using the Kappa coefficient ($0.31< \kappa < 0.44$). The results of task variables of OCRA Checklist for the inter-examiner reliability ($\rho =0.74$) were medium. In another study, 320 tasks were determined using ICC [12], reliability between OCRA checklist and SI. The SI score of task variables was obtained poor to moderate in the range of 0.16-0.60, except for the exertion variable. In the research by Rhéna *et al.* [22] on 11 examiners with 10 recorded videos of tasks, the inter-examiner reliability of OCRA Checklist was obtained for risk categorizations and task variables with poor to moderate by Kappa ($0.58 < \kappa < 0.63$). Also, the results of Rhéna *et al.* were consistent with the findings of this study for the inter-examiner reliability.

In sum, the results of previous studies on risk assessment methods showed higher inter-examiner reliability than the repeatability of the methods. For example, in another part of Rhéna *et al.*, the results of examiners were obtained in two parts with 6-week intervals. The Kappa coefficient for risk categorizations and task variables of OCRA Checklist was moderate to considerable. In the present study, the intra-examiner reliability with the between of $0.68< \kappa <0.82$ was obtained, also in the previous works, the level of scores was lower than this study. Intra-examiner reliability and repeatability of the SI method were examined by Stephens *et al.* [23]. The risk categorizations and task variables with ICC were in the range of $0.81< \rho <0.95$. In previous works, the intra-examiner reliability of risk categorizations of the SI method was excellent estimated, which is similar to the present study.

Results of inter-examiner reliability of ART method for the final score and risk categorizations using ICC were $\rho=0.73$ and $\rho =0.77$, and, using Kappa, they were $\kappa =0.67$ and $\kappa =0.72$, respectively. The intra-examiner reliability of ART was calculated using Kappa for the final score ($\kappa=0.76$), and using ICC for risk categorizations ($\rho=0.82$). The intra- and inter-examiner reliability of ART was evaluated very well, which was consistent with the results by Roodbandi *et al.* [24], who estimated proper reliability for ART method.

CONCLUSION

Generally, inter- and intra-examiner reliability of ART was higher than OCRA checklist and SI, but the results obtained of this research showed that three methods were proper for assessment in the different workplaces. The OCRA Checklist and ART for the research cases or workplaces should be used based on assessment objectives, researcher's questions, details of assessment, reliability, and validity. Also, the time required for training the people and assessing the tasks should be considered. The ART method and OCRA Checklist require more time than the SI.

ETHICAL ISSUES

Ethical issues for instance plagiarism have been considered by the authors.

CONFLICT OF INTERESTS

There are no conflicts of interest.

AUTHORS' CONTRIBUTIONS

All authors correspondingly assisted to write this manuscript.

FUNDING/ SUPPORTING

The work was financially supported by Hamadan University of medical sciences.

ACKNOWLEDGEMENTS

This work was a part of from MSc thesis, Hamadan University of Medical Sciences, Hamadan, Iran

REFERENCES

[1] Baldwin ML. Reducing the costs of work-related musculoskeletal disorders: targeting strategies to chronic disability cases. Journal of Electromyography and Kinesiology. 2004;14(1):33-41.

[2] Bongers P, Ijmker S, Van den Heuvel S, Blatter B. Epidemiology of work related neck and upper limb problems: psychosocial and personal risk factors (part I) and effective interventions from a bio behavioural perspective (part II). Journal of occupational rehabilitation. 2006;16(2):272-95.

[3] Bovenzi M. Health risks from occupational exposures to mechanical vibration. La Medicina del lavoro. 2006;97(3):535-41.

[4] Dempsey PG, McGorry RW, Maynard WS. A survey of tools and methods used by certified professional ergonomists. Applied ergonomics. 2005;36(3):489-03.

[5] Neumann WP, Village J. Ergonomics action research II: a framework for integrating HF into work system design. Ergonomics. 2012;55(2):1140-56.

[6] Takala E-P, Pehkonen I, Forsman M,. Systematic evaluation of observational methods assessing biomechanical exposures at work. Scandinavian journal of work, environment & health. 2010;36(1):3-24.

[7] David G. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. Occupational medicine. 2005;55(4):190-99.

[8] Van der Beek AJ, Frings-Dresen M. Assessment of mechanical exposure in ergonomic epidemiology. Occupational and environmental medicine. 1998;55(1):291-99.

[9] Rahman MNA, Razak NSA. Review on Pen and Paper Based Observational Methods for Assessing Work-related Upper Limb Disorders. Indian Journal of Science and Technology. 2016;9(2):1-11.

[10] Steven Moore J, Garg A. The strain index: a proposed method to analyze jobs for risk of distal upper extremity disorders. American Industrial Hygiene Association. 1995;56(4):443-58.

[11] Colombini D, Occhipinti E, Cerbai M, Battevi N, Placci M. Updating of application procedures and criteria for OCRA Checklist. Med Del Lav. 2011;102(2):1-39.

[12] Ferreira J, Gray M, Hunter L. Development of an assessment tool for repetitive tasks of the upper limbs (ART). Derbyshire: Health and Safety Executive. 2009.

[13] Borg GA. Psychophysical bases of perceived exertion. Med sci sports exerc. 1982;14(1):377-81.

[14] Spielholz P, Bao S, Howard N. Reliability and validity assessment of the hand activity level threshold limit value and strain index using expert ratings of mono-task jobs. Journal of occupational and environmental hygiene. 2008;5(1):250-57.

[15] Garg A, Kapellusch JM. Job analysis techniques for distal upper extremity disorders. Reviews of Human Factors and Ergonomics. 2011;7(3):149-96.

[16] Jones T, Kumar S. Comparison of ergonomic risk assessment output in four sawmill jobs. International Journal of Occupational Safety and Ergonomics. 2010;16(4):105-11.

[17] Cohen J. A coefficient of agreement for nominal scales. Educational and psychological measurement. 1960;20(2):37-46.

[18] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychological bulletin. 1979;86(2):420-28.

[19] Landis JR, Koch GG. The measurement of observer agreement for categorical data. biometrics. 1977;33(1):159-74.

[20] Stevens EM, Vos GA, Stephens J-P, Moore JS. Inter-rater reliability of the strain index. Journal of occupational and environmental hygiene. 2004;11(1): 745-51.

[21] Paulsen R, Gallu T, Gilkey D. The inter-rater reliability of Strain Index and OCRA Checklist task assessments in cheese processing. Appl Ergon. 2015;51(1):199-04.

[22] Rhéna I-M, Balliuc N, Forsmana M. OCRA inter- and intra-ergonomist reliability in ten video recorded work tasks. Proceedings 19th Triennial Congress of the IEA2015. p. 14.

[23] Stephens J-P, Vos GA, Stevens EM, Moore JS. Test–retest repeatability of the strain index. Applied ergonomics. 2006;37(3):275-81.

[24] Roodbandi A, Choobineh A, Feyzi V. The Investigation of Intra-rater and Inter-rater Agreement in Assessment of Repetitive Task (ART) as an Ergonomic Method. Occup Med Health Aff. 2015;3(5):1-5.