THESIS

RELIABILITY OF ERGONOMIC EXPOSURE ASSESSMENT: COMPARING THE STRAIN INDEX AND THE OCRA CHECKLIST

Submitted by

Robert James Paulsen

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2013

Master's Committee:

Advisor: David Gilkey

John Rosecrance
Raoul Reiser

ABSTRACT

RELIABILITY OF ERGONOMIC EXPOSURE ASSESSMENT: COMPARING THE STRAIN INDEX AND THE OCRA CHECKLIST

Occupational ergonomists utilize a variety of observational methods to identify jobs that elevate workers' risk for developing work-related musculoskeletal disorders. Internationally, the Strain Index (SI) and the Occupational Repetitive Actions (OCRA) Checklist are two of the most popular upper extremity exposure assessments available. Both are founded on similar biomechanical and epidemiological principles, but their approach to quantification and estimation of risk factor magnitude is quite different. The purpose of the present study was to analyze the inter-method reliability of SI and OCRA Checklist exposure assessments. An additional aim was to estimate the inter-rater reliability of both methods.

Twenty-one cheese-processing jobs were video recorded at an Italian facility. Eight individuals with occupational health training were recruited to rate every job using both the SI and OCRA Checklist. Inter-method reliability was characterized using kappa coefficients and Spearman correlations, and inter-rater reliability was characterized using intraclass correlation coefficients.

Strain Index and the OCRA Checklist assessments produced moderately reliable results, generally classifying the same job exposures to physical risk factors similarly. Systematic bias due to rater effects was observed, and a lack of individual rater familiarity with one or both methods may have contributed to this bias. Further, the assessment of multi-task jobs was associated with lower inter-method reliability than what was observed for simpler jobs.

Inter-rater reliability of the SI and OCRA Checklist summary scores and exposure classifications suggested moderate intra-method reliability. Summary scores may be a reliable measure of exposure to meet the needs of epidemiologists and occupational health and safety practitioners.

ACKNOWLEDGEMENTS

I would like to thank my committee members, Dr. John Rosecrance, Dr. David Gilkey, and Dr. Raoul Reiser, for their valuable feedback and support during the development of this manuscript. And the present study would not have been possible without the participation of Lelia Murgia and Thomas Gallu from the University of Sassari, and Dr. Rosecrance, Dr. Gilkey, Anthony Mixco, Galia Modai, and Daniel Tucker from Colorado State University. Thank you all for the excellent efforts you put forth. I would also like to thank Dr. Ann Hess and Dr. Thomas Keefe from Colorado State University for their assistance with the statistical analyses. Finally, I am very thankful for all of the love and encouragement from my wife, Mandy, and I am especially grateful for the support she provided me during the entirety of my graduate studies.

TABLE OF CONTENTS

1. Introduction

1.1 Background

Musculoskeletal disorders (MSDs) comprised a third of all work-related injuries and illnesses in the United States that involved days away from work in 2011 (BLS 2012). The incidence of work-related MSDs was 38.5/10,000 worker-years (BLS 2012). Incidence rates for MSDs attributed to repetitive motion were 2.25 times higher in manufacturing than among all private and public work sectors (BLS 2012). Additionally, repetitive motion-related cases were observed to be the most severe of any class of injury or illness, requiring a median 23 days away from work (BLS 2012). Reviews of the occupational injury and illness etiologic literature have found strong associations between physical risk factors (forceful exertions, high repetition, working in non-neutral postures, and localized vibration) with MSDs (Bernard 1997, NRC/IOM 2001). Further, the interaction between these risk factors appears to be more than additive (Silverstein *et al.* 1986, Chiang *et al.* 1993, Bernard 1997). Most recently, prospective cohort studies have shown strong relationships between exposure to physical risk factors and health outcomes (Gerr 2010, Garg *et al.* 2012).

Two of the most comprehensive and widely used measures of exposure to physical risk factors are the Strain Index (SI) (Moore and Garg 1995) and the Occupational Repetitive Actions (OCRA) Checklist (Colombini *et al.* 2000, Colombini *et al.* 2011). Both of these tools provide a single, risk summary score based on an individual's exposure to physical risk factor parameters at work. Both methods represent these parameters as task variables to which the analyst (rater) assigns a task-variable score. The SI summary exposure score is the product of six task-variable scores: (i) *intensity of exertion*, (ii) *duration of exertion*, (iii) *efforts per minute,* (iv) *hand/wrist posture*, (v) *speed of work*, and (vi) *task duration per day* (Moore and Garg 1995). The task-variable scores are selected from a five-level ordinal scale, which corresponds to the magnitude of exposure to the relevant physical parameter. The multiplicative product of the five task-variable scores is the SI summary score.

The OCRA Checklist also summarizes exposure in terms of six task-variable scores (i) *frequency of technical actions*, (ii) *force*, (iii) *awkward postures and movements*, (iv) *additional factors*, (v) *lack of sufficient recovery*, and (vi) *task duration* (Colombini 1998, Occhipinti 1998). Task-variable scores are primarily determined by measuring the percentage of time that a worker is exposed to a multiplicity of physical parameters (*e.g.* frequency of repetitive action, intensity of force exertion, extreme angles of posture, *etc.*). Increasing percentages of exposure to these parameters correspond to increasing task-variable scores along a continuous scale. The sum of the *frequency*, *force*, *posture*, and *additional factors* scores is weighted by the scores for the *lack of sufficient recovery* and *task duration* variables, producing the OCRA Checklist summary exposure score. The primary difference between the OCRA Checklist and the SI is that the former applies to the upper extremities including the shoulder, whereas the SI only applies to the distal upper extremities (Moore and Garg 1995, Colombini *et al.* 2000). Also, the two methods parameterize some of the same physical risk factors differently. For instance, the SI task variable *efforts per minute* is based on the frequency of complex upper extremity (UE) exertions. In contrast, the OCRA Checklist *frequency of technical actions* task variable is based on the frequency of fundamental groupings of UE exertion or motion. See Colombini *et al.* (2002) for the most complete description of OCRA Checklist parameters.

The SI has demonstrated ecological, cross-sectional, and longitudinal associations with upper extremity MSDs when applied to six different study samples of workers (Garg *et al.* 2007, Spielholz *et al.* 2008, Gerr 2010, Garg *et al.* 2012). The OCRA has demonstrated an association with MSDs in two cross-sectional studies of the same large sample of workers (Grieco 1998, Occhipinti and Colombini 2007, Takala *et al.* 2010). Further, the intra-method reliability of OCRA Checklist exposure estimates has never been evaluated, whereas the intra-method reliability of SI results was reported to be moderate or good (Takala *et al.* 2010).

Prior studies of the agreement between physical risk exposure classifications of the SI and OCRA Index (a more rigorous and time consuming form of the OCRA Checklist) have suggested moderate or good inter-method reliability (Takala *et al.* 2010). However, these inter-method reliability studies either evaluated a very small number of tasks (Apostoli *et al.* 2004, Jones and Kumar 2010), or a single individual or team collected all of the exposure data (Apostoli *et al.* 2004, Jones and Kumar 2010, Chiasson *et al.* 2012).

1.2 *Purpose of the present study*

The purpose of the present study was to evaluate the inter-method reliability of SI and OCRA Checklist exposure assessments completed by eight independent raters. An additional aim was to evaluate the inter-rater reliability of both methods.

1.2.1 *Specific Aims*

Evaluate the inter-method reliability of SI and OCRA Checklist exposure assessments for each rater.

Hypothesis 1a: Strain Index and OCRA Checklist scores will be well correlated (Spearman $r_s \geq 0.75$) and agree moderately beyond chance ($\kappa/\kappa_w \geq 0.40$).

Hypothesis 1b: Strain Index and OCRA Checklist assessments will classify the same exposures similarly ($p_o \geq 75\%$).

Evaluate the inter-rater reliability of SI results to OCRA checklist results.

Hypothesis 2a: The intraclass correlation coefficient for SI summary and task-variable scores will be significantly greater than 0.40, signifying at least moderate inter-rater reliability.

Hypothesis 2b: The intraclass correlation coefficient for OCRA Checklist summary and task-variable score will be significantly greater than 0.40, signifying at least moderate inter-rater reliability.

2. Review of the literature

2.1. *Three types of physical exposure assessments*

Controversy remains regarding the strength of the association between exposure to physical risk factors and MSDs. Huisstede *et al.* (2006) concluded that the strength of this association is difficult to accurately appraise given the inconsistent reporting methods for the incidence and prevalence of UE MSDs among working populations. Further, Da Costa and Vieira's (2010) review of work-related MSD epidemiology highlights that many recent publications identifying relationships between physical risk factors and disease were the product of methodologically weak research design. One of the key weaknesses according to da Costa and Vieira was inaccurate or imprecise quantification of physical risk factor exposure (or physical exposure) (2010).

Improving the quantification of physical exposure is not only important for epidemiologic research. Industry ergonomists and occupational health and safety (OHS) practitioners also need accurate and precise measures to identify the most common and the most significant exposures to MSD risk factors within a workplace (Colombini *et al.* 2001, Bao 2004, Occhipinti and Colombini 2012). Ergonomists rely on three primary types of physical exposure assessment tools: self-report methods, observational methods, and direct methods (Kilbom 1994, Spielholz *et al.* 2001, David 2005).

Self-report methods include surveys, questionnaires, interviews, and journal or diary logs (Kilbom 1994). These methods allow for relatively quick and easy assessment of many workers, can be adapted to assess a multiplicity of workplace environments, and can be administered with relatively few resources (David 2005). Of the three types of exposure assessments, self-report measures are the only ones capable of estimating non-work and past exposures (Punnett and

Wegman 2004). However, self-reports are affected by low validity and low reliability due to workers' differing perceptions of the physical risk factors inherent in their work (Kilbom 1994, David 2005). Spielholz *et al.* (2001) also notes that varying degrees of literacy and reading comprehension can bias the results of self-report measures. Thus, while some ergonomists use self-reports as the only measure of exposure, more often these qualitative methods are paired with either observational or direct measures (Spielholz *et al.* 2001).

Direct methods provide the most accurate means to quantify physical exposures. These methods require biomechanical instrumentation to measure physical parameters, such as muscle activation patterns, external force generation, joint-angle deviation, oxygen consumption, and working heart rate (Kilbom 1994, Burdorf and van der Beek 1999, Li and Buckle 1999, David 2005, Dempsey *et al.* 2005). Some of the most commonly used tools to directly measure exposures include electrogoniometers, grip dynamometers, push/pull sensors, electromyography, and heart rate monitors (Spielholz *et al.* 2001, Dempsey *et al.* 2005). Direct measures often provide extremely detailed exposure data with great accuracy, but properly interpreting the data may be difficult without advanced training and experience (Kilbom 1994, De Luca 1997, Li and Buckle 1999, David 2005). The instruments and accompanying analysis software may also be very expensive to purchase and maintain (Spielholz *et al.* 2001, David 2005). Additionally, when used outside the laboratory, the instrument data collection systems may be compromised by environmental interferences, such as strong electromagnetic fields (De Luca 1997, Li and Buckle 1999). Finally, individual workers might feel that wearing direct measurement equipment is uncomfortable or restricts their natural movements (Kilbom 1994, Li and Buckle 1999, Ebersole and Armstrong 2006).

Kilbom (1994) describes observational methods as a compromise between the time and resource intensive direct measures and the low validity and low reliability of self-reports. As such, observational methods have become the most popular of the three types of ergonomic assessments (Takala *et al.* 2010). These methods require a trained observer to systematically evaluate job-task elements using risk factor identification forms or checklists (Spielholz *et al.* 2001), and most observational methods measure multiple physical exposures (David 2005). Burt *et al.* (2000) wrote that the ideal observational method can be inexpensively and practically applied to a wide variety of jobs while still producing reliable and valid results.

Observational measures are based on a combination of objective measurements as well as subjective estimates. Objective measures are often made using time-motion study and simple tools such as a tape measures, tally counters, and scales. For instance, analysts using the NIOSH Lifting Equation directly measure the vertical distance a lifted object travels with a tape measure, but they must also subjectively estimate whether significant control of the object is required (Waters *et al.* 1993). Similarly, the SI analyst measures work-cycle time and counts the number of efforts per minute during a job-task, but the intensity of physiologic exertion during work is subjectively estimated (Moore and Garg 1995). Many observational methods can be applied at the worksite or back in the lab using video recordings of the work. The use of systematic video recording is recommended or required by a number of observational methods (Moore *et al.* 2001, Spielholz *et al.* 2001, Occhipinti and Colombini 2005, Takala *et al.* 2010), and it may reduce the effects of observer bias (Li and Buckle 1999).

While the systematic design of observational measures is intended to increase reliability and validity of results compared to subjective measures, methodological limitations must be recognized. For example, directly measured force exertion values can differ significantly from observational estimates of force exertion (Kumar 1993, Spielholz *et al.* 2001, Jones and Kumar 2010). Spielholz *et al.* (2001) compared several UE physical exposure measurement techniques applied to 18 workers completing a diversity of jobs within the tree nursery industry. The investigators assessed exposure to high force exertion, repetition, awkward postures, and rapid work pace by directly measuring physical parameters with electrogoniometers and electromyography. The same exposures were assessed using systematic video analysis (observational measure) and workers' responses to a sliding-scale questionnaire (self-report measure). In general, direct methods provided the most precise and accurate exposure estimates. Despite lower precision and accuracy of observational measures, though, the systematic video analysis produced results that were generally in agreement with the direct measures. Self-reports of exposure, in contrast, were often highly variable and likely overestimated exposure. Spielholz *et al.*(2001) concluded that while direct methods are preferable, their expensive and sometimes impractical use may be unnecessary when surrogate observational measures are available. Reliability and validity studies of observational methods must be conducted to determine if they may be used instead of direct measures and in which research/assessment contexts their use is most appropriate.

2.2. Reliability and validity of observational measures

2.2.1. *Inter-method reliability*

When an observational assessment actually measures what its users intend to measure, it is accurate (Armstrong *et al.* 1992, Kilbom 1994). Methodological accuracy is more formally referred to as the criterion validity of a method. The criterion validity of an observational measure can be established by quantifying how well the exposure assessment, when applied to a group of jobs or workers, predicts the incidence or prevalence of work-related MSDs. When a measure can predict MSD outcomes accurately, it is said to have good predictive validity (Streiner and Norman 2008). However, effective prospective study design, control of bias, and control of confounding variables can be significant challenges to the success of predictive validity studies. This is largely due to the complexity of quantifying MSD prevalence, MSD incidence, and the interaction among the many risk factors associated with MSD development (Burdorf and van der Beek 1999, da Costa and Vieira 2010).

Criterion validity of an exposure assessment method can also be established by comparing its results to those generated by a "gold standard" (*i.e.* previously validated) instrument measuring the same construct (Armstrong *et al.* 1992, Kilbom 1994, Takala *et al.* 2010). When an observational method is used to evaluate a number of exposures, and its results strongly agree with those of a different, validated measure, the assessment is said to have concurrent validity. Concurrent validity is established using a parallel testing or inter-method reliability study design (Armstrong *et al.* 1992, Kilbom 1994). The gold-standard method should measure the true exposure on average, and there should be minimal random error associated with its use (Burdorf 1995). Takala *et al.* (Takala *et al.* 2010) wrote, "There is, however, no general 'gold standard' for assessing biomechanical exposures." For instance, when Spielholz *et al.* (2001) attempted to directly measure UE joint deviation using electrogoniometers, they found that 4^{o} to 5^{o} of calibration error prevented accurate estimates of radial deviation, the observed range of which was only about 10^{o}. Similarly, the results of direct measures of force exertion using surface EMG may be obfuscated because EMG signal interpretation and analysis methods are based on isometric rather than dynamic muscle contractions (De Luca 1997, Hägg *et al.*

2004). Calibration issues and data interpretation difficulties mean that even direct measures cannot be considered error-free, gold-standard measures of exposure to physical risk factors.

Despite these limitations, inter-method reliability studies of observational measures are most successful when comparing results to those from a well understood direct method. Most of these are for posture assessments (Chiang *et al.* 1993, Burt and Punnett 1999, Juul-Kristensen *et al.* 2001, Lowe 2004), although some have examined force or repetition estimates (Ketola *et al.* 2001, Spielholz *et al.* 2001, Bao *et al.* 2006). When comparison to direct measures is not possible, inter-method reliability studies can only highlight the agreement of results generated by different methods used to assess the same exposures. Examining the agreement and strength of association between measures may still be intrinsically valuable to ergonomists who perform exposure assessments using multiple observational methods (Chiasson *et al.* 2012, Jones and Kumar 2010, Takala *et al.* 2010). For example, it may be of little value to measure physical exposures with two methods that are highly correlated and produce results that are usually in agreement. However, if those two methods evaluate exposure to physical risks differently, then the agreement in results may serve to strengthen the analyst's exposure assessment conclusions.

When comparing the assessments produced by different observational methods, raw agreement comparisons require that the results of exposure assessment be categorized or ranked according to common exposure classification criteria. These criteria are generally dichotomous or multi-level risk ranking systems provided by the developers of each method. For example, Moore and Garg (1995) classify SI scores greater than or equal to 5.0 as indicators of hazardous exposure, and scores less than 5.0 indicate safe levels of exposure. Once the common criteria has been selected, Armstrong *et al*. (1992) suggests that the methods be compared using a misclassification matrix or contingency table to calculate the percentage of overall agreement.

Percentage of overall agreement is calculated along the main diagonal of the contingency table (Fleiss and Cohen 1973). Further, the agreement between exposure classifications can be measured using weighted or unweighted kappa coefficients (Fleiss 1986, Shoukri 2004). The kappa coefficients quantify the beyond-chance overall agreement between categorical items. Alternatively, the strength of association between exposure assessment scores can determined using Pearson or Spearman correlation coefficients (Armstrong *et al.* 1992, Streiner and Norman 2008). The Spearman correlation is applied when data are non-normally distributed or when data are not continuous. Which statistics are chosen depends on the characteristics of the data collected and the aims of the inter-method reliability study.

2.2.2. Intra-method reliability

Most discussions of methodological reliability refer to intra-method reliability, that is, the coherency of results produced by a single method (Armstrong et al. 1992). When a measure yields the same result after repeated assessment of the same parameters, it is precise or reliable (Armstrong *et al.* 1992, Kilbom 1994). There are two types of intra-method reliability: inter-rater reliability and intrarater reliability. Inter-rater reliability refers to the agreement of results from repeat measures conducted by multiple analysts or raters, whereas intrarater reliability refers to the agreement of results from repeat measures conducted by a single rater at two or more points in time (Dunn 1989, Armstrong *et al.* 1992). Both types of intra-method reliability are important to consider when selecting an exposure assessment.

Streiner & Norman (2008) stress that methods are not themselves reliable, but rather that the results of a method are reliable when it is applied under a specified set of conditions. These conditions include the users of the method (e.g. the ergonomic analysts or raters), the subjects of assessment (e.g. jobs rated) and the environment in which they are measured (e.g. videos of work in a complex processing facility). Further, reliability studies of observational methods should be used under realistic conditions of their application (Dunn 1989, Kilbom 1994). For example, the SI was designed to evaluate repetitive manufacturing jobs. Therefore, an SI reliability study should not be undertaken using exposure assessments of non-repetitive jobs or exposure assessments of repetitive jobs simulated within a laboratory.

Several statistics are available for quantifying the reliability of observational measure results: proportion of agreement, weighted and unweighted kappa coefficient, and several types of the intraclass correlation coefficient (ICC) (Armstrong *et al.* 1992, Kilbom 1994, Burt and Punnett 1999). Proportion of agreement is the simplest reliability statistics (Kilbom 1994, Burt and Punnett 1999). The statistic is a valuable summary of agreement in results between analysts (raters). However, proportion of agreement is not preferred as the sole indicator of reliability because it cannot account for chance agreement (Fleiss and Cohen 1973, Ebersole and Armstrong 2006).

The kappa coefficient does account for chance by comparing raw agreement with that which would be expected by chance alone based on the marginal distributions of results (Cohen 1968, Streiner and Norman 2008). Kappa works well for dichotomous and nominal data, but ordinal data requires weighting the kappa coefficient to account for multiple levels of disagreement (Streiner and Norman 2008). Any number of weighting schemes are possible, but the use of different weights can obscure the interpretation and comparison of kappa coefficients between studies (Maclure and Willett 1987, Streiner and Norman 2008). The most commonly chosen weighting scheme, whereby disagreement weights are based on the square of the levels of discrepancy, produces a weighted kappa coefficient equivalent to an intraclass correlation coefficient (Fleiss and Cohen 1973, Streiner and Norman 2008). For this reason, some prefer to use an ICC instead (Ebersole and Armstrong 2006, Stephens *et al.* 2006, Stevens *et al.* 2004). Further, some prefer to avoid kappa when analyzing ordinal data because the value of the coefficient may fluctuate depending on the number of categorical rankings (Maclure and Willett 1987, Burt and Punnett 1999).

Many investigators characterize intra-method reliability using an intraclass correlation coefficient (Maclure and Willett 1987, Stevens *et al.* 2004, Streiner and Norman 2008). The ICC is often referred to as a statistic appropriate for the analysis of continuous data (Armstrong *et al.* 1992, Burdorf 1995), but it is equally applicable to categorical data (Berk 1979, Maclure and Willett 1987). There are several forms of the ICC, but they all characterize the proportion of variance in results due to the raters compared to the overall variance in results. Intra-method reliability studies of observational ergonomic measures most commonly rely on an agreement-based, single measures ICC (Stevens *et al.* 2004, Ebersole and Armstrong 2006, Stephens *et al.* 2006, Bao *et al.* 2009a, Xu *et al.* 2011, Dockrell *et al.* 2012). And when all raters evaluate all possible jobs, Shrout and Fleiss (1979) refer to it as ICC(2,1). Other forms of the ICC may be more appropriate depending on the design of the reliability study (Armstrong *et al.* 1992, Burt and Punnett 1999). The variance estimates required for calculating the ICC are obtained using a two-way ANOVA with random effects model.

Pearson and Spearman correlation coefficients are sometimes used to characterize agreement, but these two types of statistics are often be inappropriate for intra-method reliability studies (Armstrong *et al.* 1992, Burdorf 1995, Essendrop *et al.* 2001). They are inappropriate because they do not quantify the systematic differences between two raters, which will overestimate the true reliability (Armstrong *et al.* 1992, Streiner and Norman 2008). Streiner &

Norman (2008) explain, however, that in practice systematic error is often not great enough to alter the reliability coefficient significantly. When this is the case, Pearson or Spearman correlation coefficients will be close to the value of an ICC.

2.3. *The Strain Index (SI)*

2.3.1. *Development and application*

The SI is an observational measure that quantifies exposure to physical risk factors associated with the development of distal upper extremity (DUE) MSDs (Moore and Garg 1995). The DUE refers to muscle-tendon units and nerves of the elbow, forearm, wrist, and hand. Distal upper extremity disorders include carpal tunnel syndrome (CTS), varieties of tendinitis, synovitis, tenosynovitis, and bursitis (Tanaka *et al.* 2001). The analyst uses the SI to evaluate the magnitude of work-related physical exposure in terms of six parameters. These parameters are represented as six task variables, and the analyst assigns a task-variable score to each one based upon the characteristics of the job under evaluation. These six task variables are (i) *intensity of exertion*, (ii) *duration of exertion*, (iii) *efforts per minute,* (iv) *hand/wrist posture,* (v) *speed of work*, and (vi) *duration of task per day* (Moore and Garg 1995). There are five possible levels of exposure to each task variable, and Table 2.1 displays the verbal descriptors and physical criteria the analyst uses to determine the level and corresponding score for task-variable exposure. The SI analyst selects the most appropriate task-variable score using a combination of direct time-motion measurement techniques paired with professional judgment. The multiplicative product of the six task-variable scores determines the SI summary exposure score (also referred to as the SI summary score or SI score). Unlike the discrete task-variable scores, the SI summary score is a continuous variable, ranging from 0.1 to 1053 (Moore and Garg 1995).

Table 2.1. Five levels of physical exposure and the corresponding scores for the SI task variables.

Intensity of exertion (Borg CR-10 scale)		Duration of exertion (%)		Efforts per minute		Hand/wrist posture		Speed of work		Duration per day	
Light (0-2)	*[1]	<10	[0.5]	<4	[0.5]	Very Good	[1]	Very Slow	[1]	<1 hour	[0.25]
Somewhat hard (3)	[3]	10-29	[1]	4-8	[1]	Good	[1]	Slow	[1]	1-2 hours	[0.5]
Hard (4-5)	[6]	20-49	[1.5]	9-14	[1.5]	Fair	[1.5]	Fair	[1]	2-4 hours	[0.75]
Very Hard (6-7)	[9]	50-79	[2]	15-19	[2]	Bad	[2]	Fast	[1.5]	4-8 hours	[1]
Near Maximal (8-10)	[13]	≥80	[3]	≥20	[3]	Very Bad	[3]	Very Fast	[3]	>8 hours	[1.5]

*Task variable scores are shown in brackets (Moore and Garg 1995)

Moore and Garg (1995) developed the SI based on the principles of biomechanics, work physiology, and the epidemiologically demonstrated associations between physical risk factors and work-related DUE MSDs. The SI does not predict the risk of developing a specific DUE disorder, but rather the summary score indicates the magnitude of work-related exposure to general DUE MSD risk factors. The SI was designed with the needs of professional ergonomists and OHS practitioners in mind (Moore and Vos 2005). Additionally, epidemiologists can use SI score and task-variable scores to systematically identify job tasks that expose workers to physical risk factors. Prior to the development of the SI, neither ergonomists nor epidemiologists had access to a standardized, objective, and comprehensive DUE risk assessment method or tool (Moore and Garg 1995).

Moore and Garg (1995) also established a dichotomous SI summary score classification criterion based on preliminary validations of the method within the pork processing industry. The developers suggested that jobs assigned an SI score equal to or greater than 5.0 should be classified as unsafe or hazardous. Jobs receiving SI scores less than 5.0 were classified as safe, meaning that they likely did not elevate worker risk of developing a DUE MSD. Garg *et al.* (2007) proposed raising the exposure classification criterion to an SI score of 6.1 based on the results of additional SI validations conducted within a variety of manufacturing industries. Alternatively, many SI users have employed a three-level hazard criterion (Drinkaus *et al.* 2003, Bao 2004, Bao *et al.* 2006, Spielholz *et al.* 2008, Bao *et al.* 2009b, Jones and Kumar 2010, Chiasson *et al.* 2012). The three-level criterion classifies SI scores less than 3.0 as safe, scores of 3.0 to 7.0 as possibly unsafe and worthy of intervention or further study (*i.e.* at an action-level of exposure), and scores greater than or equal to 7.0 as hazardous. Both the dichotomous and three-level SI hazard classification criteria can inform ergonomist and OHS practitioner-led primary, secondary, and tertiary DUE MSD prevention efforts. Further, the exposure classifications readily enable epidemiologists to model the etiologic associations between work and health outcomes.

The SI was designed to assess physical exposures inherent in mono-task (cyclic, single exertion) jobs. Jobs comprised of more than one type of exertion can prove difficult to rate using Moore and Garg's (1995) original instructions (Moore and Vos 2005, Garg and Kapellusch 2011). Some attempts have been made at introducing a multi-task scoring system (Drinkaus *et al.* 2005, Garg 2006, Bao *et al.* 2009b). However, the relationship between any of the proposed multi-task scoring systems and the SI criterion validity has not been adequately studied (Bao *et al.* 2009b, Garg and Kapellusch 2011). The absence of a clear multi-task or multiple-exertion scoring system is a primary limitation of the SI. Currently, SI analysts are instructed to assess the "overall" force requirements of a job when more than one exertion is required (Garg and

Kapellusch 2011). To reduce SI score variability, Moore and Garg (1995) recommend that teams of two or more analysts rate jobs together and resolve differences in scores by consensus.

Other limitations to the SI include no consideration of mechanical compression or localized vibration risk factors, and the SI requires more training and time to use than other pen-and-paper checklists (Moore and Garg 1995, Moore and Vos 2005, Chiasson *et al.* 2012). However, the SI is second only to OCRA Checklist and OCRA Index in its comprehensive consideration of risk factors to the DUEs and their interactive effects (Garg and Kapellusch 2011). Additionally, a number of studies have concluded that the SI reliability and validity is moderate to good (Takala *et al.* 2010).

2.3.2. *Evidence supporting SI criterion validity*

Ecologic studies. Along with the publication of the SI method, Moore and Garg (1995) reported the results of a preliminary SI validation study. As a team, the authors assessed 25 pork processing jobs for exposure to physical risk factors. After assessment, jobs were classified as hazardous if they were associated with any OSHA 200-log records of DUE disorders or symptoms during a 20-month retrospective period. The SI dichotomous hazard classifications (criterion=SI score≥5.0) independently classified 24 of the 25 jobs correctly. Moore and colleagues' later used the same job classification and SI analysis methods to evaluate 56 manufacturing jobs from multiple industries. In two separate studies, they reported that the SI score was significantly related (odds ratio [OR]=50, p=0.001; OR=106.6, p=0.001) to company-recorded DUE disorders and symptoms (Knox and Moore 2001, Rucker and Moore 2002). Garg *et al.* (2007) then reanalyzed these associations using the exposure and injury/illness data from all three ecologic studies (81 jobs total). The aggregate analysis found that the SI method correctly classified 75 of the 81 jobs as hazardous. The association between the SI exposure classifications and reported DUE disorders or symptoms remained very strong (OR=108.3, 95% confidence interval [CI]=16.7–705.0, $p<0.001$). Sensitivity of the SI was reported at 0.93; specificity at 0.89; positive predictive value at 0.90; and negative predictive value at 0.93.

Moore, Rucker, and Knox (2001) performed an inter-method reliability study using the combined exposure and job hazard data from the studies of Knox and Moore (2001) and Rucker and Moore (2002). The inter-method reliability study compared the predictive strength of the SI to that of generic physical risk factors for classifying job hazards. Generic risk factors included exposure to high force exertions, repetitive exertions, awkward posture, pinch grips, glove use, hand-arm vibration, localized compression, and cold temperatures. The SI was more strongly associated with reports of DUE symptoms and disorders than were any combination of generic risk factors (OR for SI=108.3, 95% CI: 16.7–705.0; OR for highest generic risk factor association=36.0, 95% CI: 4.3–303.4)

Due to limited resources, however, the authors of these SI studies were unable to gather the necessary data to adjust their analyses for confounding variables (Moore and Garg 1995, Knox and Moore 2001, Moore *et al.* 2001, Rucker and Moore 2002). Individual factors such as age, gender, work experience, medical history, psychosocial factors, body mass index (BMI), or hobbies could have influenced whether a DUE disorder or symptom was recorded (Rucker and Moore 2002, da Costa and Vieira 2010). Further, the majority of jobs analyzed were mono-task jobs, whereas many manufacturing job-exposures occur during multi-task work. Another limitation is the use of company records to classify a job as hazardous. If one of any number of workers doing a particular job reported a DUE symptom in the recent past, that job was considered hazardous. This “one-incident trigger” may overestimate which jobs truly are hazardous, and it may underestimate which jobs are not (Garg *et al.* 2007). Despite these limitations, the strength of results suggest that the SI can at least identify accurately repetitive manufacturing jobs that are associated with the development of DUE disorders.

Cross-sectional studies. Spielholz *et al.* (2008) recruited 567 workers from 9 manufacturing facilities and 3 healthcare facilities to participate in a cross-sectional study work-related DUE associations. The jobs evaluated during the study were not all mono-task jobs. Participants' DUE health was evaluated by an occupational health physician, nurse or physical therapist. Of the 567 workers, 86 (15.2%) were clinically diagnosed with a DUE disorder. The investigators used logistic regression to quantify the association between DUE morbidity and physical exposures measured by the SI. After adjusting for age, gender, and BMI, the odds of workers developing a DUE MSD in the dominant hand were 2.33 (95% CI: 1.20-4.53) times higher among jobs with SI scores greater than 7 compared to jobs with scores less than 3. A dichotomous hazard score threshold of 7 also was associated (OR 1.82, 95% CI 1.04-3.18) with dominant hand DUE MSDs. However, no significant association between SI scores and non-dominant hand DUE MSDs was observed.

Prospective longitudinal cohort studies. Garg *et al.* (2012) studied the capacity of the SI and the American Conference of Governmental Industrial Hygienists (ACGIH®) threshold limit value (TLV®) for hand activity level (HAL) to predict carpel tunnel syndrome (CTS) development in a prospective cohort. The 6-year study included 536 male and female workers from 10 diverse manufacturing facilities. The baseline prevalence of CTS was 10.3%, and observed CTS incidence over the study period was 2.55 per 100 person-years. Multivariate statistical models adjusted for CTS covariates revealed that workers doing jobs that received SI scores between 9.0 and 45.0 were at least three to five times more likely to be diagnosed with

CTS than workers who were not exposed to physical risk factors. When treating the SI score as a continuous variable, the authors observed a significant increase in risk (hazard ratio [HR]=1.131; 95%CI= 1.02–1.26) per unit of SI score increase until a score of 13.5 (HR=5.3), after which risk decreased gradually. The insignificant elevations in risk for jobs with SI scores of 60 or more may be due to survival bias (or the healthy worker effect). They may also be the result of SI analysts misclassifying force exposures for jobs that were not strictly mono-task jobs. When treating the SI score as a dichotomous variable, though, jobs with an SI score greater than 6.1 were 2.5 times as likely to be associated with incident CTS cases (95%CI: 1.00–6.13).

Gerr *et al.* (2010) prospectively studied the development of neck and upper extremity musculoskeletal symptoms and disorders over a 3-year period at an appliance manufacturing facility. The physical job exposures of 386 male and female workers were assessed using self-reports (work logs or journals), direct methods (electromyography), and observational methods (ACGIH® HAL, the SI, and Multimedia Video Task Analysis™). The incidence rate for hand-arm symptoms was 58 per 100 person-years, and for diagnosed hand-arm disorders the rate was 19 per 100 person-years. Adjusted multivariate associations between physical risk factors and DUE outcomes revealed that the SI scores (with an SI score exposure classification criterion of ≥5.0) were the best predictors of whether a worker would report DUE symptoms (HR=1.73, 95% CI: 0.99–3.04). The relationship was somewhat weaker for DUE disorders (HR=1.93, 95% CI: 0.85–4.40). However, adjusted models of individual exposure factors (i.e. force, repetition and posture) were weaker predictors of DUE symptoms and disorders than were the dichotomized SI summary scores. The study likely underestimates the incidence of DUE disorders because many of the workers were highly experienced (mean employment duration=16 years). The authors concluded that the SI summary score and its classification are useful and accurate measures of exposure to physical risk factors.

2.3.3. *Reliability of the SI*

Inter-rater reliability. Spielholz, et al. (2008) studied the inter-rater reliability of the SI. Four raters independently assessed a total of 125 repetitive, cyclic manufacturing and healthcare jobs. The investigators analyzed the reliability of SI scores and task variable scores. Average agreement for both pairs ranged from fair to moderate. *Intensity of exertion* estimates were the least reliable (Spearman r= 0.28; kappa=0.22), and the *speed of work* estimates were the most reliable (Spearman r=0.62; kappa=0.44). Strain Index scores categorized according to a three-tiered exposure classification criterion (≤3.0, 3.1–6.9, ≥7.0) were also moderately reliable (Spearman r=0.57; kappa=0.41). The authors note that the scores for the SI and three of the task variables (*duration of exertion*, *efforts per minute*, and *hand/wrist posture*) were significantly more reliable ($p<0.05$) for rater pairs who were experts than for pairs comprised of an expert and novice rater. However, three of the raters were experts and only one was a novice, and so generalizations of how rater experience affects rater reliability were limited.

Stevens *et al*. (2004) performed the first comprehensive inter-rater reliability study of the SI. The authors recruited and trained fifteen raters of varying expertise to evaluate a total of 73 different mono-task jobs from manufacturing, meat/poultry processing, and manual material handling industries. Raters used the SI individually and also in teams of three. The investigators used an agreement-based intraclass correlation coefficient, the ICC(2,1), to characterize inter-rater reliability. The ICCs for individual ratings of the task variables ranged from 0.66 to 0.81. For teams of raters, ICCs ranged from 0.48 to 0.93. Hand-wrist posture was the most variable for individuals and teams, whereas the other four variables received ICC scores of at least 0.77 (for individuals) and 0.81 (for teams). The overall SI scores were more variable for individuals (ICC=0.43, 95% CI: 0.25-0.70), but this variability was reduced for teams (ICC=0.64; 95% CI: 0.40–0.85). The authors concluded that the reliability of the SI was moderate to good, and that higher reliability would be achieved using teams of raters rather than individuals alone.

Intrarater reliability. Stephens, Vos, Stevens, and Moore (2006) recruited the same raters from Stevens's *et al*. (2004) study to reevaluate the same 73 jobs using the SI. Intraclass correlation coefficients were used to characterize the reliability of single rater scores over time. The authors found that four of the five task-variable ICCs indicated excellent reliability (ICC ≥0.90 for individuals, and ICC≥0.87 for teams). Estimates of *hand/wrist posture* reliability received ICCs of 0.82 for individuals and 0.66 for teams. Agreement between job exposure classifications using the same dichotomous criterion was also high. In this case, the authors used tetrachoric correlation coefficients, which describe the strength of association between dichotomous variables that are categorized from a continuous scale. Tetrachoric correlation values were 0.81 for individual raters and 0.88 for teams. Overall, the results suggested that the intrarater reliability for the SI is relatively high.

2.4. *Occupational Repetitive Action (OCRA) Checklist*

2.4.1. *Development and application*

The OCRA method is an observational assessment of physical exposure risks to the hand, wrist, arm, elbow, and shoulder (Colombini 1998, Occhipinti 1998). The OCRA methods include the OCRA Index and OCRA Checklist, both of which characterize exposure in terms of six primary task 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* (Occhipinti and Colombini 2005). The *additional risk factors* variable includes exposure to vibration, precision movements, glove use, mechanical compression, and cold stress (Occhipinti and Colombini 2005). Task-variable scores are primarily determined by measuring the percentage of time a worker is exposed to a multiplicity of physical parameters (*e.g.* frequency of repetitive action, intensity of force exertion, extreme angles of posture). Increasing percentages of exposure to these parameters correspond to increasing task-variable scores along a continuous scale. The *frequency*, *force*, *posture*, and *additional factors* scores are weighted by the scores for the *lack of sufficient recovery* and *work duration* variables, producing the OCRA Checklist or OCRA Index summary score. Table 2.2 shows the most recent summary score classification criteria provided by Occhipinti and Colombini (2011).

Table 2.2: OCRA Index and OCRA Checklist exposure categories

OCRA Checklist	OCRA Index	Exposure Level
≤7.5	≤2.2	None or Acceptable
7.6-11.0	2.3-3.5	Very Low
11.1-14.0	3.6-4.5	Light
14.1-22.5	4.6-9.0	Medium
≥22.6	≥9.1	High

From Colombini *et al.* 2011

The OCRA Index was the first of the two OCRA methods developed, and its developers modeled the Index based on the design of the NIOSH Lifting Equation (LE). Occhipinti (1998) writes that the OCRA Index and NIOSH LE characterize physical exposures based on a wide range of risk factors, and elevated risk exposure is represented by a summary score (or concise index). The concise exposure index is reported in terms of the most critical task variable, which for the NIOSH LE is the weight of material lifted (Waters *et al.* 1993, 1998). For the OCRA

Index, the most critical task variable is *frequency of technical actions*, and so the OCRA Index summary score is a ratio of the total number of technical actions performed during the work shift compared to the recommended number of technical actions for that work shift (Occhipinti 1998). The default number of recommended technical actions is 30 per minute, which is adjusted downward based on the exposure to the other five task variables. Occhipinti (1998) and Colombini (1998) chose 30 repetitions per minute as the default based on both epidemiologic studies and expert ergonomic opinion.

The OCRA Index was designed to be used by occupational health and safety specialists, ergonomists, time and methods analysts, and production engineers (Colombini 1998, Occhipinti 1998). The Index is a very thorough assessment requiring slow-motion video analysis and in-depth time-motion study, and the evaluation of a single job may require 45 minutes or more (Occhipinti and Colombini 2005, Chiasson *et al*. 2012). In 2000, Colombini, Occhipinti, Cairoli and Baracco introduced the OCRA Checklist as a simplified version of the OCRA Index. The OCRA Checklist was designed to be completed using pen and paper at the worksite, taking ten to fifteen minutes to evaluate one job with a 30 second work-cycle time (Colombini *et al.* 2000, Occhipinti and Colombini 2005). In developing the OCRA Checklist, Colombini *et al.* (2011) reduced or eliminated some of the direct measurement and time-study requirements of the OCRA Index. Those eliminated scoring items were replaced with exposure scoring schemes based on expert estimation using detailed verbal descriptors. Despite the simplifications, the Checklist remains a comprehensive and systematic exposure assessment tool, and it is more similar to other observational methods such as the SI than is the OCRA Index.

Since its release, many European job analysts have adopted the OCRA Checklist (Colombini *et al.* 2011), Occhipinti and Colombini 2007). In 2011, its developers updated the exposure scoring system and clarified the verbal descriptors, and they also created several software-based versions of the Checklist available for free download (Colombini *et al.* 2011). Updated electronic and pen-and-paper versions of the Checklist are available in English and Italian languages as Microsoft Excel Workbooks on the Ergonomics of Posture and Movement Website (EPM 2012).

As an observational measure designed to assess exposure to all relevant UE disorder physical risk factors, the OCRA Checklist is relatively complex compared to similar methods. This is because determining exposure to the six main task variables requires the analyst to evaluate 14 separate task elements. For instance, the *awkward posture and movements* task variable is calculated based on separate assessments of five task elements: shoulder posture, elbow and forearm movements, hand and wrist posture, grip quality, and lack of variation in typically assumed postures and movements. Further, understanding the concept of technical actions, which is the theoretical foundation of the task variable *frequency of repetition,* may be difficult if the user is unfamiliar with Barnes' (1968) motion-time-measurement method (Occhipinti and Colombini 2005). The OCRA Checklist does provide verbal descriptors to assist with technical action frequency estimation, but a complete understanding of the concept remains necessary to the correct use of the Checklist in general. Occhipinti and Colombini (2005) suggest that OCRA Checklist or Index training be administered over the course of two eight-hour sessions, followed by an eight-hour refresher course. The intensive training requirements are a primary limitation to the OCRA method. Additional limitations include little to no consideration of psychosocial risk factors for MSDs, longer evaluation time requirements compared to similar measures, and a lack of published research supporting the validity and reliability of the Checklist (Occhipinti and Colombini 2005, Takala *et al*. 2010).

2.4.2. *Validity and Reliability*

Concurrent validity of the OCRA Checklist. To evaluate the criterion validity of the OCRA Checklist, its developers performed an inter-method reliability study between it and the OCRA Index. They assessed 46 different manufacturing jobs that had previously been assessed using the OCRA Index (Colombini *et al.* 2000). The relationship between the Index and Checklist summary scores was very well described by an exponential regression function (R^2=0.98, p<0.0001). The authors used this exponential relationship to establish the Checklist exposure score classification criterion (Table 2.2). No further studies of the OCRA Checklist validity have been conducted.

Cross-sectional studies of the OCRA Index. Grieco (1998) summarized eight studies of the association between OCRA Index scores and diagnosed work-related UE MSDs. In total, the OCRA Index scores and UE disorder prevalence associated with 462 exposed workers were compared to a control sample of 749 unexposed workers. The entire control sample was assigned an OCRA Index summary score of 0.5, indicating no physical exposure. The 462 exposed individuals performed one of seven different jobs from poultry processing, automotive parts manufacturing, ceramics finishing, and vegetable packing industries. Trained analysts assessed the 462 job exposures according to the instructions of Occhipinti (1998) and Colombini (1998). Greico used log-transformed, mean OCRA Index in each of the seven jobs scores as the independent variable in a linear regression model. The dependent variable was diagnosed UE MSD prevalence. The model revealed a significant association (R^2=0.88, p=0.0002) between OCRA Index summary scores of 4.0 or greater and increased UE MSD prevalence. Possible confounding factors were not controlled for because that data was not available.

Occhipinti and Colombini (2007) enhanced Grieco's analysis by pooling the UE MSD prevalence and OCRA Index exposure data from 16 additional groups of jobs. The sample of exposed workers was enlarged to 4624, and the same control sample of 749 workers was maintained. Occhipinti and Colombini employed a simple linear regression model to quantify the association between Index scores and UE prevalence. The association was strong (R^2 adjusted=0.92, $p<0.0001$), and the model was used to categorize jobs according to the five-level exposure classification criteria. Contingency tables and Mantel–Haenszel χ^2 tests were used to calculate ORs and 95% CIs for the exposure classification and disease associations. The results are displayed in Table 2.3.

Table 2.3. Odds of association between task with one or more UE disorder diagnoses and OCRA index exposure

OCRA Index	Odds Ratio	95% CI
≤2.2	-	-
2.3-3.5	2.16	(1.45, 3.23)
3.6-4.5	3.74	(2.38, 5.89)
4.6-9.0	5.30	(3.63, 7.78)
≥9.1	24.31	(15.99, 37.11)

From Occhipinti & Colombini 2007

The authors of these two validation studies admit that their results are limited by the study design (Grieco 1998, Occhipinti and Colombini 2007). Both are cross-sectional studies that rely on simple regression models to estimate the association between OCRA Index scores and the prevalence of one or more UE disorders. The models do not control for the effect of possible covariates, such as age, gender, work experience, medical history, psychosocial factors, BMI, or hobbies. At best, the two cross-sectional studies provide moderate assurance that the OCRA Index is an accurate estimate of exposure (Takala *et al.* 2010).

Reliability of the OCRA Checklist and OCRA Index. No studies of the inter- or intrarater reliability of the Checklist or the Index have been performed (Occhipinti and Colombini 2005, Takala *et al.* 2010). Occhipinti and Colombini (2005) admit that the lack of reliability data is a weakness of the OCRA methods.

2.5. *Strain Index and OCRA Checklist inter-method reliability studies*

2.5.1. *Commonly compared upper extremity observational exposure assessments.*

Takala's *et al.* (2010) comprehensive review of observational measures includes an overview and evaluation of nine upper limb exposure assessments. The methods include: Health and Safety Executive (HSE) Assessment of Repetitive Tasks (ART) method; Rapid Upper-Limb Assessment (RULA); Stetson's checklist for the analysis of hand and wrist; Keyserling's cumulative trauma checklist; Ketola's upper-limb expert tool; SI; ACGIH® TLV® for HAL; OCRA methods; and Washington State ergonomic checklists. However, about half of these methods have not demonstrated strong associations with MSDs (Takala *et al.* 2010). Only the SI, ACGIH® TLV® for HAL, RULA, and OCRA methods have been associated with MSDs through cross-sectional or cohort-based studies (Takala et al., 2010). For this reason, most inter-method reliability studies include one of these four methods.

2.5.2. *Exposure classification criteria for SI and OCRA Checklist comparisons*

With the exception of postural assessments, observational measures assign job-exposure scores along a variety of different scales. A common exposure score classification scheme must be used to compare the results of different observational methods (Chiasson *et* al. 2012, Drinkaus *et al.* 2003, Jones & Kumar 2010). Several combinations of exposure classification criteria have been used to compare the SI and the OCRA methods with one another and other methods. Comparisons are usually made based upon the verbal descriptions of each exposure category provided by the developers of each method, such as “safe,” “action-level,” or “hazardous” (Moore and Garg 1995, Colombini 1998, Occhipinti 1998). Table 2.4 displays the exposure classification criteria used by several investigators of the inter-method reliability of the SI or OCRA methods.

Table 2.4. Exposure classification criteria used for inter-method comparisons between the SI or OCRA methods

Authors	SI	OCRA Checklist	OCRA Index
Moore, Rucker & Knox (2001)	<5, ≥5		
Garg *et al.* (2011)	<6.1, ≥6.1		
Jones & Kumar (2010)	<3, 3-6.9, ≥7		≤.75, .76-3.9, ≥4*
Chiasson *et al*. (2012)	<3, 3-6.9, ≥7		≤1, 1.1-3.9, ≥4*
Apostoli *et al*. (2004)	<3, 3-7, ≥7.1	≤6, 6.1-11.9, ≥12*	
Ferreira *et al*. (2009)	<3.1, 3.1-6.9, 7-10, ≥10.1	≤11, 11.1-14, 14.1-22.5, ≥22.6	≤3.5, 3.6-4.5, 4.6-9.0, ≥9.1
Colombini *et al*. (2011)		≤7.5, 7.6-11, 11.1-14, 14.1-22.5, ≥22.6	≤2.2, 2.3-3.5, 3.6-4.5, 4.6-9, ≥9.1

*OCRA Checklist or OCRA Index exposure classification criteria based on Grieco’s (1998) original exposure score rankings rather than on the updated criteria of Colombini *et al.* (2011).

To ensure the best comparisons between methods, the common exposure criteria should be based upon the results of relevant validity studies (Armstrong *et al.* 1992). Unfortunately, this is not feasible for the SI and OCRA Checklist given the differences in investigator approaches to establishing criterion validity. For instance, Spielholz’s *et al.* (2008) cross-sectional study of DUE associations with SI scores revealed that gender, age, and BMI were all significant confounding variables. And high quality prospective cohort studies of SI associations with DUEs demonstrated that the effect of confounding variables further diminish the predictive power of the observational assessment (Gerr 2010, Garg *et al.* 2012). In contrast, the cross-sectional studies of the OCRA Index have relied on unadjusted or crude associations between UE disorder development and mean OCRA scores (Grieco 1998, Occhipinti and Colombini 2007). Models that do not include covariates generally support stronger associations than those of unadjusted models (Speilholz *et al.* 2008).

2.5.3. Strain Index and OCRA Checklist or OCRA Index comparisons

Jones and Kumar (Jones and Kumar 2010) assessed the physical exposures of 87 individuals performing one of four sawmill jobs at four different facilities. The exposure assessments were made using the RULA, REBA, ACGIH® TLV® for HAL, SI, and OCRA Index. In an effort to increase the accuracy of the assessment, the authors chose direct methods (electromyography and electrogoniometry) to quantify most of the physical parameters that are normally measured observationally, such as repetition frequency, duration of exertion per task cycle, and hand, wrist, and shoulder posture. Worker reported Borg CR-10 estimates of force exertion were also obtained. The summary scores for the five observational measures were classified according to three-level exposure classification criteria. Using contingency tables to estimate percentage of overall agreement between methods, the SI was found to agree best with RULA (98% agreement) and least with the ACGIH® TLV® for HAL (45% agreement). The OCRA Index agreed with the SI best (83% agreement) and least with the ACGIH® TLV® for HAL (48% agreement). For 15 individuals working a single job (saw filer) at four different facilities, Jones and Kumar (2007) evaluated the agreement between exposure classifications and the incidence of successful workers' compensation claims. The authors retrospectively analyzed compensation claims for saw filer workers at each of the four facilities. Claims rates differed between facilities, and Jones and Kumar (2007) reported that the SI and OCRA Index appeared to be sensitive to such a difference, whereas the other three methods were not. The SI and OCRA Index classified more of the workers as highly exposed at the facilities reporting higher incidence of claims rates, but the trend was not statistically significant.

Jones and Kumar (2007, Jones and Kumar 2010) recognized that the results of their inter-method reliability studies were limited by the fact that nearly all 87 workers were classified as highly exposed to physical risk factors (97% of SI scores were high hazard, and 84% were for the OCRA Index). Further, only four jobs were assessed, and the investigators used the observational methods to evaluate individual rather than job-level exposures. Moore and Garg (1995) wrote that the SI should only be used to evaluate jobs, not individuals. Occhipinti and Colombini (2005) also encouraged OCRA Index and Checklist analysts to base their exposure assessments on samples of multiple workers performing each job.

Apostoli, Sala, Gullino, and Romano (2004) compared the hazard classifications of four observational methods: the OCRA Checklist, the SI, the ACGIH® TLV® for HAL, and the OREGE tool. The four methods were used to estimate general UE MSD physical exposures among manufacturing workers performing twelve different material assembly and packaging jobs. Of the twelve jobs, the SI classified two as low hazard, four as moderate hazard, and six as high hazard. The OCRA Checklist classified nine jobs as moderate hazard, and three as high hazard. The percentage of overall agreement between OCRA Index and SI hazard classifications was 58.3%. Generalization is made difficult based on the small sample size.

The United Kingdom's HSE has recently developed an observational assessment called the ART tool (Ferreira *et al.* 2009). During the tool's development, the HSE compared the ART tool to several other observational measures, including the SI and the OCRA Checklist. Teams of three ergonomists evaluated videos of 10 different jobs, and by consensus determined summary exposure scores that were then classified according to four-level exposure criteria. Seven of the ten jobs were classified as low or high risk by both the SI and the OCRA Checklist (70% overall agreement). Jobs that were differentially classified were assigned to one of the two moderate hazard classifications by either tool. Similar to Apostoli *et al.* (2004), the small sample size prohibits generalizing from these results.

Chiasson *et al.* (2012) applied the SI and OCRA Index to a 167 different workstations across a variety of industries. Several teams of graduate students with occupational ergonomics training applied each of the observational assessments. Strain Index and OCRA Index summary scores were categorized according to three-level exposure classification criteria similar to that of Jones and Kumar's (2010) inter-method reliability study. Unlike Jones and Kumar, Chiasson *et al.* (2012) calculated OCRA Index summary scores without including task-variable scores associated with physical exposure to shoulder disorder risks. Because the SI cannot identify exposures relevant to shoulder disorders, Chiasson *et al*. expected that omitting shoulder exposure scores from the OCRA Index would permit a more appropriate comparison between the two observational methods. The authors also selectively compared hazard ratings from the ACGIH® TLV® for HAL, various forms of the QEC, the RULA, the REBA, the Finnish Institute of Occupational Health (FIOH) MSD tools, and the European Standard recommended force limits for machinery operation (EN 1005-3).

Chiasson *et al.* (2012) found that the OCRA Index exposure classifications agreed with the SI 60.1% of the time. Differential exposure classification was mostly due to a single-level of exposure misclassification (26.2% of workstations), but 13.7% of workstations were classified as safe by one method and hazardous by the other. A Pearson correlation coefficient, which represented the strength of association between the continuous SI and OCRA Checklist summary scores, was relatively low (r=0.32, $p<0.0001$). This was better than the insignificant correlation between the OCRA Index and a version of the Quick Exposure Checklist (QEC) relevant only to DUE exposures (r=0.03, $p>0.05$). While the OCRA Index did correlate best with the SI, the SI correlated best with the ACGIH® TLV® for HAL (r=0.69, $p<0.0001$). The OCRA Index and ACGIH® TLV® for HAL exposure classification correlation was low, but statistically significant (r=0.16, $p<0.05$).

In addition to comparing exposure classifications, Chiasson *et al.* (2012) reported that using the OCRA Index to rate 167 workstations required a mean analysis time of 70 minutes (standard error [SE]=28.9). The same workstations were rated in a mean 16 minutes (SE=6.49) using the SI. However, the authors used time-motion study of video collected at each workstation to identify work-cycle durations, duration of exertions per cycle, and repetition frequency. Time-motion study required a mean 170 minutes to complete (SE=92.19) for each workstation. Excessively long assessment times may be due to the abnormally long work-cycle times of some jobs; 65 workstations exhibited work-cycle times longer than 73 minutes, and two work-cycles lasted 450 minutes. The SI and OCRA Index may not yield accurate exposure assessments when jobs are not repetitive or cyclic (Moore and Garg 1995, Occhipinti and Colombini 2005). Further, the inter-method reliability results may be biased because just one time-motion study of each workstation was used to collect frequency/repetition and duration data. This approach may have artificially inflated the correlation and agreement statistics between the two assessments compared to individually rating all task variable parameters separately with each method.

Finally, Chiasson *et al.* does not clearly state whether the167 workstations assessed with the SI and OCRA corresponded to 167 different jobs or just to 167 individual workers.

3. Methods

3.1 *Jobs Rated*

Physical exposure data for 21 jobs was recorded with digital video cameras at a cheese manufacturing facility in Sardinia, Italy. The facility processes 40 megaliters of sheep milk annually, 65% of which is converted to *Pecorino Romano*, a hard cheese produced in 25 and 35 kilogram wheels. The 21 tasks evaluated represented all of the major stages of *Pecorino Romano* production, from draining curd to labeling and packaging the final product. Of the company's 130 permanent employees, 45 (44 male, 1 female) worked on the *Pecorino Romano* production line. Previous collaboration between this cheese production facility and the University of Sassari occupational health specialists revealed that employees were experiencing a variety of UE MSD symptoms. These jobs were chosen for assessment in the present study because of their reported association with UE MSD symptoms, their repetitive and cyclical character, and their varying levels of complexity. Further, these jobs generally exposed the entirety of the workers' upper extremity to physical risk factors; exposures did not appear to affect the shoulder region predominantly. This feature is relevant because the SI does not assess physical exposures to the shoulder region whereas the OCRA Checklist does. Low inter-method reliability might be expected if the jobs assessed were predominantly exposing workers to shoulder disorder risks.

University of Sassari occupational health specialists recorded the 21 jobs using handheld digital video camcorders, capturing upper extremity activity from sagittal and frontal perspectives. Video recordings captured a minimum of five work-cycles. Job and break/recovery duration data were identified by direct observation and interviews at the facility. Employees consented to being recorded, and Institutional Review Board approval was obtained from the University of Sassari.

3.2 *Raters*

Eight individuals were recruited to evaluate the 21 cheese production jobs using the SI and OCRA Checklist. Raters included university faculty from occupational health programs and graduate students in ergonomics. At least one rater was an experienced user of either method, having assessed 100 or more exposures in the field. Prior to participating in the present study, four of the eight raters had observed some or all of the 21 jobs in person at the cheese-production facility.

All raters received method-specific training on the SI and OCRA Checklist prior to evaluating any of the cheese production jobs. Strain Index training was administered separately from OCRA Checklist training. Training occurred at least several weeks prior to cheese production job evaluation using either method. 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. Trainees continued practicing until they consistently assigned exposure ratings that were similar to those of an experienced rater.

3.3 *Exposure Assessments*

3.3.1. *General procedure*

Each rater assessed physical exposures during cheese production jobs based on the 21 video recordings. Raters were provided with digital copies of the 21 video recorded jobs, Microsoft Excel worksheets of the SI and OCRA Checklist, and general instructions for job rating procedures. Each rater was also provided with a stopwatch, tally counter and software enabling slow and normal speed video playback. Raters were instructed to assess the videos in alphabetic order according to the job name. Job names bore no inherent relationship to any of the physical exposure parameters, work-cycle time, or job complexity.

Raters analyzed each job using the SI and OCRA Checklist separately. Three of the raters completed SI assessments first, while the other five raters administered the OCRA Checklist first. Repeat measures using either method were not made. Raters waited at least three weeks after completing the first exposure assessment before beginning the next. Rater assessments were completed individually, and raters did not communicate the results of their exposure assessments with one another until ratings were completed. The job duration time and the distribution of work-rest cycles were provided to raters based on company personnel interviews and direct field observation. These times were used to determine the score for the SI *task duration per day* variable and the scores for the OCRA Checklist *lack of sufficient recovery* and *task duration* variables. The scores for these task variables could not be calculated from the 21 video files alone.

3.3.2. *Strain Index procedure*

Raters used Excel software to record exposure assessment scores for the cheese production jobs according to the original SI method developed by Moore and Garg (1995). Raters measured or estimated exposure levels to all of the task variables except for *duration of task per day*. When raters recorded the task-variable scores into the Excel spreadsheet, the software automatically calculated the summary SI exposure score. The print version of the SI used in the present study is attached as Appendix I.

The *intensity of exertion* variable was estimated using both the Borg CR-10 scale (Borg 1982) and the verbal descriptors (*e.g.* “light,” “somewhat hard,” “hard,” *etc.*) provided by Moore and Garg (1995). The *duration of exertion* and *efforts per minute* variables were directly measured using a stopwatch and counter. The *hand/wrist posture* and *speed of work* variables were estimated based on Moore and Garg's (1995) verbal descriptors (e.g. "good/near-neutral," "fair/non-neutral," "bad/marked deviation," *etc.*) and posture criteria (*i.e.* ranges of wrist flexion, extension and radial/ulnar deviation).

3.3.3. *OCRA Checklist procedure*

Raters administered the OCRA Checklist according to Occhipinti and Colombini's instructions (Occhipinti and Colombini 2001, Colombini *et al.* 2011). Raters used an Excel version of the OCRA Checklist based on Colombini's *et al.* (2011) most recent updates to the method. Colombini's *et al.* (2011) Checklist was originally developed in Italian, and various English translations were provided on the EPM website (EPM 2012). For the present study, bilingual English and Italian (native language) speakers familiar with the OCRA Checklist ensured that the English translations of verbal descriptors and rating instructions were consistent with those of the 2011 Italian version. Similar to the SI Excel spreadsheet, the OCRA Checklist summary exposure score was automatically calculated as the raters recorded scores for the task-variables. The print version of the Checklist used in the present study is attached as Appendix II.

Raters used a stopwatch, verbal descriptors, and physical criteria to evaluate the percent time exposed to the relevant OCRA parameters for the *frequency*, *force*, and *awkward posture and movements* task variables. Raters determined scores for the task variables, which ranged from 0 up to 32 on a continuous scale. Raters used a counter and stopwatch to assess technical actions per minute, and used the verbal anchors of the Checklist to assign dynamic technical action frequency scores. Technical action criteria were based on the definitions and examples provide by the OCRA method developers (Occhipinti and Colombini 2001, Colombini *et al.* 2002). Raters estimated force intensity using the Borg CR-10 scale (Borg 1982) and verbal descriptors. Estimates for the *additional factors* and *awkward postures* task variables were determined using verbal descriptors. For greater clarification on the verbal descriptors associated with all of the task variables, refer to Appendix II.

3.4. ***Data Analysis and Statistical Methods***

3.4.1. *Exposure data*

Strain Index and OCRA Checklist summary scores, which are continuous variables, are not directly comparable. However, the developers of both methods categorize summary scores according to similar exposure classification criteria. These are also referred to as hazard cut-points or thresholds, and they indicate the magnitude of exposure to physical risk factors during job completion. Common categorical exposure classifications were used to perform inter-method reliability analyses. The exposure classifications chosen for inter-method reliability analyses are shown in Table 3.1, and they were similar to those used in previous SI and OCRA Checklist studies (Apostoli *et al.* 2004, Garg *et al.* 2007, Spielholz *et al.* 2008, Jones and Kumar 2010, Colombini *et al.* 2011, Chiasson *et al.* 2012, Garg *et al.* 2012).

Table 3.1. Strain Index and OCRA Checklist exposure classification criteria used for inter-method comparisons.

Dichotomous Exposure Criteria			Three-Level Exposure Criteria		
SI		OCRA	SI		OCRA
<6.1	Safe	<11.1	<3	Safe	<7.6
≥6.1	Hazardous	≥11.1	3-6.9	Action*	7.6-14
			≥7	Hazardous	≥14.1

*Action-level indicates a magnitude of exposure that may elevate a worker's risk of developing an upper extremity disorder, and employer action should be taken to reduce the exposure.

The 21 cheese production jobs were reviewed *a priori* to categorize the upper-limb activity during the work-cycle as symmetric or asymmetric. Asymmetric jobs were those that required the worker to complete exertions and movements with the left limb that were distinctly different in quality of magnitude from the actions of the right limb. In contrast, symmetric jobs were defined as those in which the left and right limbs performed similar actions during the work-cycle. For symmetric jobs, mean task-variable and summary exposure scores were used in the reliability analyses, and these jobs were represented by a single exposure identifier (*e.g.* E01). For asymmetric jobs, scores assigned to each limb were analyzed independently, represented by an exposure identifier for each limb (*e.g.* E02 for the left limb and E03 for the right limb). In total, 11 jobs were classified as asymmetric and 10 were classified as symmetric, thus permitting the use of 32 upper extremity exposures in the reliability analyses.

Descriptive statistics (mean, median, standard deviation, and range) for the 32 job exposures across all raters were calculated for the SI and the OCRA Checklist summary scores. Descriptive statistics were also provided for individual raters across all job exposures. Standard tests of normality and for homogenous variance were performed. Analysis of variance (ANOVA) identified group level differences between job exposure scores and mean rater scores. Individual rater differences were identified using a two-way ANOVA with a Tukey-Kramer multiple comparison adjustment.

3.4.2. *Inter-method reliability analysis*

Inter-method reliability analyses of SI and OCRA Checklist exposure classifications were performed using median and raw rater scores. Additionally, reliability analyses were performed for groups of raters or jobs stratified according to possible covariate effects (job symmetry, job complexity, rater experience, and rater profession/education).

A combination of statistics were used to characterize the inter-method reliability: proportion of overall agreement, McNemar's or Bowker's test of symmetry, Spearman's rank-order coefficient, and Cohen's simple and weighted kappa coefficient. Tests of symmetry were performed to evaluate whether rater disagreements were significantly clustered above or below the main diagonal of the contingency table. A significant symmetry test ($p<0.05$) indicated that a rater tended consistently to assess exposures as more hazardous when using either the SI or the OCRA Checklist. Spearman correlations represented the strength of association between inter-method exposure classifications. Kappa coefficients were calculated to characterize chance-corrected agreement. Weighted kappa was calculated using quadratic weights, also referred to as Fleiss-Cohen weights. Chi-square tests were used to evaluate the equivalency of kappa coefficients between individuals and groups of raters. Landis and Koch's (1977) verbal criteria of kappa coefficient magnitude were used: $\kappa < 0.20$, poor or slight agreement; $0.21 \leq \kappa \leq 0.40$, fair agreement; $0.41 \leq \kappa \leq 0.60$, moderate agreement; $0.61 \leq \kappa \leq 0.80$, substantial agreement; and $\kappa > 0.80$, almost perfect agreement. Confidence intervals for kappa coefficients were obtained, and their lower limits were interpreted in terms of Landis and Koch's verbal criteria.

3.4.2. *Inter-rater reliability analysis*

The eight raters and the 32 upper extremity exposures were considered to be random samples selected from a larger population of possible raters and exposures. As such, variance parameters for the SI and OCRA Checklist task-variable scores, summary scores, and exposure classifications were estimated using a two-way analysis of variance (ANOVA) with random effects model (Shrout and Fleiss 1979).

Inter-rater reliability was measured using an intraclass correlation coefficient (ICC). The ICC was a single-measure, agreement-based version, which Shrout and Fleiss (1979) labeled as ICC(2,1). This is the most common form of the ICC for general rater reliability studies of observational ergonomic methods (Ebersole and Armstrong 2002, Stevens *et al.* 2004, Stephens *et al.* 2006, Bao *et al.* 2009a, Xu *et al.* 2011, Dockrell *et al.* 2012), regardless of whether data are continuous or categorical variables (Berk 1979, Maclure and Willett 1987, Streiner and Norman 2008). Confidence intervals for each reliability coefficient were estimated using the ICC(2,1) standard error formula (α=0.05) presented by Streiner and Norman (2008).

Verbal criteria used to interpret ICCs were based on Fleiss's (1986) recommendations: $\rho < 0.40$, poor reliability; $0.40 \leq \rho \leq 0.75$, fair to good reliability; and $\rho > 0.75$, excellent reliability. Lower confidence interval limits were interpreted in terms of Fleiss's verbal criteria.

Inter-rater reliability of categorical exposure classifications was also evaluated in terms of the average proportion of agreement for the 32 analyzed cheese production exposures. This was a measure of total agreement divided by total number of exposures rated.

All statistical analyses were completed using SAS/STAT® Software (SAS Institute, Cary, NC) version 9.3 (2012) on a Windows PC. Confidence intervals and hypothesis tests were conducted with $\alpha = 0.05$.

4. Results

4.1 *Descriptive statistics and summary of data*

4.1.1. *Study Jobs*

Jobs were predominantly machine-paced and represented a spectrum of repetitive upper extremity activity, with work-cycle times ranging from 6 to 106 seconds (mean=41.5, SD= 31.2). *A priori* descriptions of the 21 videos were made, identifying the symmetry of upper extremity technical actions during the work-cycle, the number of tasks comprising each job, and the overall quality of each video recording. Table 4.1 shows the job symmetry and job complexity for each job and its corollary exposure identification number.

- Eleven of the 21 jobs were characterized by asymmetric upper limb technical actions. Thus, 22 of the exposures assessed were left and right limb exposures for same 11 jobs. The left and right limbs of the 10 symmetric jobs were assessed by all raters, but the analyses of inter-method and inter-rater reliability were based on the average of the left and right limb exposure scores.
- Six jobs were comprised of a single task, nine jobs were comprised of two separate tasks, and the remaining six jobs were comprised of three separate tasks. These jobs were labeled as mono-task, dual-task, or multi-task jobs.

4.1.2. *Raters*

Three raters were PhD university faculty from occupational health programs, and five raters were graduate students in ergonomics. The majority of raters were novice users of either the SI or the OCRA Checklist, but all were familiar with the principles of observational exposure assessment prior to participating in the present study. Three raters were considered experienced users, having assessed at least 100 exposures in the field using either the SI or OCRA Checklist. Table 4.2 lists each rater's profession/education and experience. For the present study, rater training using the OCRA Checklist required more than twice as much time as SI training required (10 hours compared to 4 hours). All raters assessed all possible jobs using both the SI and the OCRA Checklist.

Table 4.1. Descriptions of cheese processing jobs and upper extremity exposures.

Exposure ID	Job description	Job symmetry	Job complexity
*E01	Bagging cheese	Symmetric	Dual-task
**E02, E03	Boxing grated cheese	Asymmetric	Dual-task
E04, E05	Coating cheese	Asymmetric	Mono-task
E06, E07	Collecting plastic molds	Asymmetric	Dual-task
E08	Draining/pressing curd	Symmetric	Mono-task
E09, E10	Dressing cheese	Asymmetric	Dual-task
E11, E12	Entry of cheese in salt machine	Asymmetric	Dual-task
E13, E14	Entry washing for salt machine	Asymmetric	Multi-task
E15, E16	Exit washing for salt machines	Asymmetric	Multi-task
E17	Insert marking band	Symmetric	Multi-task
E18, E19	Label application	Asymmetric	Multi-task
E20	Loading metal molds	Symmetric	Mono-task
E21, E22	Loading salt board	Asymmetric	Multi-task
E23	Removing metal molds	Symmetric	Mono-task
E24	Removing plastic molds	Symmetric	Mono-task
E25	Stacking cheese	Symmetric	Mono-task
E26	Turning cheese	Symmetric	Dual-task
E27	Vacuum bagging	Symmetric	Dual-task
E28, E29	Washing cheese	Asymmetric	Multi-task
E30, E31	Washing sheets	Asymmetric	Dual-task
E32	Wrapping plastic molds	Symmetric	Dual-task

*Identifies mean of left and right limb exposures from symmetric jobs.
**Identifies left and right limb exposures separately from asymmetric jobs.

Table 4.2. Rater profession/education and experience with the SI and OCRA Checklist.

Rater ID	Profession/ education	OCRA Experience	SI Experience
A	PhD Student	Novice	Novice
B	*CPE, PhD Faculty	Novice	Novice
C	MS Student	Novice	Novice
D	PhD Student	Novice	Novice
E	CPE, PhD Faculty	Novice	Experienced
F	PhD Faculty	Novice	Novice
G	MS Student	Novice	Experienced
H	PhD Student	Experienced	Novice

*CPE, Certified Professional Ergonomist

4.1.3. *Exposure Assessment Scores*

Visual inspection of the data and tests of residuals revealed that continuous SI and OCRA Checklist exposure scores were neither normally distributed nor did they exhibit homogeneous variance. However, the present study is balanced (all raters rated all possible exposures using both methods). The results of ANOVA analyses are fairly robust to violations of normality and homoscedasticity when using a balanced study design. Exposure scores were transformed using a square-root or cube-root function, and the resulting data did not violate the ANOVA assumptions. Parallel inter-method and intra-method analyses were made using transformed and untransformed data. Transforming the data did not affect the interpretation of results significantly, therefore all reliability analyses were performed using untransformed data. However, transformed summary exposure scores were used to conduct post-hoc multiple comparison testing among mean rater scores. The transformed scores revealed more significant differences between raters than did the untransformed scores.

The overall mean SI score was 22.8 (SD=28.75), the overall median SI score was 10.1, and the range of SI scores was from 0.1 to 161.9. The overall mean OCRA Checklist score was 19.4 (SD=14.7), the overall median Checklist score was 15.7, and the range of Checklist scores was from 0.0 to 76.7. One-way ANOVA and Kruskal-Wallis tests revealed that significant differences between exposures scores ($p<0.0001$) were observed for both the SI and OCRA Checklist scores. Figures 4.1(a) and 4.1(b) display the distribution of SI and OCRA Checklist scores according to the 32 exposures rated. The figures highlight that mean scores across all tasks are generally higher than median scores.

Greater variance in exposure scores was observed for SI assessments than for OCRA Checklist assessments. Further, the majority of mean exposure scores were greater than 7.0 for the SI and 14.1 for the OCRA Checklist (62.5% for the SI, 53.1% for the Checklist). The SI scores for exposures E10, E14, and E29 were associated with the greatest score ranges and variance ($\text{range}>100.0$, $SD\geq41.8$). For the OCRA Checklist, exposures E09, E10, E26, and E29 were associated with the greatest score ranges and variance ($\text{range}>51$, $SD\geq11.0$). For both methods, all but one of these exposures with high ranging scores was associated with asymmetric tasks.

Figures 4.2(a) and 4.2(b) display the distribution of SI and OCRA Checklist scores according to each rater. Variance in SI scores was greatest for rater F, and the least variable SI scores were from rater C. Variance in OCRA Checklist scores was greatest for rater C, and the least variable OCRA Checklist scores were from rater H.

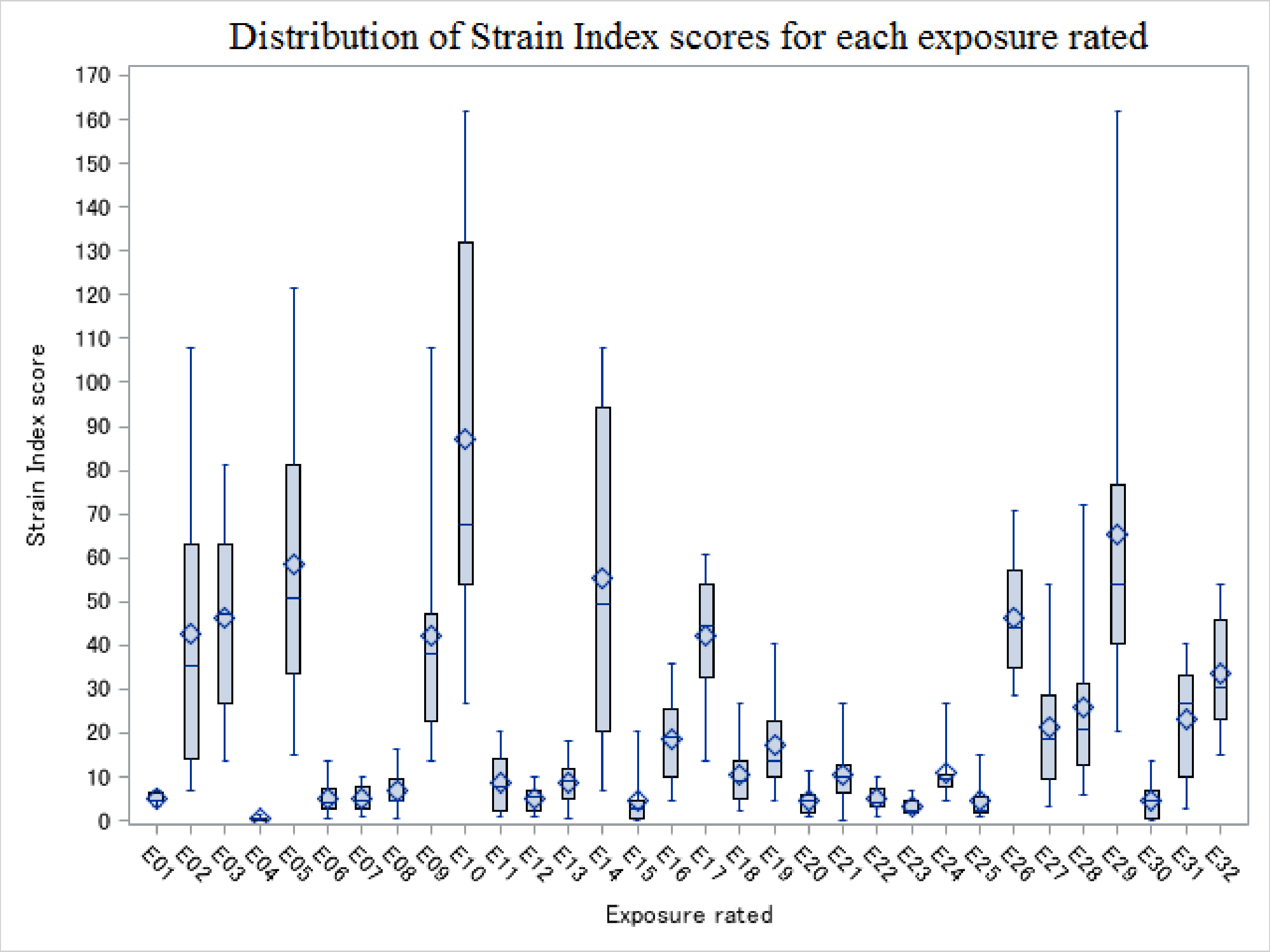


Figure 4.1(a). Distribution of raw SI scores by exposure rated. Within the boxes, diamonds represents mean scores, bars represent median scores, and shaded regions represent the 25th through 75th percentiles. The whiskers represent the lowest and highest quartiles.

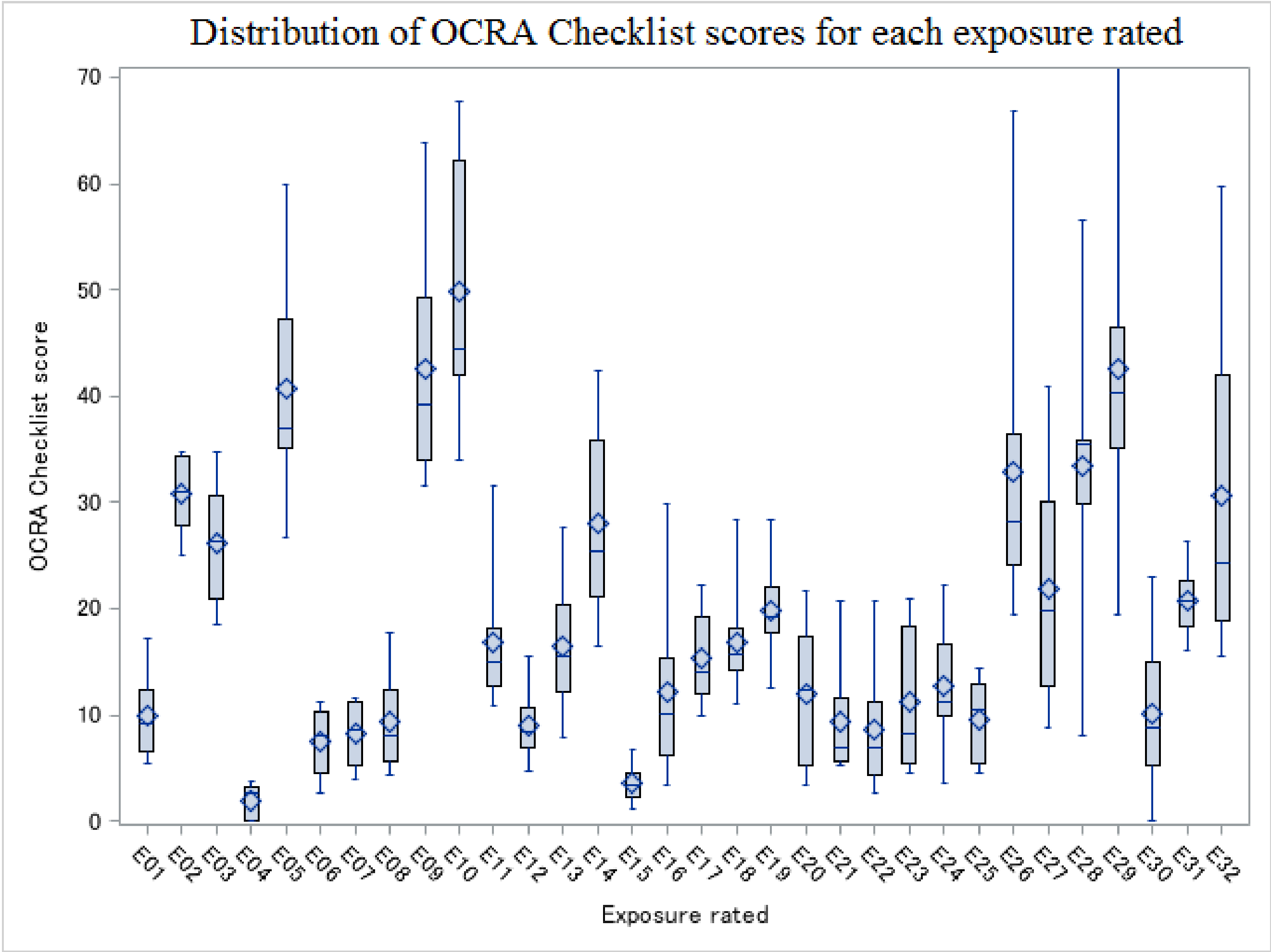


Figure 4.1(b). Distribution of raw OCRA Checklist scores by exposure rated. Within the boxes, diamonds represents mean scores, bars represent median scores, and shaded regions represent the 25th through 75th percentiles. The whiskers represent the lowest and highest quartiles.

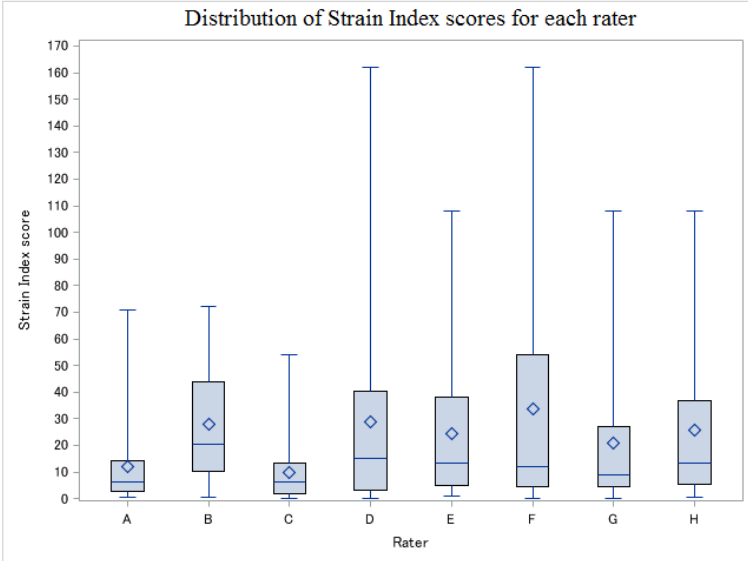


Figure 4.2(a). Distribution of raw SI scores by rater. Within the boxes, diamonds represents mean scores, bars represent median scores, and shaded regions represent the 25th through 75th percentiles. The whiskers represent the lowest and highest quartiles.

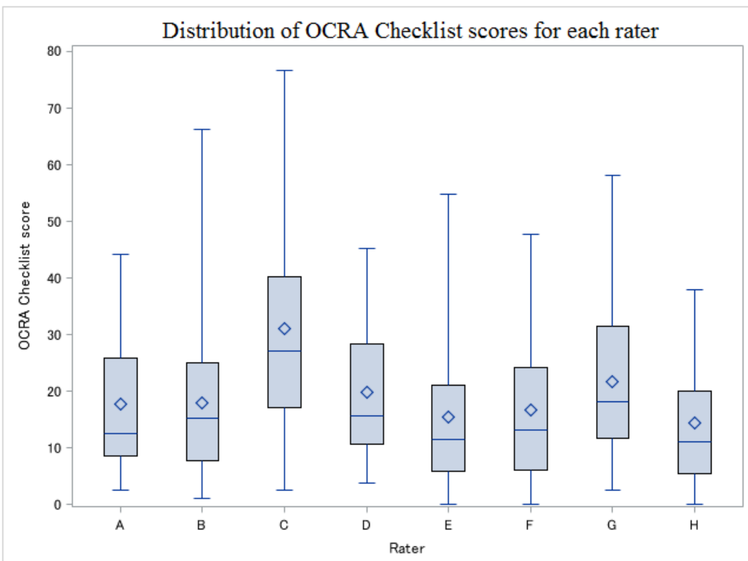


Figure 4.2(b). Distribution of raw OCRA Checklist scores by rater. Within the boxes, diamonds represents mean scores, bars represent median scores, and shaded regions represent the 25th through 75th percentiles. The whiskers represent the lowest and highest quartiles.

Differences between raters were tested using a two-way ANOVA with a Tukey-Kramer adjustment for multiple comparisons. The test revealed that mean SI and OCRA Checklist scores differed significantly ($p<0.05$) between some raters. The mean scores of rater C were the least similar to all other raters' mean SI and OCRA Checklist scores. Table 4.3 lists these rater differences by method and across all exposures.

Table 4.3. Significant differences between mean transformed summary exposure scores for all raters across 32 cheese production job-exposure assessments.

Rater	Differs from raters
Strain Index scores	
A	B D E F H
B	A C G
C	B D E F G H
OCRA Checklist scores	
C	A B D E F G H
D	C E H
G	E F H

Tables A1 and A2 in Appendix III show the raw SI and OCRA Checklist hazard scores that the eight raters each assigned to the 32 job exposures. Median scores for each exposure are listed in the far right column, and median scores for each rater are listed in the bottom row. Standard deviation estimates are also provided.

4.2. *Inter-method reliability*

4.2.1. *Group-level analyses*

Group analyses were performed using raw SI and OCRA Checklist exposure classifications stratified by rater and by job complexity. Stratifications by all other group-level traits (*i.e.* job symmetry, rater education, rater experience) did not reveal any statistically significant associations or trends.

Inter-method agreement was significantly different (p=0.0256) for dichotomous classification comparisons, and approaching significance (p=0.0740) for three-level classification comparisons. Tables 4.4 and 4.5 show that kappa, Spearman and overall agreement statistics were lowest for multi-task exposure classification comparisons. Tests of symmetry were not significant for any job complexity grouping, but observed p-values were lowest for multi-task jobs.

Table 4.4. Inter-method reliability statistics by job complexity for dichotomous exposure classification criteria.

Job complexity	Overall agreement, p_o	*Test of symmetry	Spearman, r_s	Kappa, κ	95% CI for κ
Mono-task	69.6%	p=0.6291	0.37	0.37	0.12-0.62
Dual-task	82.1%	p=0.8238	0.55	0.55	0.38-0.73
Multi-task	68.2%	p=0.3449	0.24	0.24	0.03-0.45
Overall	74.6%	p=1.0000	0.44	**0.42	0.30-0.53

* McNemar's test.
**Chi-square test of differences among kappa coefficients was approaching significance (p=0.0740).

Table 4.5. Inter-method reliability statistics by job complexity for three-level exposure classification criteria.

Job complexity	Overall agreement, p_o	*Test of symmetry	Spearman, r_s	Kappa, κ_w	95% CI for κ_w
Mono-task	60.7%	p=0.9309	0.63	0.40	0.22-0.59
Dual-task	68.8%	p=0.1328	0.64	0.42	0.28-0.55
Multi-task	55.7%	p=0.0557	0.29	0.16	0.00-0.32
Overall	64.1%	p=0.0434	0.54	**0.53	0.43-0.62

* Bowker's test.
**Chi-square test of differences among weighted kappa coefficients was significant (p=0.0256).

4.2.2. *Rater-level analyses*

The agreement in each rater's exposure classifications using the SI and OCRA Checklist was analyzed. In Appendix III, Tables A3(a) through A3(h) show the within-rater agreement of dichotomous exposure classification using the SI and OCRA Checklist. Raters A, C, D, and G assessed more jobs as hazardous when using the OCRA Checklist, whereas raters B, E, F, and H assessed more jobs as hazardous when using the SI.

Tables A4(a) through (h) in Appendix III show the within-rater agreement of three-level exposure classification for the SI and OCRA Checklist. Raters A, C, and D assessed more jobs as hazardous or at the action-level when using the OCRA Checklist, whereas raters B, E, F, G, and H assessed more jobs as hazardous or at the action-level when using the SI.

Tables 4.6 and 4.7 summarize the between-method reliability in exposure classification. Generally, inter-method reliability was low for individual raters; the lower confidence interval level (LCL) for kappa was below 0.40 for all raters except F and G. For both hazard classification criteria comparisons, Raters F and G exhibited the highest inter-method agreement ($p_o \geq$ 65.6%; $\kappa \geq 0.68$, 95% LCL$\geq$0.42) while also exhibiting the strongest inter-method hazard classification associations ($r_s \geq 0.63$). For the three-level criteria comparisons, rater F did appear to be slightly biased (p=0.0534 for symmetry), rating more jobs at the action-level or as hazardous with the SI. The lowest measures of inter-method reliability were associated with rater C ($p_o \geq$ 62.5%; $\kappa \geq 0.25$, 95% LCL=0.00; $r_s \geq 0.32$).

Tables 4.6. Inter-method reliability statistics by rater for dichotomous exposure classification criteria.

Rater	Overall agreement, p_o	*Test of symmetry	Spearman, r_s	Kappa, κ	95% CI for κ
A	65.6%	p=0.2266	0.33	0.31	0.00-0.63
B	71.9%	p=0.0039	0.42	0.29	0.00-0.57
C	62.5%	p=0.0063	0.32	0.25	0.00-0.51
D	75.0%	p=0.7266	0.43	0.42	0.09-0.75
E	78.1%	p=0.1250	0.58	0.55	0.27-0.83
F	84.4%	p=1.0000	0.68	0.68	0.42-0.94
G	87.5%	p=0.1250	0.75	0.71	0.46-0.97
H	71.9%	p=0.0391	0.49	0.43	0.16-0.72
Overall	74.6%	p=1.0000	0.44	**0.47	0.37-0.57

* McNemar's test.
**Chi-square test of differences among kappa coefficients was not significant (p=0.1034).

Table 4.7. Inter-method reliability statistics by rater for and three-level exposure classification criteria.

Rater	Overall agreement, p_o	*Test of symmetry	Spearman, r_s	Kappa, κ_w	95% CI for κ_w
A	65.6%	p=0.3916	0.64	0.63	0.39-0.87
B	81.3%	p=0.0293	0.54	0.50	0.17-0.83
C	65.6%	p=0.0503	0.41	0.33	0.00-0.63
D	65.6%	p=0.5433	0.57	0.61	0.36-0.85
E	46.9%	p=0.0117	0.58	0.43	0.19-0.68
F	65.6%	p=0.0534	0.78	0.69	0.50-0.87
G	68.8%	p=0.0897	0.63	0.69	0.48-0.90
H	53.1%	p=0.0367	0.63	0.49	0.26-0.71
Overall	64.1%	p=0.0434	0.54	**0.57	0.49-0.66

*Bowker's test.
**Chi-square test of differences among weighted kappa coefficients was not significant (p=0.4055).

4.3. *Inter-rater Reliability*

For the SI, the proportion of agreement in exposure classifications among all raters was 82.8% and 76.2% for the dichotomous and three-level criteria, respectively. For the OCRA Checklist, proportion of agreement was 82.4% and 71.9% for dichotomous and three-level criteria.

Intraclass correlation coefficient (ICC) measures for task variable scores, exposure scores, and exposure classifications are shown in Table 4.8. The ICCs for the SI and OCRA Checklist scores were 0.54 and 0.70, respectively. Of the task variables, the lowest reliability was associated with the SI *hand/wrist posture* (ρ=0.17) and the OCRA Checklist *additional factors* (ρ=0.19). The highest task variable reliability was associated with SI *efforts per minute* (ρ=0.59) and the OCRA Checklist *frequency of technical action* (ρ=0.70).

Table 4.8. Intraclass correlation coefficient calculations for SI and OCRA Checklist exposure scores, task-variable scores, and exposure classifications.

Task-variable and exposure scores	*ICC, ρ	95% CI for ρ
Strain Index		
SI score	0.54	0.47-0.62
Intensity of exertion	0.35	0.28-0.43
Duration of exertion	0.39	0.32-0.47
Efforts per minute	0.59	0.52-0.66
Hand/wrist posture	0.17	0.11-0.23
Speed of work	0.31	0.24-0.39
OCRA Checklist		
OCRA Checklist score	0.70	0.63-0.75
Frequency of technical actions	0.70	0.63-0.76
Force exertion	0.31	0.24-0.39
Awkward posture/movement	0.48	0.40-0.56
Additional factors	0.19	0.13-0.26
Exposure classifications		
Dichotomous SI	0.45	0.37-0.53
Three-level SI	0.50	0.42-0.58
Dichotomous OCRA Checklist	0.47	0.39-0.55
Three-level OCRA Checklist	0.56	0.48-0.63

*ICC, Intraclass correlation coefficient

5. Discussion

The present study evaluated the inter-method and inter-rater reliability of the SI and the OCRA Checklist. Inter-method reliability was moderate to good for most raters , indicating that the SI and OCRA Checklist classified the same job exposures to physical risk factors similarly. Inter-rater reliability of the SI and OCRA Checklist summary scores and exposure classifications suggested moderate intra-method reliability.

5.1. *Inter-method reliability*

5.1.1. *Group-level analyses*

Strain Index and OCRA Checklist assessments of multi-task jobs appear to classify exposures less consistently than do assessments of mono- or dual-task jobs. Agreement and strength of association measures were strongest for dual-task jobs ($p_o \geq 68.8\%$, $r_s \geq 0.55$, $\kappa \geq 0.42$) regardless of the exposure classification criteria used to compare methods. Inter-method reliability statistics for mono-task jobs were very similar to those of dual-task jobs when exposures were classified dichotomously. For three-level classifications, mono-task jobs were appreciably more reliable than multi-task jobs, and the chi-square test of kappa coefficient equality indicated that multi-task agreement was significantly lower than agreement among scores for less complex jobs (p=0.0256). This suggests that SI and OCRA Checklist inter-method reliability is strongly affected by the complexity of jobs evaluated.

Other group-level analyses did not reveal any significant associations between inter-method reliability and the rater experience, rater profession/education, or job symmetry.

5.1.2. *Rater-level analyses*

Individual-level analyses of inter-method reliability revealed that the strength of association and agreement between SI and OCRA Checklist scores varied widely according to rater. For instance, the kappa coefficient lower confidence intervals (κ LCLs) for rater C included 0.0, indicating that any agreement in assessment was no better than what would be expected by chance alone. In contrast, raters F and G both demonstrated at least moderate agreement beyond chance (κ LCLs≥0.42), and they most reliably classified the same job exposures as safe, at the action-level, or hazardous. Overall, raters A, B, D, E, and H demonstrated poor to moderate levels of agreement between methods (κ ranged from 0.25 to 0.63, κ LCLs ranged from 0.00 to 0.39) and moderate levels of association between methods (r_s ranged from 0.32 to 0.64).

Despite their lack of field experience, raters F and G were both more familiar with the principles, theory, and literature associated with the OCRA Checklist than were the other novice Checklist users. And rater F, a novice user of the SI, was more familiar with the SI than were other novice users. Thus, the poor, fair, or moderate agreement among SI and OCRA Checklist exposure classifications for the other six raters may improve if their familiarity with the methods improves. And increasing familiarity may require additional study of the principles and theory underlying the observational methods, not just additional practice performing the exposure assessments.

For dichotomous and three-level exposure classification comparisons, all but rater G consistently ranked more job exposures as action-level or hazardous with one method rather than the other. Rater G assessed more job exposures as hazardous using the OCRA Checklist dichotomous criteria and more as safe using the OCRA Checklist three-level criteria. For three-level classification criteria, five raters assigned higher exposure classifications when using the SI. Overall, the results suggest that the SI may rank more jobs as hazardous than does the OCRA Checklist.

5.2. *Inter-rater reliability*

The proportion of agreement in SI and OCRA Checklist exposure classifications among raters was moderately high (≥71.9%). Also, ICC analyses showed that inter-rater reliability of SI and OCRA Checklist summary scores was moderate to good. The majority of task-variable scores were characterized by low ICCs and, thus, exhibited high between-rater variance. The *efforts per minute*, *frequency of technical actions*, and *awkward postures*/*movements* variables were at least moderately reliable (ρ LCL≥0.40). Yet, the remaining variables all received reliability coefficient LCLs that included poor reliability ranges ($0.00<\rho< 0.40$). The inter-rater reliability of dichotomous exposure classifications also included ICC LCLs associated with poor reliability. Compared to dichotomous criteria, three-level SI and OCRA Checklist exposure classifications were more reliable (ρ LCLs≥0.42).

Overall, higher inter-rater reliability, was associated with the OCRA Checklist task variables scores, exposure scores, and exposure classifications than compared to the SI. Only the OCRA Checklist summary score ICC was significantly higher than the analogous SI scores, which the lack of confidence interval overlap indicates.

5.3. *Comparisons to other studies*

5.3.1. *Inter-method reliability*

Prior studies of inter-method reliability between the SI, OCRA Checklist, or OCRA Index have only reported overall agreement or correlation statistics. This is the first study of the two methods to report chance-corrected agreement using kappa and weighted kappa coefficients. None of the studies have reported tests of symmetry, but they all report that the SI ranks more job exposures as hazardous than do the OCRA Checklist or Index (Apostoli *et al.* 2004, Jones and Kumar 2010, and Chiasson *et al.* 2012).

The observed SI and OCRA Checklist inter-method reliability is similar to or stronger than results reported by prior investigators. Apostoli *et al*. (2004) used the SI and OCRA Checklist to assess 12 repetitive job exposures using three-level exposure classification criteria. The authors reported a low proportion of overall agreement (p_o=41.7%), with all of the disagreement due to SI scores rating more jobs as hazardous or at the action-level than the OCRA Checklist. Of the eight rater participants in the present study, seven exhibited an overall proportion of agreement in ratings greater than or equal to 65.6%. Higher agreement percentages may have been achieved simply due to the greater number of exposures analyzed (32 compared 12 for Apostoli *et al.*).

Using three-level exposure classification criteria, Jones and Kumar (2010) reported 83% overall agreement between the SI and the OCRA Index. This is higher than the observed proportion in the present study, but Jones and Kumar evaluated 87 individuals performing just four different high hazard sawmill jobs. Thus, the agreement in SI and OCRA Index scores may not hold for jobs characterized by more varied levels of exposure.

Chiasson *et al.* (2012) reported 60.1% overall proportion of agreement between OCRA Index and SI three-level exposure classifications. The authors assessed 167 different job exposures in a variety of industries, many exposures were likely associated with multi-task jobs. The authors do not describe the complexity of the jobs rated in terms of the constituent number of tasks, but average cycle times for all of the jobs ranged from 1.1 to 450 minutes. The probability that these jobs were complex might explain Chiasson's *et al.* observation of weak correlation between the SI and OCRA Index exposure classifications (Pearson r=0.32). Correlation analyses for three-level exposure classifications in the present study were higher on average (overall r_s=0.54), and ranged higher (maximum r_s=0.78).

5.3.2. *Inter-rater reliability*

This is the first study of the inter-rater reliability of the OCRA Checklist. The estimate of OCRA Checklist rater reliability was good, and is higher than reports of SI score reliability (Stevens *et al* 2004, Spielholz *et al.* 2008).

Stevens's *et al.* (2004) study of the SI showed good to excellent levels of inter-rater reliability of task-variable scores for fifteen raters individually evaluating 73 mono-task job exposures. Task-variable reliability coefficients ranged from 0.77 to 0.81 (ρ LCLs≥0.63) for all but the *hand/wrist posture* variable, which was still associated with moderate reliability (ρ=0.66, LCL=0.45). For the overall inter-rater reliability of the SI score, Stevens' *et al.* reported poor to moderate reliability (ρ=0.43, LCL=0.25). The inter-rater reliability of task-variable scores observed by the present study is less impressive than Stevens's *et al.* results. However, the assessment of dual-task and multi-task jobs likely increased task-variable score variance between raters, thereby reducing inter-rater reliability.

Stevens's *et al.* (2004) selected jobs for the reliability study that exhibited a full and balanced range of physical work exposures. In contrast, the jobs rated in the present study were a random sample of possible exposures in the food manufacturing industry. No effort was made to ensure that jobs contained a balanced number of physical exposures. Visual inspection of the task-variable scores selected by all the raters confirmed that many task-variables were underutilized. For instance, about 75% of the *intensity of exertion* variable scores were either 3.0 or 6.0. Less than 2% of all of the exposures rated received very-hard or near-maximal SI *intensity of exertion* scores (9.0 and 13.0). In contrast, the *efforts per minute* task variable scores were more equally distributed throughout the range of possible scores, but 40% of exposures were still rated at the maximum-level of 20 efforts per minute. Generally, inter-rater reliability estimates are lower when the range of possible scores is small (Stevens *et al.* 2004, Streiner & Norman 2008).

The SI inter-rater reliability results of Spielholz's *et al*. (2008) are similar to those observed in the present study. The agreement of assessments made by four raters evaluating 125 cyclic task-exposures was measured using an equally weighted kappa coefficient, which ranged from 0.22 to 0.44 for task-variable and SI scores. Similar to the present study, the exposures rated were from a random sample of jobs in the manufacturing and healthcare sectors. Of interest is that Spielholz *et al.* reported a significant difference in the reliability of ratings provided by a pair of expert raters compared to a rater-pair comprised of an expert and a novice. The present study was not designed to evaluate the effects of rater experience on score reliability, but stratifying SI raters by experience-level did not produce significantly improved reliability results. Further, only one experienced OCRA Checklist rater participated in the present study, and so no such comparison could be made for within the method.

5.4 *Practical implications of findings*

Observational methods are a popular type of ergonomic exposure assessment due to their low cost, systematic design, and moderate or good validity and reliability (Spielholz *et al.* 2001, David 2005, Takala *et al.* 2010). Further, recent studies have demonstrated that observational methods are strong predictors of work-related upper extremity disorder development (Violante *et al.* 2007, Spielholz *et al.* 2008, Bonfiglioli *et al.* 2012, Garg *et al.* 2012), and they may even be more strongly predictive than direct measures of individual physical risk factors (Gerr *et al.* 2010).

The present study suggests that SI and OCRA Checklist exposure assessments produce moderately similar results when applied to repetitive food manufacturing jobs. Yet, the inter-method reliability is probably too low to assume that the two ergonomic tools produced interchangeable exposure assessments of the same jobs. This is not surprising because the two methods summarize physical exposure scores differently; the SI considers the *intensity of exertion* to be the central predictor of risk, whereas the OCRA Checklist considers the *frequency of technical action* to be the most important predictor. Nevertheless, based on the results of previous inter-method reliability studies of the SI and OCRA methods, the SI and OCRA Checklist may more similar to each other than to any other observational UE assessment method (Drinkaus *et al.* 2003, Apostoli *et al.* 2004, Spielholz *et al.* 2008, Jones and Kumar 2010, Chiasson *et al.* 2012). An additional factor that may influence inter-method reliability is job exposure to physical shoulder disorder risks. The OCRA Checklist raters must assess physical exposures affecting the shoulder, whereas the SI raters only assess exposures to DUE MSD risks. This difference in the anatomical focus of each method may have negatively influenced the agreement and strength of association between summary scores. However, these cheese production jobs were chosen for assessment in the present study because they did not appear to expose workers predominantly to shoulder disorder risks. Had only the DUE been assessed with both methods, one might reasonably expect some increase in inter-method reliability.

The present study is a reminder that the inter-rater reliability of the SI and the OCRA Checklist is simultaneously sensitive and robust to the effects of rater bias. For example, neither method may be a satisfactory measure of any individual physical exposure (*e.g.* force intensity, high repetition, awkward postures, *etc.*) as these individual elements had generally poor reliability. However, the reliability of summary scores ranged from moderate to good, suggesting that the use of a summary score may mask the inherent rater-related variability. Alternatively, variability due to rater effects will decrease if average scores are used (Streiner and Norman 2008) or if single-measure scores are based on consensus ratings from multiple analysts (Stevens *et al.* 2004). Additionally, the results of the inter-method reliability analyses indicated that multi-task jobs may be associated with greater variance in individual summary scores. It is reasonable to assume that the inter-rater reliability of summary scores would also be reduced when multi-task jobs are assessed. Researchers and practitioners should consider these limitations when deciding which measures of physical exposure are most appropriate for a given work environment.

5.5. *Strengths, limitations and future work*

The strengths of the present study include: group and individual-level comparisons of SI and OCRA Checklist exposure assessment classifications, stratification of inter-method reliability results according to possible covariates, and the participation of multiple raters with varied background and experience applying exposure assessments. Also all raters assessed all physical parameters of the 21 cheese production jobs, and a multitude of reliability statistics were used to measure the agreement and strength of association between methods and raters.

Generalizability of results may be limited by the fact that only a small number of jobs were rated (21 jobs, 32 independent upper extremity exposures), and all came from a single manufacturing facility. The results may not be applicable to job exposures in non-manufacturing industries. Further, the widely varying SI and OCRA Checklist scores for each rater suggest that training did not eliminate systematic bias. The magnitude or direction of this bias is uncertain. Authors of similar studies estimated the effects of systemic bias by using the scores of an expert to serve as the "true" or "benchmark" score (Waters *et al.* 1993, Ketola *et al.* 2001). However, at least in the case of the SI, the reliability of expert ratings is not perfect and systemic bias is to be expected (Spielholz *et al.* 2008). This is because the results of observational methods are strongly influenced by the subjective estimation of physical exposure parameters (Kilbom 1994, Takala *et al.* 2010). The findings of the present study suggest that increased rater familiarity with the SI and OCRA Checklist may improve the inter-method reliability, but the effect of method familiarity and training on inter-rater reliability is uncertain.

Future studies of the SI and OCRA Checklist should be designed to evaluate the effect that multi-task jobs have on the rater and method reliability. While some authors have used multi-task scoring systems for the SI (Bao *et al.* 2009b, Gerr *et al.* 2010, Garg *et al.* 2012) or OCRA Checklist (Colombini and Occhipinti 2012), the reliability and validity of these systems requires further testing. Additionally, studies of the inter-method reliability of the SI and OCRA Checklist should quantify what effect, if any, rater training and familiarity with the methods have on reliability outcomes.

6. Conclusion

Strain Index and OCRA Checklist assessments classified job exposures similarly, and the strength of inter-method reliability may depend on rater familiarity with the methods and the complexity of jobs assessed. The summary scores of both methods were characterized by moderate or good inter-rater reliability, but the reliability of SI and OCRA Checklist task-variable scores was poor. If the inter-rater reliability of exposure assessment scores is a concern, researchers and occupational health and safety practitioners should choose to use average or consensus-based scores. Future SI and OCRA Checklist research should investigate whether inter-method and inter-rater reliability increases with enhanced rater training or the use of multi-task scoring systems.

.

7. 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]. *Giornale italiano di medicina del lavoro ed ergonomia,* 26 (3), 223-241.

Armstrong, B.K., White, E. & Saracci, R., 1992. *Principles of exposure measurement in epidemiology* Oxford: Oxford University Press.

Bao, S., 2004. Hand-intensive jobs: Effectively evaluating risk levels. *Professional Safety,* 49 (1), 28-34.

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 (4), 381-392.

Bao, S., Howard, N., Spielholz, P., Silverstein, B. & Polissar, N., 2009a. Interrater reliability of posture observations. *Human Factors,* 51 (3), 292-309.

Bao, S., Spielholz, P., Howard, N. & Silverstein, B., 2009b. Application of the strain index in multiple task jobs. *Applied Ergonomics,* 40 (1), 56-68.

Berk, R.A., 1979. Generalizability of behavioral observations - clarification of interobserver agreement and interobserver reliability. *American Journal of Mental Deficiency,* 83 (5), 460-472.

Bernard, B.P., 1997. Musculoskeletal disorders and workplace factors: A critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back. U.S. Department of Health and Human Services.

Bls, 2012. *Nonfatal occupational injuries and illnesses requiring days away from work* [online]. http://www.bls.gov/news.release/osh2.toc.htm[Accessed 2013/02/01].

Bonfiglioli, R., Mattioli, S., Armstrong, T., Graziosi, F., Marinelli, F., Farioli, A. & Violante, F., 2012. Validation of the acgih tlv for hand activity level in the octopus cohort: A two-year longitudinal study of carpal tunnel syndrome. *Scandanavian Journal of Work Environment & Health,* Available from: http://www.sjweh.fi/show_abstract.php?abstract_id=3312.

Borg, G.A., 1982. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc,* 14 (5), 377-81.

Burdorf, A., 1995. Reducing random measurement error in assessing postural load on the back in epidemiologic surveys. *Scandinavian Journal of Work Environment & Health,* 21 (1), 15-23.

Burdorf, A. & Van Der Beek, A., 1999. Exposure assessment strategies for work-related risk factors for musculoskeletal disorders. *Scand J Work Environ Health,* 25 Suppl 4, 25-30.

Burt, S. & Punnett, L., 1999. Evaluation of interrater reliability for posture observations in a field study. *Applied Ergonomics,* 30 (2), 121-135 Available from: <Go to ISI>://WOS:000078665800003.

Chiang, H.C., Ko, Y.C., Chen, S.S., Yu, H.S., Wu, T.N. & Chang, P.Y., 1993. Prevalence of shoulder and upper-limb disorders among workers in the fish-processing industry. *Scandinavian Journal of Work Environment & Health,* 19 (2), 126-131.

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. *International Journal of Industrial Ergonomics,* 42 (5), 478-488.

Cohen, J., 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin,* 70 (4), 213-220.
Colombini, D., 1998. An observational method for classifying exposure to repetitive movements of the upper limbs. *Ergonomics,* 41 (9), 1261-1289.
Colombini, D., Occhipinti, E., Cairoli, S. & Baracco, A., 2000. [proposal and preliminary validation of a check-list for the assessment of occupational exposure to repetitive movements of the upper limbs]. *Medicina Del Lavoro,* 91 (5), 470-485.
Colombini, D., Occhipinti, E., Cerbai, M., Battevi, N. & Placci, M., 2011. [updating of application procedures and criteria for ocra checklist]. *Medicina Del Lavoro,* 102 (1), 1-39.
Colombini, D., Occhipinti, E., Delleman, N., Fallentin, N., Kilbom, A. & Grieco, A., 2001. Exposure assessment of upper limb repetitive movements: A consensus document developed by the technical committee on musculoskeletal disorders of international ergonomics association (iea) endorsed by international commission on occupational health (icoh). *G Ital Med Lav Ergon,* 23 (2), 129-42.
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* Oxford: Elsevier.
Da Costa, B.R. & Vieira, E.R., 2010. Risk factors for work-related musculoskeletal disorders: A systematic review of recent longitudinal studies. *American Journal of Industrial Medicine,* 53 (3), 285-323.
David, G.C., 2005. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational Medicine-Oxford,* 55 (3), 190-199.
De Luca, C.J., 1997. The use of surface electromyography in biomechanics. *Journal of Applied Biomechanics,* 13 (2), 135-163.
Dempsey, P.G., Mcgorry, R.W. & Maynard, W.S., 2005. A survey of tools and methods used by certified professional ergonomists. *Applied Ergonomics,* 36 (4), 489-503.
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. *Applied Ergonomics,* 43 (3), 632-636.
Drinkaus, P., Sesek, R., Bloswick, D., Bernard, T., Walton, B., Joseph, B., Reeve, G. & Counts, J.H., 2003. Comparison of ergonomic risk assessment outputs from rapid upper limb assessment and the strain index for tasks in automotive assembly plants. *Work,* 21 (2), 165-72.
Drinkaus, P., Sesek, R., Bloswick, D.S., Mann, C. & Bernard, T., 2005. Job level risk assessment using task level acgih hand activity level tlv scores: A pilot study. *Int J Occup Saf Ergon,* 11 (3), 263-81.
Dunn, G., 1989. *Design and analysis of reliability studies the statistical evaluation of measurement errors* New York: Oxford University Press.
Ebersole, M. & Armstrong, T., Year. Inter-rater reliability for hand activity level (hal) and force metricsed.^eds. *Human Factors and Ergonomics Society 46th Annual Meeting,* Baltimore, MD, 1037-1040.
Ebersole, M.L. & Armstrong, T.J., 2006. Analysis of an observational rating scale for repetition, posture, and force in selected manufacturing settings. *Human Factors,* 48 (3), 487-498.

Epm, 2012. *Ergonomics of posture and movement* [online]. http://www.epmresearch.org/index.php [Accessed 2013/02/01].
Essendrop, M., Schibye, B. & Hansen, K., 2001. Reliability of isometric muscle strength tests for the trunk, hands and shoulders. *International Journal of Industrial Ergonomics,* 28 (6), 379-387.
Ferreira, J., Gray, M., Hunter, L. & Riley, D., 2009. Development of an assessment tool for repetitive tasks of the upper limbs (art). Derbyshire: Health and Safety Executive.
Fleiss, J.L., 1986. *The design and analysis of clinical experiments* New York: Wiley.
Fleiss, J.L. & Cohen, J., 1973. Equivalence of weighted kappa and intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement,* 33 (3), 613-619.
Garg, A., Year. Composite strain indexed.^eds. *16th International Ergonomics Association world conference proceedings*, Maastricht, the Netherlands.
Garg, A. & Kapellusch, J., 2011. Job analysis techniques for distal upper extremity disorders. *Reviews of Human Factors and Ergonomics,* 7 (1), 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 (4), 396-414.
Garg, A., Moore, S. & Kapellusch, J.M., Year. The strain index to analyze jobs for risk of distal upper extremity disorders: Model validation. *In*: Helander, M., Xie, M., Jaio, M. & Tan, K.C., ed.^eds. *2007 IEEE International Conference on Industrial Engineering and Engineering Management, Vols 1-4*, 497-499.
Gerr, F., 2010. Musculoskeletal disorders among manufacturing workers: A prospective epidemiological study. *American College of Occupational and Environmental Medicine, Annual Conference.* Orlando, FL.
Grieco, A., 1998. Application of the concise exposure index (ocra) to tasks involving repetitive movements of the upper limbs in a variety of manufacturing industries: Preliminary validations. *Ergonomics,* 41 (9), 1347-1356.
Hägg, G.M., Melin, B. & Kadefors, R., 2004. Applications in ergonomics. *In* Merletti, R. & Parker, P. eds. *Electromyography - physiology, engineering, and noninvasive applications.* John Wiley & Sons, 343-363.
Huisstede, B.M.A., Bierma-Zeinstra, S.M.A., Koes, B.W. & Verhaar, J.a.N., 2006. Incidence and prevalence of upper-extremity musculoskeletal disorders. A systematic appraisal of the literature. *Bmc Musculoskeletal Disorders,* 7.
Jones, T. & Kumar, S., 2007. Comparison of ergonomic risk assessments in a repetitive high-risk sawmill occupation: Saw-filer. *International Journal of Industrial Ergonomics,* 37 (9-10), 744-753.
Jones, T. & Kumar, S., 2010. Comparison of ergonomic risk assessment output in four sawmill jobs. *International Journal of Occupational Safety and Ergonomics,* 16 (1), 105-111.
Juul-Kristensen, B., Hansson, G.A., Fallentin, N., Andersen, J.H. & Ekdahl, C., 2001. Assessment of work postures and movements using a video-based observation method and direct technical measurements. *Applied Ergonomics,* 32 (5), 517-524.
Ketola, R., Toivonen, R. & Viikari-Juntura, E., 2001. Interobserver repeatability and validity of an observation method to assess physical loads imposed on the upper extremities. *Ergonomics,* 44 (2), 119-131.

Kilbom, A., 1994. Assessment of physical exposure in relation to work-related musculoskeletal disorders - what information can be obtained from systematic observations. *Scandinavian Journal of Work Environment & Health*, 20, 30-45.
Knox, K. & Moore, J.S., 2001. Predictive validity of the strain index in turkey processing. *Journal of Occupational and Environmental Medicine,* 43 (5), 451-462.
Kumar, S., 1993. Perception of posture of short-duration in the spatial and temporal domains. *Applied Ergonomics,* 24 (5), 345-350.
Landis, J.R. & Koch, G.G., 1977. Measurement of observer agreement for categorical data. *Biometrics,* 33 (1), 159-174.
Li, G. & Buckle, P., 1999. Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics,* 42 (5), 674-695.
Lowe, B.D., 2004. Accuracy and validity of observational estimates of wrist and forearm posture. *Ergonomics,* 47 (5), 527-554.
Maclure, M. & Willett, W.C., 1987. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol,* 126 (2), 161-9.
Moore, J.S. & Garg, A., 1995. The strain index - a proposed method to analyze jobs for risk of distal upper extremity disorders. *American Industrial Hygiene Association Journal,* 56 (5), 443-458.
Moore, J.S., Rucker, N.P. & Knox, K., 2001. Validity of generic risk factors and the strain index for predicting nontraumatic distal upper extremity morbidity. *Aihaj,* 62 (2), 229-235.
Moore, J.S. & Vos, G.A., 2005. The strain index. *In* Stanton, N. ed. *Handbook of human factors and ergonomics methods.* Boca Raton: CRC Press.
Nrc/Iom, 2001. *Musculoskeletal disorers and the workplace: Low back and upper extremities* Washington, D.C.: National Academy Press,.
Occhipinti, E., 1998. Ocra: A concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics,* 41 (9), 1290-1311.
Occhipinti, E. & Colombini, D., 2001. The ocra method: Assessment of exposure to occupational repetitive actions of the upper limbs. *In* Karwowski, W. ed. *International encyclopedia of ergonomics and human factors.* London: Taylor & Francis.
Occhipinti, E. & Colombini, D., 2005. The occupational repetitive actions (ocra) methods: Ocra index and ocra checklist. *In* Stanton, N. ed. *Handbook of human factors and ergonomics methods.* Boca Raton: CRC Press.
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 (11), 1727-1739.
Occhipinti, E. & Colombini, D., 2012. Iea/who toolkit for wmsds prevention: Criteria and practical tools for a step by step approach. *Work-a Journal of Prevention Assessment & Rehabilitation,* 41, 3937-3944.
Punnett, L. & Wegman, D.H., 2004. Work-related musculoskeletal disorders: The epidemiologic evidence and the debate. *Journal of Electromyography and Kinesiology,* 14 (1), 13-23.
Rucker, N.P. & Moore, J.S., 2002. Predictive validity of the strain index in manufacturing facilities. *Applied Occupational and Environmental Hygiene,* 17 (1), 63-73.
Shrout, P.E. & Fleiss, J.L., 1979. Intraclass correlations - uses in assessing rater reliability. *Psychological Bulletin,* 86 (2), 420-428.

Silverstein, B.A., Fine, L.J. & Armstrong, T.J., 1986. Hand wrist cumulative trauma disorders in industry. *British Jouranl of Industrial Medicine,* 43, 779-784.

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. *Journal of Occupational & Environmental Hygiene,* 5 (4), 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 (6), 588-613.

Stephens, J.P., Vos, G.A., Stevens, E.M. & Moore, J.S., 2006. Test-retest repeatability of the strain index. *Applied Ergonomics,* 37 (3), 275-281.

Stevens, E.M., Vos, G.A., Stephens, J.P. & Moore, J.S., 2004. Inter-rater reliability of the strain index. *Journal of Occupational & Environmental Hygiene,* 1 (11), 745-751.

Streiner, D.L. & Norman, G.R., 2008. *Health measurement scales a practical guide to their development and use* Oxford: Oxford University Press.

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. *Scandinavian Journal of Work Environment & Health,* 36 (1), 3-24.

Tanaka, S., Petersen, M. & Cameron, L., 2001. Prevalence and risk factors of tendinitis and related disorders of the distal upper extremity among us workers: Comparison to carpal tunnel syndrome. *American Journal of Industrial Medicine,* 39 (3), 328-335.

Violante, F.S., Armstrong, T.J., Fiorentini, C., Graziosi, F., Risi, A., Venturi, S., Curti, S., Zanardi, F., Cooke, R.M., Bonfiglioli, R. & Mattioli, S., 2007. Carpal tunnel syndrome and manual work: A longitudinal study. *J Occup Environ Med,* 49 (11), 1189-96.

Waters, T.R., Putzanderson, V., Garg, A. & Fine, L.J., 1993. Revised niosh equation for the design and evaluation of manual lifting tasks. *Ergonomics,* 36 (7), 749-776.

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. *Human Factors,* 53 (4), 371-382.

Appendix I.

Strain Index

Analyst Name:		Date:	
Start Time:			
Line/work task/job name:			
Description of job task and notes:			

For left and right limbs each, select appropriate exposure level from orange drop-down box. Auto. multiplier calculation.

Task Variable	Exposure Level	Observation and Calculation Guidelines	Variable Multiplier	Exposure Multiplier	Exposure Multiplier
Intensity of Exertion	Light	Barely noticeable or relaxed effort (BS: 0-2)	1	Left	Right
	Somewhat Hard	Noticeable or definite effort (BS: 3)	3		
	Hard	Obvious effort; Unchanged facial expression (BS: 4-5)	6		
(BS = Borg CR-10 Scale)	Very Hard	Substantial effort; Changes expression (BS: 6-7)	9	-	-
	Near Maximal	Uses shoulder or trunk for force (BS: 8-10)	13		
Duration of Exertion (% of Cycle)	< 10%		0.5	Left	Right
	10 - 29%	$\%Duration\ of\ Exertion = 100 * \frac{duration\ of\ all\ exertions\ [secs]}{total\ observaton\ time\ [secs]}$	1.0		
	30 - 49%		1.5		
	50 - 79%		2.0	-	-
	> 80%		3.0		
Efforts Per Minute	< 4		0.5	Left	Right
	4-8	$Efforts\ per\ minute = \frac{number\ of\ exertions}{total\ observation\ time\ [mins]}$	1.0		
	9-14		1.5		
	15-19		2.0	-	-
	> 20		3.0		
Hand/Wrist Posture	Very Good	Perfectly Neutral = Ext: 0-10°; Flex: 0-10°; UlnDev: 0-10°	1.0	Left	Right
	Good	Near Neutral = Ext: 11-25°; Flex: 6-15°; UlnDev: 11-15°	1.0		
	Fair	Non-Neutral = Ext: 26-40°; Flex: 16-30°; UlnDev: 16-20°	1.5		
	Bad	Marked Deviation = Ext: 41-55°; Flex: 31-50°; UlnDev: 21-25°	2.0	-	-
	Very Bad	Near Extreme = Ext: >55°; Flex: >50°; UlnDev: >25°	3.0		
Speed of Work	Very Slow	Extremely relaxed pace	1.0	Left	Right
	Slow	Taking one's own time	1.0		
	Fair	Normal speed of motion	1.0		
	Fast	Rushed, but able to keep up	1.5	-	-
	Very Fast	Rushed and barely/unable to keep up	2.0		
Duration of Task Per Day (hours)	< 1		0.25	Left	Right
	1 - 2		0.50		
	2 - 4	*Note: Variable is predetermined. Do not adjust multiplier.*	0.75		
	4 - 8		1.00	-	-
	> 8		1.50		

	Left DUE	Right DUE
Product of all multipliers = Strain Index Scores:	-	-

End time:		**Total Time:**	

Appendix II.

OCRA Checklist

Start time:		Analyst Name:	

Company:	F.lli PinnaIndustria Casearia
Department:	Produzione Pecorino Romano
Line/work task/job name:	
Number of male workers:	Number of female workers:
Description of job task:	

1. DURATION OF WORK			Net Duration multiplier. Choose multiplier based on value of (1-E).	
Description of work shift components		**Minutes**	Minutes	Multiplier
(1-A) Shift Duration Note official/gross duration, too:	Actual/Net duration=			
(1-B) Breaks Note official duration, too:	Actual/Net duration=		60-120	0.5
			121-180	0.65
(1-C) Lunch break (if paid within shift) Note official duration, too:	Actual/Net duration=		181-240	0.75
			241-300	0.85
(1-D) Length of non-repetitive tasks (e.g. cleaning, supplying, etc.) Note official duration, too:	Actual/Net duration=		301-360	0.925
			361-420	0.95
(1-E) Total actual/net time worked in repetitive tasks (minutes) *Note: to calculate (1-E), subtract actual values of (1-B) thru (1-D) from the actual shift duration (1-A). Compare to multiplier ranges and record score.*	**0**		421-480	1
			>480	1.5
(1-F) Stated or calculated–Number of units produced per work shift (or number of work cycles per shift)	planned			
	actual			
(1-G) Net cycle time calculated [(1-F) actual*60/(1-E)] (seconds)			**Duration Multiplier**	
(1-H) Observed cycle time (seconds)				
(1-I) Difference between planned and actual work time		**#VALUE!**		
Analyst note: When (1-I) >5%, then reassessment of work organization is important.				

2. OPPORTUNITIES FOR RECOVERY (breaks must last at least 8 minutes)	Points due to presence	*Select most appropriate Recovery Score based on verbal anchors. Intermediate scores can be chosen.*
For any shift duration, there is an interuption of at least 8 minutes every hour during the repetitive work (also counting the lunch break).	0	
During a 7-8 hour shift, there are 2 breaks in the morning and 2 in the afternoon (plus the lunch break); **OR** there are four breaks plus the lunch break; **OR** four breaks during a 6 hour shift without a lunch break.	1	
During a 7-8 hour shift, there are 3 breaks plus a lunch break; **OR** there are two breaks during a 6-hour shift without a lunch break.	3	
During a 7-8 hour shift, there are 2 breaks plus the lunch break; **OR** 3 breaks without a lunch break; **OR** one break during a 6 hour shift without a lunch break.	4	**Recovery Score**
During an 8 hour shift, there is only an unpaid lunch break; **OR** in a 7 hour shift without a lunch break, there is only a single break lasting at least 10 mins.	6	
During a 7-8 hour shift, there are no actual breaks except for a few minutes insufficient for recovery (less than 5).	10	

page 1

Appendix II (continued).

3. FREQUENCY OF TECHNICAL ACTIONS (TAs)

(3-A) Dynamic Actions	Points due to presence
Arm movements are slow; frequent, short interruptions are possible. **(20 actions per minute).**	0
Arm movements are not too fast--constant and regular; short interuptions are possible. **(30 actions per minute).**	1
Arm movements are quite fast and regular; only occasional and irregular short pauses are possible. **(~40 actions per minute).**	3
Arm movements are fast and less reugular/rythmic; only occasional and irregular short pauses are possible. **(~40 actions per minute).**	4
Arm movements are fast; only occasional and irregular short pauses are possible. **(~50 actions per minute).**	6
Arm movements are very fast; the lack of interupptions makes it difficult to hold the pace. **(~60 actions per minute).**	8
Arm movements are extremely fast; No interuptions in work activity are possible. **(70 or more actions per minute).**	10

Intermediate scores can be chosen.

Counted # of TAs/min (if possible)

# LEFT	# RIGHT

(3-A) Scores

LEFT	RIGHT

(3-B) Static Actions (those TAs held statically for at least 5 seconds)	
The sum duration of static TAs comprises 50% or less of the work cycle.	0
The sum duration of static TAs comprises 51-80% of the work cycle.	2.5
The sum duration of static TAs comprises more than 80% of the work cycle.	4.5

(3-B) Scores

LEFT	RIGHT

Frequency of technical actions score: Select highest score for each limb from sections (3-A) & (3-B).

Frequency Score

LEFT	RIGHT
0	0

4. PRESENCE OF REPEATED FORCE EXERTIONS BY THE UEs

*Intermediate scores can be chosen. Choose one score for **EACH** force category.*

		Points due to presence or % of work cycle
Analyst note: All work activities only require **MINIMAL FORCE** (1 or 2 on Borg CR-10 scale). *Circle or highlight 'yes' or 'no' answer.*	No	0
	Yes	0
(4-A) Work activities require **MODERATE FORCE** exertions (3 or 4 on Borg CR-10 scale). When:	None	0
	~33%	2
	~50%	4
	~66%	6
	~100%	8
(4-B) Work activities require **INTENSE FORCE** exertions (5, 6 or 7 on Borg CR-10 scale). When:	None	0
	2 secs every 10 mins	4
	1%	8
	5%	16
	> 10%	24
(4-C) Work activities require **MAXIMAL FORCE** exertions (8, 9 or 10 on Borg CR-10 scale). When:	None	0
	2 secs every 10 mins	6
	1%	12
	5%	24
	> 10%	32

(4-A) Scores

LEFT	RIGHT

(4-B) Scores

LEFT	RIGHT

(4-C) Scores

LEFT	RIGHT

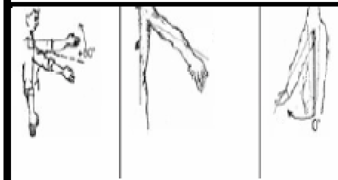
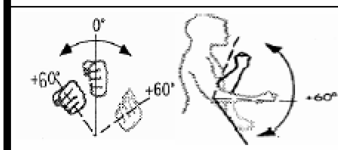
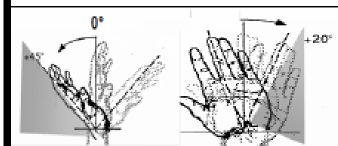
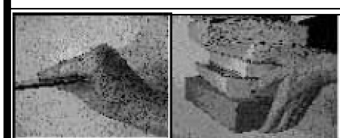
Force exertion score: Sum scores for (4-A) thru (4-C) for each limb.

Force Score

LEFT	RIGHT
0	0

page 2

Appendix II (continued).

5. PRESENCE OF AWKWARD POSTURE, AWKWARD MOVEMENT OR LACK OF VARIATION			Select highest score for each section [(5-A) thru (5-E)]. Intermediate scores can be chosen.
(5-A) Shoulder Posture	Points due to presence or % of work cycle		
The arms are not supported on a workbench and are a little uplifted for at least half of the time	None	0	
	Yes	1	
 •Arm more or less at shoulder height (≥80 degrees flexion); •Other awkward arm postures (≥80 degrees lateral abduction, ≥20 degrees extension).	None	0	(5-A) Scores
	~10%	2	LEFT RIGHT
	~33%	6	
	~50%	12	
	~100%	24	
(5-B) Elbow & Forearm Movement	Points due to % of work cycle		
 •Wide object rotation (prono-supination ≥ 60 degrees); •Wide flexion-extension (≥ 60 degrees) motion; •Jerking or striking motions.	None	0	(5-B) Scores
	25-50%	2	LEFT RIGHT
	51-80%	4	
	> 80%	8	
(5-C) Wrist Posture	Points due to % of work cycle		
 •Extreme wrist deviations (≥45 degrees flexion or extension; ≥ 20 degrees ulnar or radial deviation).	None	0	(5-C) Scores
	25-50%	2	LEFT RIGHT
	51-80%	4	
	> 80%	8	
(5-D) Grip (hand and finger posture)	Points due to % of work cycle		
 •Grips object with fingers (pinch, palmer or hook grip) instead of gripping with the whole hand & wrapping the thumb wrapped (power grip).	None	0	(5-D) Scores
	25-50%	2	LEFT RIGHT
	51-80%	4	
	> 80%	8	
Partial posture score: Select highest score for each limb from sections (5-A) thru (5-D).			Partial Posture Scores LEFT 0 RIGHT 0
(5-E) Lack of Variation (stereotypy/stereotypical actions)	Points due to presence		
•Performs same UE technical actions for 51-80% of the time; •OR the work cycle is 8 - 15 seconds in length and full of technical actions.	No	0	(5-E) Scores
	Yes	1.5	LEFT RIGHT
•Performs same UE technical actions 81% or more of the time; •OR the work cycle is less than 8 seconds in length and full of technical actions.	No	0	
	Yes	3	
For each limb, add (5-E) to the partial posture score. This yields the final awkward posture, awkward movement and lack of variation score.			Final Posture Scores LEFT 0 RIGHT 0

page 3

Appendix II (continued).

6. PRESENCE OF ADDITIONAL FACTORS		
(6-A) Physical Factors	Points due to presence	*For each (6-A) & (6-B), record the highest score associated with the additional factors identified.*
No physical additional factors are present	0	
Gloves worn for ≥ 50% of the time, and are indequate or uncomfortable	2	
Vibrating tools are used for ≥ 33% of the time	2	
Tools used compress the muscle tendons (leaving callouses, redness, etc.)	2	
Repeated impacts by the hand (using the hand as a tool) ≥ 10x per hour	2	
Use of hammers or mauls for hitting for ≥ 50% of the time	2	
Tasks require sudden movements, tears or kickback with frequency ≥ 2 per minute	2	
Precision tasks (over a 2 -3 mm area) that require close viewing distance for ≥ 50% of the time	2	(6-A) Scores: LEFT ___ RIGHT ___
Contact with cold objects/surfaces (≤ 0 °C), or working in cold environment (e.g. refrigerated storage) for ≥ 50% of the time	2	
More than one additional factor is present, and they are present all of the time	3	
(6-B) Organizational Factors Present for ≥ 50% of the Work Cycle	Points due to presence	
Work pace is not set by a machine	0	
Working pace is set by a machine, but there are "buffers" by which the pace can be slowed or accelerated.	1	(6-B) Scores: LEFT ___ RIGHT ___
Work pace is completely determined by the machine	2	

Additional factors score: Sum scores from sections (6-A) & (6-B) for each limb.Note: Max Additional Factor score is 5. →

Add'l Factor Score	
LEFT	RIGHT
0	0

Partial Total OCRA Simple Checklist Score: Sum section scores (the large grey boxes) for sections 3, 4, 5 and 6. Do this for each limb seperately. →

PARTIAL TOTAL SCORE	
LEFT	RIGHT

End time:		Net Rating Time (mins):	

Automatic calculation of total OCRA Checklist score based on Recovery and Duration factors. →

OCRA CHECKLIST SCORE	
LEFT	RIGHT
0	0

page 4

Appendix III.

Table A1. Individual and median SI scores by exposure ID and by rater.

	Rater Score									
Exposure ID	A	B	C	D	E	F	G	H	Median	SD
E01	5.7	6.8	3.4	4.5	5.1	4.5	4.5	6.8	4.8	1.19
E02	18.0	72.0	10.1	6.8	40.5	54.0	108.0	30.4	35.4	34.6
E03	13.5	72.0	13.5	40.5	54.0	54.0	81.0	40.5	47.3	24.5
E04	1.5	0.4	0.6	0.2	0.8	0.2	0.3	0.4	0.4	0.4
E05	15.2	40.5	27.0	121.5	40.5	81.0	60.8	81.0	50.6	34.9
E06	0.5	10.1	4.5	13.5	4.5	2.3	3.0	3.4	3.9	4.4
E07	1.0	10.1	6.8	3.0	9.0	4.5	3.0	4.5	4.5	3.1
E08	0.5	13.5	5.6	16.4	4.5	4.5	5.1	4.5	4.8	5.3
E09	13.5	40.5	18.0	40.5	36.0	108.0	27.0	54.0	38.3	29.7
E10	27.0	54.0	54.0	162.0	81.0	156.0	54.0	108.0	67.5	50.3
E11	10.1	20.3	0.8	18.0	2.3	5.1	10.1	2.3	7.6	7.4
E12	4.5	10.1	1.1	0.8	6.8	3.4	6.8	6.8	5.6	3.2
E13	6.0	10.1	0.5	4.5	18.0	9.0	9.0	13.5	9.0	5.4
E14	6.8	27.0	13.5	108.0	108.0	72.0	27.0	81.0	49.5	41.8
E15	1.5	20.3	0.3	1.0	4.5	4.5	0.3	4.5	3.0	6.6
E16	27.0	20.3	4.5	24.0	13.5	36.0	6.8	18.0	19.2	10.5
E17	42.2	54.0	13.5	46.5	60.8	40.5	25.3	54.0	44.3	15.8
E18	4.5	27.0	13.5	2.3	9.0	13.5	6.0	9.0	9.0	7.7
E19	4.5	27.0	18.0	40.5	13.5	13.5	6.8	13.5	13.5	11.7
E20	1.1	2.3	5.1	11.3	6.0	4.5	1.2	6.0	4.8	3.3
E21	10.1	7.6	0.2	10.1	27.0	10.1	5.1	15.2	10.1	7.9
E22	6.8	7.6	0.8	3.4	5.1	3.4	3.4	10.1	4.2	3.0
E23	1.7	2.3	2.3	2.3	5.1	1.7	4.5	6.8	2.3	1.9
E24	4.5	27.0	8.4	10.5	6.8	10.1	9.0	10.1	9.6	6.9
E25	2.3	15.2	7.9	1.4	1.1	2.3	3.4	2.3	2.3	4.8
E26	70.9	47.3	33.8	28.5	40.5	67.5	47.3	36.0	43.9	15.5
E27	3.4	54.0	7.6	27.0	30.4	20.3	11.4	16.9	18.6	16.1
E28	12.0	72.0	6.0	24.0	18.0	36.0	13.5	27.0	21.0	20.8
E29	27.0	72.0	20.3	54.0	54.0	162.0	54.0	81.0	54.0	44.0
E30	4.5	6.8	0.1	0.2	6.8	0.8	13.5	5.1	4.8	4.5
E31	6.8	13.5	3.0	40.5	27.0	36.0	30.4	27.0	27.0	13.7
E32	27.0	24.0	15.2	54.0	33.8	54.0	22.5	38.0	30.4	14.4
Median	6.4	20.3	6.4	15.0	13.5	11.8	9.0	13.5		
SD	14.8	23.0	11.5	37.9	25.6	43.2	25.9	28.3		

Appendix III (continued).

Table A2. Individual and median OCRA Checklist scores by exposure ID and by rater.

	Rater Score									
Exposure ID	A	B	C	D	E	F	G	H	Median	SD
E01	12.2	10.5	7.7	12.7	5.5	6.6	17.1	6.6	9.1	4.0
E02	33.9	27.5	34.7	28.3	34.7	29.9	32.3	25.0	31.1	3.7
E03	30.7	24.2	34.7	28.3	18.6	23.4	30.7	18.6	26.2	5.9
E04	2.6	2.6	2.6	3.8	0.0	0.0	3.8	0.0	2.6	1.7
E05	26.8	37.0	59.9	39.5	54.8	35.1	37.0	35.1	37.0	11.1
E06	9.3	5.0	10.6	11.3	6.7	4.0	10.0	2.7	8.0	3.3
E07	4.0	11.6	11.3	7.3	11.3	6.0	10.0	4.7	8.6	3.2
E08	4.4	8.5	14.0	17.8	10.8	6.3	7.8	5.0	8.1	4.6
E09	40.4	46.8	63.8	34.7	33.1	38.0	51.7	31.5	39.2	11.0
E10	43.6	66.2	67.8	45.2	42.8	41.2	58.1	33.9	44.4	12.5
E11	13.0	16.4	31.6	19.5	16.9	12.5	13.4	10.8	14.9	6.6
E12	12.1	7.8	6.1	7.8	4.8	9.1	15.6	9.1	8.4	3.4
E13	12.4	11.9	27.7	15.8	7.9	15.3	22.0	18.7	15.5	6.3
E14	24.9	17.5	39.6	42.4	16.4	25.4	32.2	25.4	25.4	9.5
E15	3.4	1.1	6.8	4.5	2.3	4.5	3.4	2.3	3.4	1.8
E16	6.8	14.7	30.0	10.2	5.7	10.2	15.8	3.4	10.2	8.4
E17	21.5	17.0	22.3	10.0	13.8	12.5	14.3	11.3	14.0	4.6
E18	12.6	15.7	28.3	11.0	16.5	19.7	15.7	15.7	15.7	5.3
E19	12.6	18.9	28.3	22.0	16.5	19.7	22.0	18.9	19.3	4.6
E20	10.6	3.4	20.4	21.7	7.2	14.0	3.4	14.5	12.3	7.0
E21	7.8	7.8	20.7	15.6	6.1	5.2	6.1	5.2	6.9	5.7
E22	7.8	7.8	20.7	14.7	6.1	5.2	2.6	3.5	6.9	6.2
E23	11.1	5.5	21.0	15.5	5.5	4.4	21.0	5.5	8.3	7.1
E24	12.0	3.7	22.3	10.3	14.3	10.0	19.0	10.0	11.1	5.8
E25	11.6	14.4	11.1	4.4	4.4	9.9	14.4	6.6	10.5	4.0
E26	43.6	26.0	66.9	29.3	28.2	22.1	28.2	19.3	28.2	15.5
E27	16.6	32.6	40.9	23.2	11.6	13.8	27.6	8.8	19.9	11.2
E28	35.5	36.3	56.5	35.5	8.1	33.1	35.5	26.6	35.5	13.3
E29	19.4	32.3	76.7	42.0	38.8	47.6	45.2	38.0	40.4	16.4
E30	6.8	10.6	23.0	13.6	0.0	5.1	16.2	5.5	8.7	7.3
E31	16.2	17.0	26.4	21.7	23.4	21.7	19.6	20.0	20.8	3.3
E32	44.2	17.7	59.7	15.5	23.8	24.9	39.8	19.9	24.3	15.6
Median	12.5	15.2	27.0	15.7	11.5	13.2	18.0	11.0		
SD	12.8	14.3	20.5	11.9	13.3	12.4	14.3	10.7		

Appendix III (continued).

Tables A3(a)–(h). Inter-method agreement frequencies for each rater. Dichotomous exposure classifications were used.

Table A3(a). Rater A

OCRA	Strain Index: Safe	Strain Index: Hazardous	Total
Safe	8	3	11
Hazardous	8	13	21
Total	16	16	32

Table A3(b). Rater B

OCRA	Strain Index: Safe	Strain Index: Hazardous	Total
Safe	3	9	12
Hazardous	0	20	20
Total	3	29	32

Table A3(c). Rater C

OCRA	Strain Index: Safe	Strain Index: Hazardous	Total
Safe	5	1	6
Hazardous	11	15	26
Total	16	16	32

Table A3(d). Rater D

OCRA	Strain Index: Safe	Strain Index: Hazardous	Total
Safe	6	3	9
Hazardous	5	18	23
Total	11	21	32

Table A3(e). Rater E

OCRA	Strain Index: Safe	Strain Index: Hazardous	Total
Safe	9	6	15
Hazardous	1	16	17
Total	10	22	32

Table A3(f). Rater F

OCRA	Strain Index: Safe	Strain Index: Hazardous	Total
Safe	11	3	14
Hazardous	2	16	18
Total	13	19	32

Table A3(g). Rater G

OCRA	Strain Index: Safe	Strain Index: Hazardous	Total
Safe	8	0	8
Hazardous	4	20	24
Total	12	20	32

Table A3(h). Rater H

OCRA	Strain Index: Safe	Strain Index: Hazardous	Total
Safe	8	8	16
Hazardous	1	15	16
Total	9	23	32

Appendix III (continued).

Tables A4(a)-(h). Inter-method agreement frequencies for each rater. Three-level exposure classifications were used.

Table A4(a). Rater A

OCRA	Strain Index: Safe	Action	Hazardous	Total
Safe	4	1	1	6
Action	4	7	2	13
Hazardous	0	3	10	13
Total	8	11	13	32

Table A4(b). Rater B

OCRA	Strain Index: Safe	Action	Hazardous	Total
Safe	3	0	3	6
Action	0	2	6	8
Hazardous	0	0	18	18
Total	3	2	27	32

Table A4(c). Rater C

OCRA	Strain Index: Safe	Action	Hazardous	Total
Safe	3	0	0	3
Action	0	4	1	5
Hazardous	6	4	14	24
Total	9	8	15	32

Table A4(d). Rater D

OCRA	Strain Index: Safe	Action	Hazardous	Total
Safe	3	1	0	4
Action	3	1	4	8
Hazardous	1	3	16	20
Total	7	5	20	32

Table A4(e). Rater E

OCRA	Strain Index: Safe	Action	Hazardous	Total
Safe	2	8	2	12
Action	0	1	5	6
Hazardous	1	1	12	14
Total	3	10	19	32

Table A4(f). Rater F

OCRA	Strain Index: Safe	Action	Hazardous	Total
Safe	4	5	1	10
Action	1	3	4	8
Hazardous	0	0	14	14
Total	5	8	19	32

Table A4(g). Rater G

OCRA	Strain Index: Safe	Action	Hazardous	Total
Safe	3	2	0	5
Action	0	3	1	4
Hazardous	0	7	16	23
Total	3	12	17	32

Table A4(h). Rater H

OCRA	Strain Index: Safe	Action	Hazardous	Total
Safe	2	7	3	12
Action	1	1	3	5
Hazardous	0	1	14	15
Total	3	9	20	32