Photograph-based ergonomic evaluations using the Rapid Office Strain Assessment (ROSA)

J. Liebregts, M. Sonne*, J.R. Potvin

Department of Kinesiology, McMaster University, 1280 Main St. W, Hamilton, Ontario L8S 4L8, Canada

A B S T R A C T

The Rapid Office Strain Assessment (ROSA) was developed to assess musculoskeletal disorder (MSD) risk factors for computer workstations. This study examined the validity and reliability of remotely conducted, photo-based assessments using ROSA. Twenty-three office workstations were assessed on-site by an ergonomist, and 5 photos were obtained. Photo-based assessments were conducted by three ergonomists. The sensitivity and specificity of the photo-based assessors' ability to correctly classify workstations was 79% and 55%, respectively. The moderate specificity associated with false positive errors committed by the assessors could lead to unnecessary costs to the employer. Error between on-site and photo-based final scores was a considerable ~2 points on the 10-point ROSA scale (RMSE = 2.3), with a moderate relationship ($\rho = 0.33$). Interrater reliability ranged from fairly good to excellent (ICC = 0.667 −0.856) and was comparable to previous results. Sources of error include the parallax effect, poor estimations of small joint (e.g. hand/wrist) angles, and boundary errors in postural binning. While this method demonstrated potential validity, further improvements should be made with respect to photo-collection and other protocols for remotely-based ROSA assessments.

1. Introduction

Recent developments in information technology have resulted in dramatic increases in occupational computer use (Blatter and Bongers, 2002). In 2000, approximately 60% of Canadian workers reported computer usage as part of their job duties, while 82% of those workers reported daily occupational computer use (Lin and Popovic, 2003; Marshall, 2001). In the United States, 2003 census data showed that over 50% and 60% of all employed men and women, respectively, used a computer as part of their job (Day et al., 2005). This trend of increasing workplace computer use has been associated with an increase in work-related musculoskeletal disorders (WMSDs) and symptoms among workers (Mani and Gerr, 2000). Meta-analyses (Gerr et al., 2006; IJmker et al., 2007) found an overall positive association between computer usage and WMSDs.

Computer use has been identified as a risk for the development of WMSDs. Risk factors of computer use include prolonged non-neutral postures of the: 1) hands and wrists (Jensen et al., 2002; Keir et al., 1999; Marcus et al., 2002), 2) head and neck (Gerr et al., 2002; Marcus et al., 2002) and 3) shoulder, elbow and lower back (Burdorf et al., 1993; Juul-Kristensen et al., 2004).

The commonly used Rapid Upper Limb Assessment (RULA) (McAtamney and Corlett, 1993) was created to identify postures that can lead to WMSDs. While some computer workstation risk factors may be assessed using RULA, it does not necessarily assess risk factors specific to those associated with the set-up of office equipment, such as office chairs, monitors, phone, etc. (Sonne et al., 2012). Computer workstation-specific evaluation tools, such as the modified Rapid Upper Limb Strain Assessment for computer workstations (Lueder,1996; Lueder and Corlett, 1996) and the University of California Computer Checklist (Janowitz et al., 2002), are rapid ergonomics assessment tools for the office workstation. However, both were found to be inadequately associated with symptoms of WMSDs and were inconsistent in their prediction of these symptoms (Menendez et al., 2009). The Strain Index (Moore and Garg, 1995) has also been applied to the office workstation setting, however, it does not appropriately take into account all risk factors in the office, including the equipment and work habits of the user (e.g. telephone location and telephone user strategy). Finally, many generic computer workstation checklists involve dichotomous questions and answers (e.g. 'yes' or 'no' responses) to identify risk factors that do not account for the magnitude of the

* Corresponding author. Department of Kinesiology, McMaster University, 1280 Main St. W, Hamilton, Ontario L8S 4K1, Canada.
E-mail address: sonnemw@mcmaster.ca (M. Sonne).

risk factors, and are not validated against WMSD symptoms (eg. OHSA Office ergonomics checklist, Ontario MOL Office ergonomics checklist).

The Rapid Office Strain Assessment (ROSA) (Sonne et al., 2012) allows ergonomists to quickly quantify risk factors specific to the computer workstation through workplace posture and equipment assessment. Risk factors are weighted based on increasing WMSD risk, and provide users with risk scores for subsets of the workstation (i.e. chair, monitor/telephone, mouse/keyboard) as well as an aggregate ROSA final score from 1 to 10. Original validity testing of the tool (Sonne et al., 2012) found a significant correlation of ROSA final scores with reported discomfort, with a proposed action level score of 5 indicating an increased risk of discomfort for workstations with final scores equal or above that score. In addition, original on-site interrater reliability testing was shown to be high (ICC = 0.88, 0.91, respectively) (Sonne et al., 2012), further demonstrating the effectiveness of the tool.

Conventionally, field office ergonomic assessments have been limited to an in-person, one-on-one assessment format. While this is considered to be a reliable and proven methodology, it may become costly and inefficient for an ergonomist faced with assessing a large number of workstations. As a potential alternative, Sonne and Andrews (2012) investigated the validity of worker self-assessment using ROSA via an online interface; however there was a noted issue with the untrained workers' ability to properly assess their own mouse and keyboard risk factors.

In addition to checklist entries concerning the presence/absence of workstation characteristics (e.g. documents holder, screen glare), a variety of ROSA's checklist entries require the estimation and classification of posture angles into interval bins, including; knee flexion, trunk flexion/extension, neck flexion angles etc. The practicality of office assessments, based on static photos and/or video observations, has seldom been investigated in past literature. However, image observation has been widely investigated in occupational biomechanics experimental settings to analyze the validity and phenomena surrounding joint angle estimation (Li and Buckle, 1999; Abu-Rajab et al., 2010; Baluyut et al., 1995; Bao et al., 2007; Genaidy et al., 1993; Lau and Armstrong, 2011; Liu et al., 1997; Paul and Douwes, 1993). Researchers found satisfactory accuracy of joint angle estimates, with errors ranging from ~2 to 10% (Baluyut et al., 1995; Covalla, 2003; Genaidy et al., 1993; Paul and Douwes, 1993). Also, joint angle bin widths (in degrees) were positively correlated with posture rating reliability (Bao et al., 2009) and accuracy (Van Wyk et al., 2009), with the optimal angle interval widths being found to be 30°.

Bao et al. (2007) compared two static image analysis methods used for postural estimations ("worst-case" posture assessment vs. frame-by-frame posture sampling assessment) using RULA (McAtamney and Corlett, 1993). However, they only compared and contrasted the results between the two image selection approaches, and did not analyze the accuracy of the methods for estimating actual postures. Covalla (2003) compared static and dynamic image observation to evaluate the validity of using of RULA and the Strain Index (Moore and Garg, 1995). It still remains, however, that no studies have investigated the use of static images while using risk analysis tools for office workstations.

On-site office ergonomics assessments are currently the preferred method with ergonomic practitioners. We have proposed that the use of static image observation may be a valid method to improve the efficiency of these assessments. To investigate this proposal, we need to compare remotely performed photo-based assessment scores against on-site assessment scores using the Rapid Office Strain Assessment tool. The purpose of this study is to determine the validity and reliability of remotely performed photo-based ROSA assessments.

2. Methods

2.1. Participants

One ergonomist performed the on-site assessments (Section 2.2), while three ergonomists were recruited to participate in the remote photo-based assessment portion of this study (Section 2.4). In addition, computer workstation users (n = 23; 11 males, 12 females) were recruited to participate in this study as the subjects of the assessments. Each workstation user was required to read a letter of information and sign a consent form before participation in the study. All procedures in this study approved in advance by the university's research ethics board.

2.2. On-site assessment

On-site assessments of workstations were performed by an expert ergonomist trained in using the Rapid Office Strain Assessment (ROSA) tool (Sonne et al., 2012). This ergonomist, henceforth called the 'on-site assessor', had experience working as a consultant in the public sector conducting office ergonomic evaluations of government worker offices. Since the interrater reliability of on-site assessments using ROSA has already been established in previous studies (Sonne et al., 2012; Sonne and Andrews, 2011), just one assessor was chosen to do the on-site assessments.

The on-site assessor was given a 60-min tutorial on using the ROSA tool, and afterwards performed 20 ROSA assessments of office mock-ups using vacant office spaces at the university. This practice allowed the on-site assessor to gain a consistent comfort level and understanding of the tool for a wide variety of workstations. The 23 desktop workstations were evaluated while being operated by the workstation user, and three ROSA sub-scores: 1) chair, 2) monitor/telephone and 3) mouse/keyboard, and a ROSA final score were obtained for each workstation. Estimated work duration was reported by the workstation user. This value was used as the assumed duration of use for all components of the workstation. The on-site assessment scores were considered most accurate, and were used as the reference against which the photo-based assessment errors were calculated.

2.3. Photographs

Five photos were taken of each of the 23 workstations during completion of the on-site assessment, by the on-site assessor. The perspectives were chosen to capture as much workstation information as possible within the fewest number of photos. It was determined that five photos would provide a minimum of two viewing angles for all, or most, body postures during a ROSA office assessment, while including visual information of all workstation features.

The first photo (Fig. 1a) was taken from the sagittal perspective while the workstation user operated the keyboard. The photo captured a fully inclusive view of the workstation user and workstation components (i.e. chair, monitor, keyboard, etc.) and was to be used for the estimation of all chair-related characteristics (e.g. seat pan depth/height, back rest angle, arm rest height, space under desk, etc.), monitor-related factors and postures, keyboard factors (with the exception of wrist deviation), and the presence of an overhead cabinet. The second photo (Fig. 1b) was taken from the same sagittal perspective as the first photo; however the workstation user demonstrated a reach to the telephone. The third photo (Fig. 1c) was taken from the coronal perspective with the workstation user performing a typing task and included the workstation user's head, shoulders, and arms, as well as all workstation components found on the desk surface. This angle was chosen for the
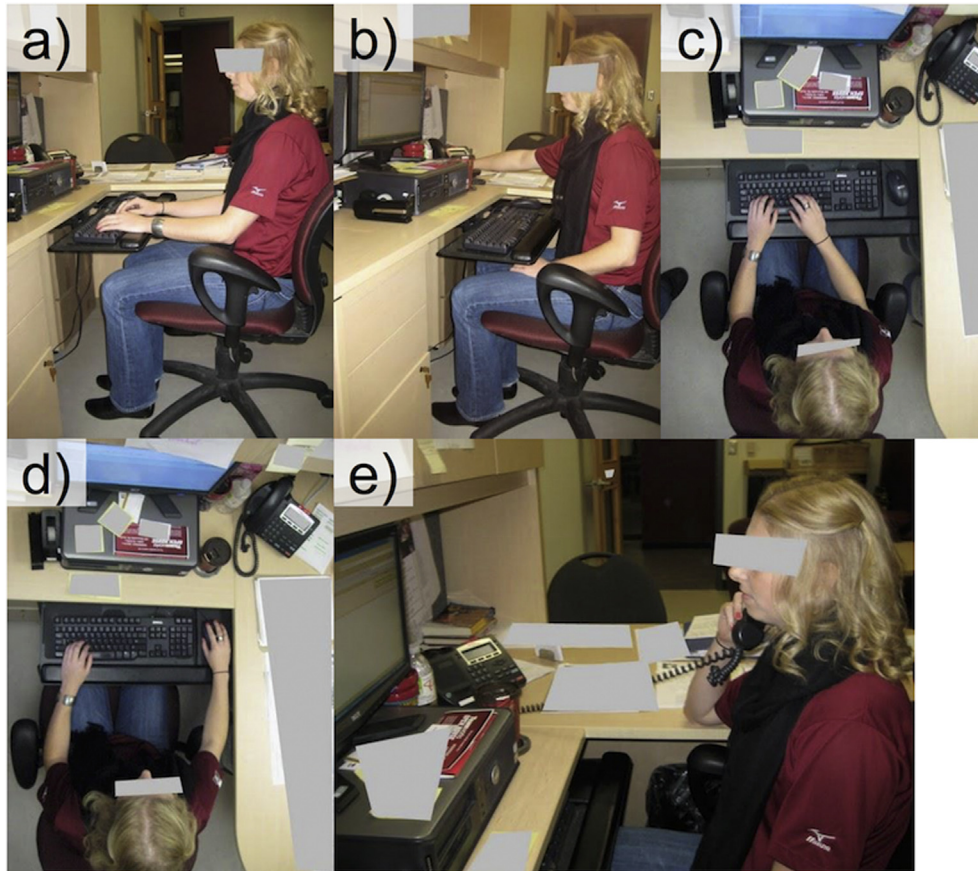
Fig. 1. A sample set of 5 photos from a sample workstation (#10) used for off-site photo-based expert assessments, which were designed to effectively capture all aspects of the office workstation. This particular workstation resulted in a final on-site ROSA score of 4. The photos provided include: a) a full sagittal view of the workstation user performing a typing task, capturing most chair, keyboard, and monitor related factors, b) another full sagittal view of the workstation user demonstrating a reach for their communication device, c) a fully inclusive coronal perspective of workstation user performing a typing task, showing additional keyboard, chair, and monitor related factors, d) another coronal perspective of the workstation user performing a mousing task, and e) workstation user demonstrating typical use of their communication device.

purpose of assessing wrist postures while typing (including wrist deviation), shoulder abduction, neck twist due to monitor position, the presence of a document holder, and monitor glare. The fourth (Fig. 1d) was taken from the same perspective as the third; however, this photo captured a view while the workstation user performed a mousing task. This allowed determination of risk factors/postures associated with the mouse (e.g. reach, grip, etc.). Both the third and fourth photos were taken by the on-site assessor while they stood on a stepstool in the workstation. The fifth photo (Fig. 1e) was taken of the workstation user using their telephone per the workstation user's preference to evaluate telephone technique. All photo sets were then sent electronically to three photo-based assessors.

2.4. Photo-based assessments

Three ergonomists, trained in performing ROSA assessments, used the five-photo set for all 23 workstations (referred to as 'photo-based' assessors). The photo-based assessors were kinesiology graduate students that had experience using the ROSA tool. They also participated as raters in the original interrater reliability testing of ROSA (Sonne et al., 2012). The assessments were performed at an off-site location. To assess work duration, values obtained during the on-site assessment were used by the photo-based assessors as the assumed duration for all workstation components. As with the on-site assessments described above, four outcome

scores were derived for each of the 23 workstations, from each of the three photo-based assessors, which includes three sub-scores (chair, monitor/telephone, mouse/keyboard) and the ROSA final score.

2.5. Statistical analyses

2.5.1. Comparison between on-site and photo-based assessments

Four root mean squared error (RMSE) values were calculated to examine the proximity of the photo-based scores to the on-site assessment scores. This was calculated for each of the ROSA sub-scores (chair, monitor/telephone, mouse/keyboard) as well as the ROSA final scores. For each of the 23 workstations, an RMSE value was calculated by taking the RMS difference between each of the three photo-based scores and the on-site score. For each sub-score and the final score, an overall RMSE value was calculated as the average RMS difference across the 23 workstations.

To further examine the relationship between photo-based scores and the on-site scores, four correlation coefficients were calculated: one for each of the ROSA sub-scores (chair, monitor/telephone, mouse/keyboard), and another for the final score. Since ROSA scores are ordinal data, Spearman's Rho rank-correlation coefficients (ρ) were calculated. This was calculated with the following method: 1) each set of 23 ROSA sub- and final scores (for both the on-site and photo-based assessors), were ranked, relative to the average score for the given set, giving us sets of 23 rank-

values, 2) for each of the three sub-scores and the final score, a ρ value was calculated by correlating the 23 rank-values for each photo-based assessor to the corresponding on-site rank-value set and 3) these three ρ values were then averaged, resulting in an overall ρ-value for each of the four ROSA scores. For interpretation, coefficients of ±0.1–0.3, ±0.3–0.5, and >±0.5 were considered weak, moderate, and strong relationships, respectively (Cohen, 1988).

A factorial analysis of variance (ANOVA) was performed to investigate significant effects of independent variables: assessor (on-site, photo-based assessors 1, 2 and 3) and ROSA score type (chair, monitor/telephone, mouse/keyboard, and final), on the dependent variable of ROSA final and sub-score values. A Tukey's post hoc test was performed for any significant main effects and interactions. Significance level was set to $p < 0.05$ for the analysis.

2.5.2. Interrater reliability

To examine interrater reliability within the photo-based assessments, we used an intraclass correlation coefficient (ICC), with two-way random analysis and absolute agreement (Burt and Punnett, 1999). These values were calculated between all assessor scores within each of the final and three ROSA sub-scores, across the 23 workstations. These values were compared against the interrater ICC values seen in Sonne et al. (2012), which determined the reliability of assessment scores across three assessors. Correlation coefficients were interpreted based on Fleiss's (1986) recommendations, where ICCs greater than 0.75 were considered excellent, between 0.40 and 0.75 were considered fair to good, and less than 0.40 were considered poor (Sutherland et al., 2007).

2.5.3. Functional classification: sensitivity and specificity

From Sonne et al. (2012), ROSA final scores can be classified as either 1) no immediate action required (<5), or 2) further action required (≥5). It is important to examine how a photo-based assessor's ability to correctly classify a workstation at this action level (AL) of ROSA final score ≥ 5. Therefore, sensitivity and specificity tests were performed to determine the likelihood that photo-based final ROSA scores were correctly classified (i.e. ≥AL versus < AL) based on the on-site final ROSA scores. Sensitivity tests determined the probability of a true positive classification, i.e. the chance of a ≥AL photo-based ROSA score if the on-site score is ≥AL for a given workstation. Conversely, specificity tests determined the probability of a true negative classification, i.e. the chance of a <AL photo-based ROSA score if the on-site score is <AL for a given workstation.

3. Results

3.1. Descriptive statistics

The on-site assessment evaluated 11 of the 23 workstations to have a ROSA final score of 5 or higher, indicating further evaluation was required. The photo-based assessors rated ROSA final scores of 5 or higher for 58% of the workstations. The sub-score with highest score was the chair, with an average (±SD) ROSA score of 3.8 (±1.3) for photo-based assessments, and 4.1 (±1.0) for on-site assessments. The lowest score was seen in the monitor/telephone sub-score for the on-site assessments (3.1 (±1.7)), while the lowest score for photo-based assessments was reported for the mouse/keyboard sub-score (3.3 (±1.1)). Table 1 summarizes the results of all ROSA scores for all on-site and photo-based assessments.

3.2. Comparison between on-site and photo-based assessments

The ROSA final score had an RMSE of 2.29, which implies that photo-based assessors erred from the on-site score by approximately

2.3 points (23%) on the 10-point ROSA scale. The monitor/telephone sub-scores showed the greatest amount of photo-based assessment error (Table 2). The average monitor/telephone sub-score was 3.7 (±1.6), compared to the on-site assessment value of 3.1 (±1.7) (Table 1). The lowest amount of error was found with the chair sub-score, with an RMSE of 2.1 (±1.0) (Table 2). The photo-based assessments average was 3.8 (±1.3), while the on-site assessment average was 4.1 (±1.0).

On-site and photo-based assessment scores were most strongly correlated for the monitor/telephone sub-score ($\rho = 0.57$). The weakest relationship between assessment types was in the chair sub-score, with a $\rho = 0.40$. The relationship between assessment types was moderate for the ROSA final score, with a rank-correlation of $\rho = 0.38$ (Table 2). Examination of chair assessments illustrated that seat pan depth was incorrectly assessed by photo-assessments 47.8% of the time.

Based on the ANOVA results, no significant differences were found between assessors. There was a significant effect of score type, however this was an expected result due to the nature of ROSA; the sub-scores are always smaller in value by design compared to the final score.

3.3. Interrater reliability

ROSA final scores showed fair to good reliability between photo-based assessments from three assessors, yielding an ICC of 0.74. The highest reliability was found with the monitor/telephone sub-score, whereas the lowest reliability was found with the mouse/keyboard sub-score (Table 3). The reliability scores for the chair, and monitor/telephone were higher for photo-based assessments than those seen when comparing the three on-site assessors (from Sonne et al., 2012), but were lower for the mouse/keyboard sub-score and ROSA final score (Table 3). Meanwhile, in the original ROSA validation study, Sonne et al. (2012) found corresponding ICC's of 0.91 for final scores, and 0.51, 0.74, and 0.83 for each respective sub-score.

3.4. Functional classification: specificity and sensitivity

Photo-based assessments had a 55% specificity rating and a 79% sensitivity rating. These values represent the probability that a photo-based assessment would yield either a true negative (specificity) or a true positive (sensitivity) ROSA score. Across the 69 total photo-based ROSA scores, a "false negative" (score of 4 or less when the on-site score was 5 or more) was committed 9 times, while a "false positive" (scoring 5 or more when the on-site score was 4 or less) was committed 16 times (Table 4).

However, when looking into the data in detail, we see that 7 of the 9 false negatives committed by the photo-based assessors were a ROSA score of 4. Similarly, 10 of the 16 false positives were a ROSA score of 5, while the remaining 6 false positives were a ROSA score of 6.

4. Discussion

4.1. Validity of photo-based assessments

The most important finding in this study is that remotely performed, photo-based ergonomic assessments showed potential as a valid assessment method; however, improvements must be made before an intervention program is implemented in the workplace. While RMSE results indicated that a photo-based assessor could commit significant error on a final ROSA score, the sensitivity and specificity results (Section 3.4) demonstrate that the photo-based assessors were more likely to correctly classify a workstation than

Table 1
Summary of mean (SD) ROSA sub-scores and final scores. Photo-based assessments were conducted remotely by three assessors by viewing five photographs from each of 23 workstations. The on-site assessments were conducted at the workstation by an assessor, who also photographed the workstation for photo-based assessment.

	Chair	Monitor/Telephone	Mouse/Keyboard	Final
RMSE (SD)	2.11 (1.00)	2.29 (1.55)	2.15 (1.27)	2.29 (1.12)
Spearman's Rho (ρ)	0.32[a]	0.60[b]	0.40[a]	0.33[a]

[a] Moderate relationship.
[b] Strong relationship.

Table 2
Summary of RMSE (SD) and Spearman's Rho rank-correlation coefficients (ρ), comparing the photo-based assessment ROSA scores against the on-site assessment ROSA score.

	Chair	Monitor/Telephone	Mouse/Keyboard	Final
On-site assessor	4.1 (1.0)	3.1 (1.7)	3.4 (1.4)	4.7 (1.4)
Photo-based assessor 1	3.7 (1.4)	3.7 (1.7)	3.2 (1.1)	4.7 (1.4)
Photo-based assessor 2	3.7 (0.9)	4.1 (1.7)	3.7 (1.4)	4.8 (1.1)
Photo-based assessor 3	4.0 (1.6)	3.3 (1.3)	2.8 (1.3)	4.4 (1.4)
All photo-based	3.8 (1.3)	3.7 (1.6)	3.3 (1.3)	4.6 (1.3)
All assessors	3.9 (1.3)	3.6 (1.6)	3.3 (1.3)	4.7 (1.3)

Table 3
Summary of ICC results (ICC), compared against ICC results from the Sonne et al. (2012) original ROSA validation study (Original ICC).

	Chair	Monitor/Telephone	Mouse/Keyboard	Final
ICC	0.74[b]	0.86[c]	0.67[b]	0.74[b]
Original ICC[a]	0.51[b]	0.74[b]	0.83[c]	0.91[c]

[a] Sonne et al. (2012).
[b] Fair to good.
[c] Excellent reliability.

to incorrectly classify one. We argue that, above all other metrics, the most important indicator of practical performance is whether a photo-based assessor could correctly classify workstation risk (i.e. further/no further assessment needed).

The majority of photo-based assessors' misclassifications were "false positives" (i.e. workstation was incorrectly deemed to require further assessment), rather than false negatives. Therefore, one could argue that photo-based assessments are valid and practical, since false positives are less consequential as they are more conservative and they do not pose undue risk on workstation users in

Table 4
Table of ROSA final scores, across all 23 workstations, as reported by the on-site assessor ("on-site") and the three photo-based assessors (PB1, PB2, PB3), and sorted by the value of the on-site ROSA scores in ascending order. Scores that are <5 are shown in green, while scores that are ≥5 are shown in red. False negatives and false positives are denoted using bold text.

Workstation	On-Site	PB1	PB2	PB3
On-Site ROSA Final Value <5				
1	3	2	2	1
2	3	2	4	4
3	3	**5**	4	**5**
4	3	**5**	**6**	**6**
5	4	3	3	2
6	4	3	4	4
7	4	4	**5**	4
8	4	**5**	**5**	4
9	4	**6**	**5**	3
10	4	**6**	4	**5**
11	4	**5**	**6**	4
12	4	4	**6**	**5**
On-Site ROSA Final Value ≥ 5				
13	5	5	**4**	**3**
14	5	**4**	5	**4**
15	5	5	5	5
16	5	6	6	6
17	6	5	5	**3**
18	6	**4**	6	**4**
19	6	5	5	5
20	6	5	5	7
21	6	8	6	5
22	7	5	**4**	7
23	8	6	6	**4**

terms of potential injury. However, employers would be negatively impacted due to false positive scores, which create an undue number of further assessments on what are actually "low-risk" workstations. In conclusion, the tested method, of remotely performed photo-based ROSA assessments, showed promise. However, real-world ergonomic intervention programs are often implemented to benefit both the employees *and* the employer; while this method showed that the benefits for the latter may be mitigated due to "false positive" assessment scores.

Measures comparing on-site and photo-based ROSA scores may indicate further issues in the photo-based method. As stated in the Results, the photo-based assessors committed up to 23% error based on the RMSE, or the equivalent of 2.3 points on the ROSA final scale (Table 2). We see this as a significant proportion of the 10-point ROSA scale, and must be improved on before a photo-based method can be recommended. The correlation results showed that the relationship between on-site and photo-based assessments was moderate for the chair and mouse/keyboard sub-scores and strong for monitor/telephone sub-scores (Table 2). Most importantly, however, the relationship was only moderate for the ROSA final score and was among the weakest relationship of all ROSA scores. Although these rating errors did not necessarily have a substantial effect on the assessors' ability to classify workstations, they do signify that there is room for improvement for gaining accuracy in scoring.

4.2. Interrater reliability

Based on the ANOVA results, our assessors were not significantly different in their ROSA sub-score or final score ratings. Furthermore, ICC results showed that the overall reliability between expert assessor final and sub-scores were generally fairly good to excellent, with the chair, monitor/telephone, and final scores showing excellent ICC values (Portney and Watkins, 2000) (Table 3). Meanwhile, a moderate reliability was reported for mouse/keyboard sub-scores.

In comparison to previous studies, Sonne et al. (2012) reported comparable values during original validation and reliability testing of the ROSA tool. Meanwhile, interrater reliability was previously been shown to be strong in video-based reliability testing, as seen in the development of the Rapid Upper Limb Assessment tool (McAtamney and Corlett, 1993). Other visual observation studies have found equivalent or slightly lower values of intraclass correlation (Bao et al., 2009; Lowe, 2004; Sutherland et al., 2007); however this is an expected result. Since past studies have analyzed reliability of absolute joint angle estimation, or used higher resolution joint angle categorization binning, it is safe to assume that ROSA's use of such low resolution bins would likely result in higher interrater reliability.

In a review, Covalla (2003) reported assessor reliability ranged from moderate to high in static image studies, while dynamic image analyses ranged from low to high. Based on the above, it can be proposed that use of photos has a strong reliability comparable to other modes, including dynamic image analysis, and proven modes of assessment such as on-site assessments.

4.3. Potential sources of error

Our literature review suggests that photo collection problems may have had a larger influence on our photo-based posture assessment than initially suspected. Primarily, it is likely that perspective error had a large effect on the estimation errors. Any deviation of the camera angle from the orthogonal plane of a joint may result in perspective error due to the parallax effect (Fig. 2). The parallax effect occurs when a viewing angle deviates from an "orthogonal" perspective (i.e. deviation from a 0° or 90° viewing angle). The effect has been shown to negatively affect our ability to correctly estimate joint angles (Lau and Armstrong, 2011; Liu et al., 1997; Paul and Douwes, 1993; Perry et al., 2008; Sutherland et al., 2007). While the photographer in our study attempted to keep as close to the orthogonal plane as possible, it is safe to assume that all photos provided at least some perspective error (or parallax effect) due to the conditions of the data collection environment.

The distance between the workstation user and the camera during photo collection is another potential source of error. Paul and Douwes (1993) found that full-body images were optimally observed when taken from a distance of 4.5 m; far enough not to distort the head and feet due to their extreme positions at the top and bottom of the image. However, we had many instances where obtaining photos from this recommended minimum distance was not possible due to spatial office constraints. Therefore, there are some spatial requirements necessary for capturing an optimal photo-set for these assessments.

Photo-based assessor estimation errors were another potential source for ROSA score errors. First, studies have shown that neutral joint postures are more accurately estimated than those with greater joint deviations (Abu-Rajab et al., 2010; Lau and Armstrong, 2011). Thus, it is likely that some workstations may not have been as accurately rated due to an increased presence of extreme body postures. Second, larger joints (e.g. hip, trunk, and knee, etc.), such as those associated with the chair sub-scores, have been found to be easier to estimate in past studies compared to smaller joints (e.g. wrists, etc.) (Baluyut et al., 1995). Thus, entries associated with the mouse/keyboard sub-score, such as wrist flexion/extension angle or presence of wrist deviation, can difficult to accurately rate for photo-based assessors. Studies have also shown that small joint estimation has only to moderate interrater reliability (Bao et al., 2009; Lowe, 2004). The literature indicates that extreme postures and smaller joint postures may be hard to visually estimate with static images. Future studies must take these factors into account when investigating static image posture estimation.

The boundary issue in posture estimation and binning assessments has been investigated in past literature, and may be of concern when performing posture binning assessments like those required when using ROSA (Andrews et al., 2008). The boundary issue is often present when an observed joint angle is close to the boundary point between two posture bin intervals, such that bin classification errors may increase regardless of rater experience. It was also stated, in the original RULA validation study, that discrepancies between assessors were most often found when body segment angles were at the border between two ranges (McAtamney and Corlett, 1993). While a benefit of using low resolution bins, such as with ROSA, might be beneficial for avoiding the boundary issue, it remains a concern for a posture assessment of this nature.

4.4. Limitations & recommendations

There were some limitations in our study that should be noted. Photo-based assessors were not able to directly measure distances (i.e. seat pan edge to back of knee, distance between face and monitor, distance to phone) from the photos, or judge some characteristics (hardness of arm rests, adjustability of features, hands-free option of phone, screen glare, sufficiency of under-desk space for free movement of legs) while doing photo-based assessments. However, fulfillment of these ROSA entries is a requirement for a thorough checklist completion. To overcome this issue, perhaps it may be necessary for the workstation user to perform a brief supplementary questionnaire. This would provide the assessor with all the aforementioned information that would otherwise be
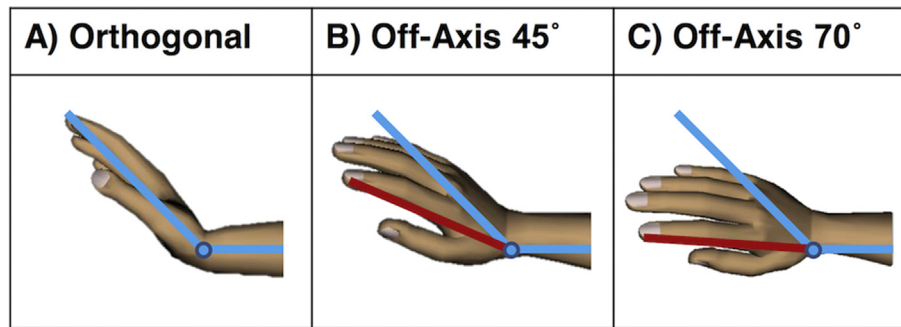
Fig. 2. Sample illustration of the parallax effect using images taken from digital human modeling program Jack 7.1 (Siemens, Ann Arbor, MI). All three images show a human wrist in the exact same extended posture. The far left image shows the wrist from an orthogonal perspective, which would theoretically be the most accurately estimated posture due to zero parallax effect. Meanwhile, the middle and right images show the same wrist posture from perspectives that are approximately 45° and 70° deviated, respectively, from the orthogonal plane. The blue line shows the actual angle of extension, while the red lines in the middle and right images show the perceived angle. It is evident that, due to the change in perspective, our determination of the wrist posture may be affected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

impossible to obtain or be accurately rated simply using photographs. To expand on this point, ROSA has been shown to be a valid self-assessment tool (Sonne and Andrews, 2012), so perhaps a completed ROSA performed by the workstation user could be provided to the assessor as a supplement to the photos. This would provide the assessor with ROSA entries that would be perhaps more accurately rated by the workstation user than the assessor using photos. This process would need to be validated in future studies.

The photos taken in this study only provided a total of five frames of reference of the activities performed by the workstation user. The photos do not account for any possible changes in posture throughout the day, changes in workstation setup that the workstation user may routinely perform, or even changes in posture mousing, keyboarding etc. Furthermore, the workstation user was aware that photos were being taken, and may have been inclined to statically "pose" for the photos, rather than perform their work activities naturally. It is possible that these issues could be resolved with more sampling, or even video analysis. While Ortiz et al. (1997) concluded that postural variability is low enough for computer users throughout the day to justify a single measurement of posture, this is not unanimous amongst past studies.

The spatial constraints (e.g. walls, furniture, irregularly shaped desks, etc.) of the office were a more severe limitation, to our photo collection protocol, than initially expected. As previously mentioned, it was sometimes very difficult to get the camera at a proper angle and sufficient distance to be able to acquire the five photos that would best capture the entire worker while minimizing the effects of parallax due to these constraints. For this study, it was important to assess real-life office environments, rather than constraint-free, lab-simulated environments, since they best reflect what an ergonomist may experience when performing remotely-based assessments. We have learned from this experiment that a future photo, or even video, collection protocol should be designed to account for the spatial constraints of an office environment.

5. Conclusions

The purpose of our study was to evaluate the validity and reliability of ROSA office workstation ergonomic assessments performed remotely using a series of photographs. The photo-based assessment results were compared the results of on-site assessments. The functional classification results indicated potential validity of the tested method, but showed that misclassifications could create undue costs to employers. Meanwhile, RMSE results and moderate correlation results showed that there is room for

improvement in the accuracy of photo-based scoring. On a positive note, fairly good to excellent interrater reliability values showed that there was adequate consistency between photo-based raters when required to perform these remote, photo-based, assessments.

At this point, we propose that remotely performed, photo-based assessments would only be most appropriate when applied as a "triage" method for a large group of office workstations. However, issues including; the parallax effect, small joint angle estimation, and photographic collection protocol must be accounted for in future protocols. This may include an improved photo, or perhaps video, collection protocol, and even supplementary information provided by the workstation user. These improvements should be made to continue towards developing a valid, remotely-performed office ergonomics assessment method that can reduce the burden on proactive ergonomists who are responsible for a large number of computer workstation users.

References

Abu-Rajab, R.B., Marsh, A., Young, D., Rymaszewski, L.A., 2010. Visual estimation of joint angles at the elbow. Eur. J. Orthop. Surg. Tr. 20 (6), 463—467.

Andrews, D.M., Arnold, T.A., Weir, P.L., van Wyk, P.M., 2008. Errors associated with bin boundaries in observation-based posture assessment methods. Occup. Ergon. 8, 11—25.

Baluyut, R., Genaidy, A.M., Davis, L.S., Shell, R.L., Simmons, R.J., 1995. Use of visual perception in estimating static postural stresses: magnitudes and sources of errors. Ergonomics 38 (9), 1841—1850.

Bao, S., Howard, N., Spielholz, P., Silverstein, B., Polissar, N., 2009. Interrater reliability of posture observations. Hum. Factors 51 (3), 292—309.

Bao, S., Howard, N., Spielholz, P., Silverstein, B., 2007. Two posture analysis approaches and their application in a modified rapid upper limb assessment evaluation. Ergonomics 50 (12), 2118—2136.

Blatter, B.M., Bongers, P.M., 2002. Duration of computer use and mouse use in relation to musculoskeletal disorders of neck or upper limb. Int. J. Ind. Ergon. 30, 295—306.

Burdorf, A., Naaktgeboren, B., de Groot, H.C.W.M., 1993. Occupational risk factors for low back pain among sedentary workers. Occup. Environ. Med. 35 (12), 1213—1220.

Burt, S., Punnett, L., 1999. Evaluation of interrater reliability for posture observations in a field study. Appl. Ergon. 30, 121—135.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. Lawrence Erlbaum Associates, Hillsdale, New Jersey, USA.

Covalla, E.D., 2003. Visual Posture Observation Error and Training. North Carolina State University.

Day, J.C., Janus, A., Davis, J., 2005. Computer and Internet Use in the United States, 2003. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Fleiss, J.L., 1986. The Design and Analysis of Clinical Experiments. John Wiley and Sons, Toronto.

Genaidy, A.M., Simmons, R.J., Guo, L., Hidalgo, J.A., 1993. Can visual perception be used to estimate body part angles? Ergonomics 36 (4), 323—329.

Gerr, F., Marcus, M., Ensor, C., Kleinbaum, D., Cohen, S., Edwards, A., Gentry, E., Ortiz, D.J., Monteilh, C., 2002. A prospective study of computer users: I. Study

design and incidence of musculoskeletal symptoms and disorders. Am. J. Ind. Med. 41, 221—235.

Gerr, F., Monteilh, C.P., Marcus, M., 2006. Keyboard use and musculoskeletal outcomes among computer users. J. Occup. Rehabil. 16, 265—277.

IJmker, S., Huysmans, M.A., Blatter, B.M., van der Beek, A.J., van Mechelen, W., Bongers, P.M., 2007. Should office workers spend fewer hours at their computer? A systematic review of the literature. J. Occup. Environ. Med. 64 (4), 211—222.

Janowitz, I., Stern, A., Morelli, D., 2002. Validation and field testing of an ergonomic computer use checklist and guidebook. Human factors and Ergonomics Society annual meeting proceedings. Ind. Ergon. 5, 1007—1011.

Jensen, C., Finsen, L., Søgaard, K., Christensen, H., 2002. Musculoskeletal symptoms and duration of computer and mouse use. Int. J. Ind. Ergon. 30, 265—275.

Juul-Kristensen, B., Søgaard, K., Stroyer, J., Jensen, C., 2004. Computer users' risk factors for developing shoulder, elbow and back symptoms. Scand. J. Work Environ. Hea. 30 (5), 390—398.

Keir, P.J., Bach, J.M., Rempel, D., 1999. Effects of computer mouse design and task on carpal tunnel pressure. Ergonomics 42, 1350—1360.

Lau, M.H., Armstrong, T.J., 2011. The effect of viewing angle on wrist posture estimation from photographic images using novice raters. Appl. Ergon. 42 (5), 634—643.

Lueder, R., 1996. A proposed RULA for computer users. In: Proceedings of the Ergonomics Summer Workshop. UC Berkley Center for Occupational and Environmental Health Continuing Education Program, San Francisco August 8e9, 1996.

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.

Lin, Z., Popovic, A., 2003. Working with Computers in Canada: an Empirical Analysis of Incidence, Frequency and Purpose. Final report. Human Resources Development Canada. Catalog: RH63—1/574-05-03E.

Liu, Y., Zhang, X., Chaffin, D., 1997. Perception and visualization of human posture information for computer-aided ergonomic analysis. Ergonomics 40 (8), 818—833.

Lowe, B.D., 2004. Accuracy and validity of observational estimates of wrist and forearm posture. Ergonomics 47 (5), 527—554.

Lueder, R., Corlett, E.N., 1996. A proposed RULA for computer users. In: Proceedings of the Ergonomics Summer Workshop, UC Berkeley Center for Occupational and Environmental Health Continuing Education Program, San Francisco.

Mani, L., Gerr, F., 2000. Work related upper extremity musculoskeletal disorders. Occup. Environ. Med. 27, 845—864.

Marcus, M., Gerr, F., Monteilh, C., Ortiz, D.J., Gentry, E., Cohen, S., Edwards, A., Ensor, C., Kleinbaum, D., 2002. A prospective study of computer users: II. Postural risk factors for musculoskeletal symptoms and disorders. Am. J. Ind. Med. 41, 236—249.

Marshall, K., 2001. Working with Computers. Statistics Canada, pp. 9—15. Perspectives. Catalog: 75-001.

McAtamney, L., Corlett, E.N., 1993. RULA: a survey method for the investigation of work-related upper limb disorders. Appl. Ergon. 24, 45—57.

Menendez, C.C., Amick (III), B.C., Chang, J., Harrist, R.B., Jenkins, M., Robertson, M., Janowitz, I., Rempel, D.M., Katz, J.N., Dennerlein, J.T., 2009. Evaluation of two posture survey instruments for assessing computing postures among college students. J. Work 34 (4), 421—430.

Moore, J.S., Garg, A., 1995. The strain index: a proposed method to analyze jobs for risk of distal upper extremity disorders. Am. Ind. Hyg. Assoc. J. 56 (5), 433—458.

Ortiz, D.K., Marcus, M., Gerr, F., Jones, W., Cohen, S., 1997. Measurement variability in upper extremity posture among VDT users. Appl. Ergon. 28 (2), 139—143.

Paul, J.A., Douwes, M., 1993. Two-dimensional photographic posture recording and description: a validity study. Appl. Ergon. 24 (2), 83—90.

Perry, M., Smith, A., Straker, L., Coleman, J., O'Sullivan, P., 2008. Reliability of sagittal photographic spinal posture assessment in adolescents. Adv. Physiother. 10 (2), 66—75.

Portney, L.G., Watkins, M.P., 2000. Foundation of Clinical Research Applications to Practice. Prentice Hall, Inc., New Jersey.

Sonne, Michael, Andrews, David, M., 2011. The rapid office strain assessment (ROSA): validity of online worker self-assessments and the relationship to worker discomfort. Occup. Ergon. 10 (3), 83—101.

Sonne, M., Andrews, D.M., 2012. Validity of online worker self-assessments and the relationship to worker discomfort. Occup. Ergon. 10, 83—101.

Sonne, M., Villalta, D.L., Andrews, D.M., 2012. Development and evaluation of an office ergonomic checklist: ROSA — Rapid Office Strain Assessment. Appl. Ergon. 43 (1), 98—108.

Sutherland, C., Albert, W.J., Wrigley, A.T., Callaghan, J.P., 2007. The effect of camera viewing angle on posture assessment repeatability and cumulative spinal loading. Ergonomics 50 (6), 877—889.

Van Wyk, P.M., Weir, P.L., Andrews, D.M., Fiedler, K.M., Callaghan, J.P., 2009. Determining the optimal size for posture categories used in video-based posture assessment methods. Ergonomics 52 (8), 921—930.