DISSERTATION

ERGONOMIC EXPOSURE ASSESSMENT: A STUDY OF RATER RELIABILITY, METHOD RELIABILITY, AND SAMPLING STRATEGY

Submitted by

Angela L. Dartt

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2010

COLORADO STATE UNIVERSITY

June 30, 2010

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY ANGELA L. DARTT ENTITLED ERGONOMIC EXPOSURE ASSESSMENT: A STUDY OF RATER RELIABILITY, METHOD RELIABILITY, AND SAMPLING STRATEGY BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Peter Chen

David Gilkey

William Brazile

David Douphrate

Advisor: John C. Rosecrance

Department Head: Jac Nickoloff

ABSTRACT OF DISSERTATION

ERGONOMIC EXPOSURE ASSESSMENT: A STUDY OF RATER RELIABILITY, METHOD RELIABILITY, AND SAMPLING STRATEGY

Poor characterization of exposures due to inadequately tested ergonomics exposure assessment tools contributes to the skepticism regarding the work-relatedness of musculoskeletal disorders. Due to their ability to capture individual exposures for large populations, observational methods have been commonly used to assess awkward postures in occupational settings. However, use of observation-based methods is complicated due to infrequent assessment of reliability and validity. While direct instrumentation is typically recommended for assessment of awkward postures, application of direct instrumentation in large field studies has been limited.

Evaluation of reliability, validity, and sampling strategies are critical for ergonomic exposure assessment tools, particularly for research that attempts to establish a causal relationship between ergonomic risk factors and musculoskeletal outcomes. The results of this dissertation research addressed rater reliability, method reliability, and sampling strategy concerns for a computer-based observation tool and direct measurement devices known as an inclinometer. In general, the results from this dissertation research indicated: observation of postures using a video-based assessment tool demonstrated moderate to high inter- and intra-rater reliability for the majority of anatomical areas and body parts evaluated; comparison of a video-based posture assessment tool and inclinometry demonstrated moderate to high correlation for the majority of anatomical areas and body parts evaluated; and, evaluation of sampling strategies of posture assessment using inclinometry demonstrated that two to four hours of sampling may be sufficient when assessing postures of the upper arms and trunk. This dissertation research provided critical information regarding the need for improved exposure assessment techniques in the field of ergonomics.

Angela L Dartt
Environmental and Radiological Health Sciences
Colorado State University
Fort Collins, CO 80523
Summer 2010

ACKNOWLEDGMENTS

I would like to acknowledge my advisor, Dr. John Rosecrance, for his support and guidance during my graduate research.

I would like to acknowledge my committee members, Dr. David Douphrate, Dr. Peter Chen, Dr. Bill Brazile, and Dr. David Gilkey, for their support and guidance during my graduate research.

I would like to thank NIOSH for partial financial support during my dissertation research.

I would like to thank the University of Iowa researchers that made my project possible: Dr. Dan Anton, Dr. Fred Gerr, and Linda Merlino. I would also like to thank Jason Tomsic and Andy Troccoli of New Belgium Brewing for allowing part of this research to be collected at their facility.

I would like to especially acknowledge my husband and family. My husband has provided constant support, patience, and love throughout my graduate career. My parents' love and support have been a constant force in attaining by goals in life and in academia.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SECTION FIVE
STUDY III: INTER-METHOD RELIABILITY

SECTION SIX
STUDY IV: SAMPLING STRATEGY

SECTION ONE

INTRODUCTION

BACKGROUND

In two comprehensive reviews of the epidemiological literature, the National Institute for Occupational Safety and Health (NIOSH, 1997) and the National Research Council (NRC/IOM, 2001) reported significant associations between physical occupational risk factors (awkward postures, force, repetition, and vibration) and musculoskeletal disorders (MSDs) of the upper extremities. These reports also reported strong evidence of a causal relationship between awkward postures and neck/shoulder disorders and a combination of physical risk factors and upper extremity disorders. Despite these findings, skepticism remains about the work-relatedness of MSDs due to conflicting study results, flaws in exposure assessment techniques, and little evidence of a concrete causal relationship between risk factors and musculoskeletal outcomes.

Poor characterization of exposures due to inadequately tested exposure assessment techniques contributes to the skepticism regarding the work-relatedness of MSDs. Many ergonomic exposure assessment tools have been developed to help assess occupational risk factors, including awkward postures. These tools range from indirect methods that provide qualitative estimates to direct methods that yield quantification of physiologic responses. Awkward postures have been assessed by analyzing the frequency of extreme joint motion, duration in a specific posture, and magnitude of joint angle (Karhu et al., 1997; Keyserling, 1986; van der Beek et al., 1992; McAtamney and Corlett, 1993; Wiktorin et al., 1995; Fransson-Hall et al., 1995; Ergonomics Analysis and Design Research Consortium, 2003). Due to their low cost and ability to capture individual exposures for large populations, observational methods have been commonly used to assess awkward postures in occupational settings. Observational tools range from simple checklists and diagrams to computer-based programs (Priel, 1974; Karhu et al., 1977; Keyserling, 1986; van der Beek et al., 1992; McAtamney and Corlett, 1993; Fransson-Hall et al., 1995; Wiktorin et al., 1995; Hignett and McAtamney, 2000). Based on research investigating the utility of self-report, video observation, and direct measurements, Spielholz et al. (2001) concluded that video analysis is the most reasonable choice for large epidemiological studies based on analysis and cost requirements. In addition, observational methods for assessing posture are more widespread in industry than any other ergonomic exposure assessment tool (Genaidy et al., 1993). Recently, the technology of small accelerometry devices has been applied to wearable inclinometers or tilt meters. These direct measurement tools can be utilized to capture body postures during working tasks.

The application of observation methods for assessing physical exposure depends on their reliability and validity. Reliable and valid exposure measurements are critical when used to determine causal relationships (or even associations) between occupational risk factors and health outcomes. Few occupational health studies have evaluated or reported the reliability and validity of exposure assessment techniques that have been utilized to predict adverse health outcomes (Burt and Punnett, 1999). Evaluation of exposure to awkward postures is not only dependent on the usability of the measurement tool, but also on the sampling strategy applied.

The proposed investigation is novel because little research exists regarding the application of current computer-based ergonomics exposure assessment tools particularly in regards to comprehensive reliability and validity studies. In addition, this study will explore exposure variability and sampling strategy using direct technical measurements over full work shifts. The proposed study is significant because the lack of well-defined exposure assessment methods and exposure sampling methodologies is a primary issue associated with epidemiological studies of musculoskeletal disorders. The results of this study will address rater reliability, method reliability, and sampling strategy concerns for a computer-based observation tool and direct measurement device and provide critical information regarding the need for improved exposure assessment techniques in the field of ergonomics.

SUMMARY OF THE DISSERTATION STUDIES

Sections Three through Six of this dissertation document discuss each of the dissertation studies. Each dissertation study was built on the previous study in an effort to evaluate current exposure assessment tools and methodologies in the field of ergonomics. While much research exists for observational methods, there is little research regarding reliability and validity for observational methods that employ video-based observation assessment programs. In addition, there is little research that has collected full-shift direct assessment of exposures. The dissertation studies are described below and a schematic has been provided (Figure 1.1).

I Rater Reliability

II Rater Reliability Expanded

III Inter-Method Reliability

IV Sampling Strategy

Study I evaluated the inter- and intra-rater reliability of assessing upper limb postures of workers performing cyclic manufacturing tasks. Assessment of neck, shoulder, and wrist postures of 20 manufacturing employees was conducted by two raters observing continuous digital video footage using a computer-based observation software program. Study I has been published in a peer reviewed journal (Dartt et al., 2009).

Study II expanded the methodologies of Study I. Study II attempted to further evaluate inter- and intra-rater reliability, particularly the variables leading to lower reliability when assessing postures of the upper limbs from video. Assessment of neck, shoulder, and wrist postures of 39 manufacturing workers performing cyclic and non-cyclic tasks was conducted by four raters observing continuous digital video footage using a computer-based observation software program. Study II was funded was funded as a pilot research project by the Rocky Mountain Center for Occupational and Environmental Health in 2008.

Study III evaluated two exposure assessment tools utilized in ergonomics research. The two exposure assessment tools evaluated consisted of a computer-based observation software program used to analyze and log postures from video-recorded tasks taken in the field and an inclinometer used to obtain direct readings of postures in the field. While direct measures are generally recommended, their use has been limited in large field studies. The two tools were used simultaneously to analyze postures of the upper arms and trunk in a manufacturing environment. Results obtained from the two tools were analyzed to evaluate inter-method reliability.

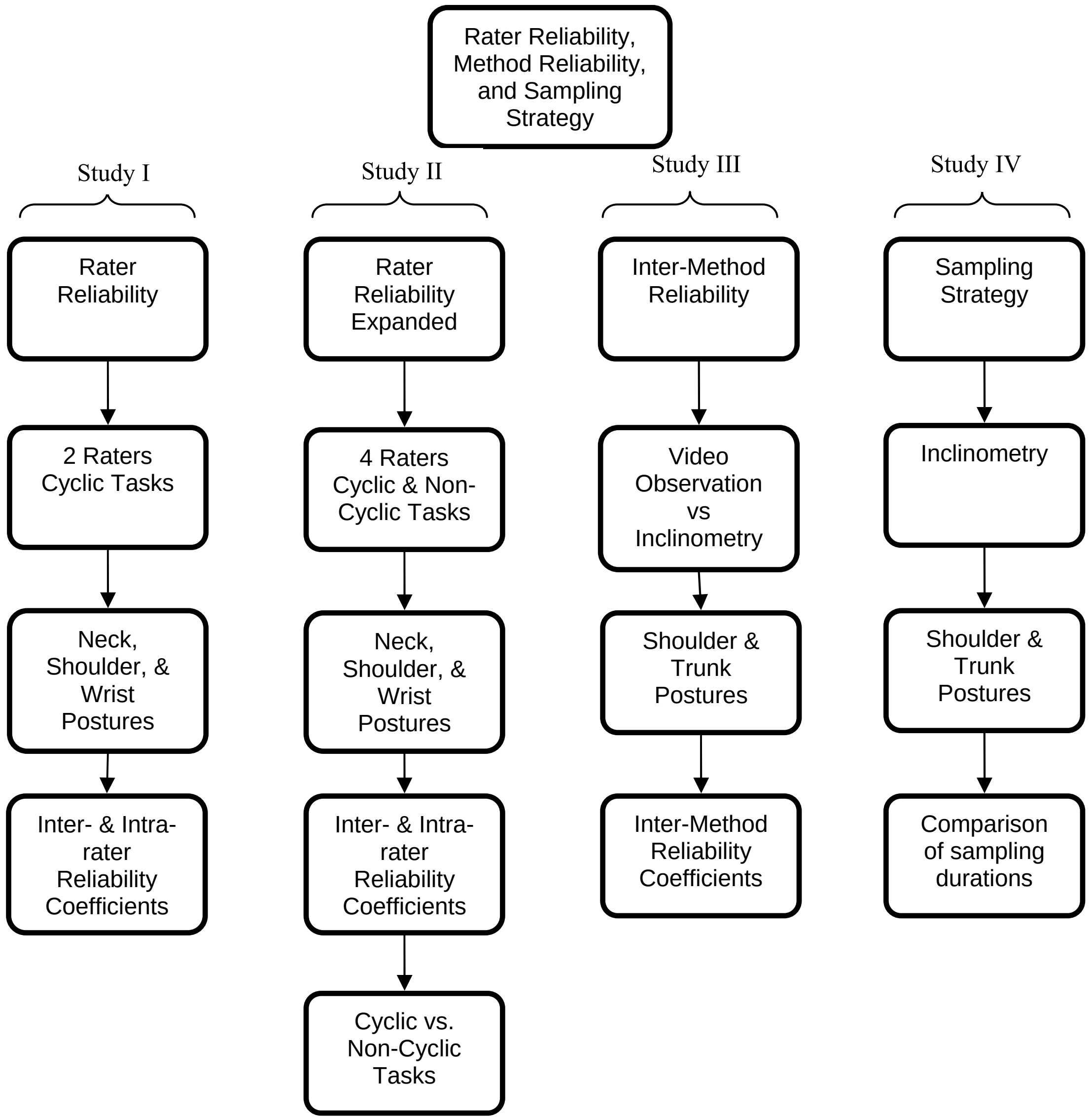
Study IV evaluated exposure variability and sampling strategy for an inclinometer used to assess posture in ergonomics research. The direct reading instrument evaluated was a pager-sized inclinometer that measures position relative to gravity. The purpose of this study was to compare full-shift inclinometry of the upper arms and trunk with shorter sampling durations. The issues of how much and how long to sample continue to be an issue for ergonomic exposure assessments since few tools explicitly state how to sample exposures. Full-shift measures were obtained for 17 workers over three work areas and 7 tasks. The full-shift measures were used to explore exposure variability, compare exposures in the three work areas assessed, and evaluate sampling strategy.

Study Schematic

The following schematic (Figure 1.1) provides a general overview of the three dissertation studies and the progression of the dissertation research over time.

FIGURE 1.1

Overview of the Dissertation Studies



Aims of the Dissertation Studies

The goal of this research was to:

1. Evaluate current exposure assessment tools and methodologies in the field of ergonomics.
2. Establish reliable and valid exposure assessment procedures for assessing awkward postures.

The specific aims of the dissertation studies were to:

- Investigate the <u>inter- and intra-rater reliability</u> of raters assessing upper limb postures of workers performing manufacturing tasks using a video-based observation technique (Studies I and II).
- Investigate the <u>inter-method reliability</u> of assessing upper limb postures of workers performing manufacturing tasks using a video-based observation technique and direct technical measurements (Study III).
- Investigate <u>sampling strategies</u> of upper limb postures of workers performing manufacturing tasks using a direct technical measurement tool (Study IV).

SIGNIFICANCE OF THE DISSERTATION STUDIES

Posture assessment using video observation is currently being utilized in several large prospective cohort studies throughout the United States and has widespread use in industry. Demonstration of reliability and validity of exposure assessment techniques is critical when investigating a causal relationship between postures and musculoskeletal outcomes, particularly when large prospective field studies utilize video observation as a means to quantify exposures to awkward postures. Results from these studies addressed reliability and validity concerns of observational methods and provided insight to improve observational exposure assessment techniques.

Based upon literature review, the dissertation studies addressed several areas either not previously researched or where little research exists. This included the investigation of rater reliability with multiple raters, analysis of non-cyclic tasks, comparison of inclinometry measures with video analysis measures, evaluation of inclinometry data using exposure variation analysis, obtainment of full-shift direct measures of posture, subsequent statistical analyses that evaluate individual postures, and evaluation of sampling duration strategies for posture assessment using inclinometry. This also included investigation of inter-method reliability of posture analysis for a more sophisticated analysis technique (computer-based program) not previously evaluated in the literature. In addition, these studies evaluated exposure variability in an attempt to enhance exposure assessment and sampling strategies in ergonomics research.

The dissertation studies addressed several specific goals as outlined in the National Occupational Research Agenda (NORA). This included the NORA 1st Decade priority areas of 'musculoskeletal disorders of the upper extremities' and the call for improved 'exposure assessment methods' and the NORA 2nd Decade 'Manufacturing Sector' and the 'Cross-Sector Programs' (exposure assessment and musculoskeletal disorders). The dissertation studies addressed r2p objectives by transferring and translating results to other research professionals in an effort to improve exposure assessment techniques and encourage standardization of methodologies when using observational methods and tools.

SCOPE OF THE DISSERTATION STUDIES

Studies I and II were part of a larger epidemiological prospective cohort study that investigated the relationship between hand-intensive work and upper extremity MSDs in a home appliance manufacturing facility. The prospective cohort study had two specific aims: (1) to determine the incidence of self-reported upper extremity MSDs among household appliance manufacturing workers over a three year period; and (2) to estimate the effects of hand and arm forceful exertions, frequency of repeated stereotypical motions, postures assumed while performing work, and occupational psychosocial stressors on future MSD outcomes. Each participant's work tasks were characterized by measurement of forceful exertions using surface electromyography, repetition using the Hand Activity Level (HAL) (ACGIH, 2003), and postural assessment using MVTA. Daily exposures for each participant were calculated from these data. To prevent exposure misclassification, new sets of measures were conducted for every new task assigned to participants during the study period. Participants also recorded upper extremity symptoms on a daily basis. Approximately 500 study participants at the facility were enrolled over several years.

Studies III and IV were performed as separate research projects in a different manufacturing environment from Studies I and II. While these studies were not part of a larger research project, they are part of a similar set of projects being performed at this particular facility.

SECTION TWO

REVIEW OF LITERATURE

ORGANIZATION OF THE LITERATURE REVIEW

The literature review begins with a discussion of the musculoskeletal disorders as they relate to manufacturing environments. This is followed by a discussion of ergonomic exposure assessment methods for evaluating posture, followed by a discussion of research in the ergonomics literature that has investigated rater reliability, inter-method reliability, and exposure variability. Lastly, a discussion of statistical analyses is provided.

MUSCULOSKELETAL DISORDERS AND MANUFACTURING

Musculoskeletal disorders (MSDs) of the upper extremities are painful and potentially disabling conditions that affect the hands, arms, shoulders, and neck. Over 14 million workers in the manufacturing sector are at risk for occupational injuries and illnesses. In 2006, the largest percentage of occupational injuries and illnesses were in the manufacturing sector, 20% and 36% respectively (BLS, 2007). More than half of these cases (55%) resulted in days away from work, job transfer, or restriction. The manufacturing sector accounted for 18% of all MSD cases and had the highest incident rate per 10,000 full-time workers for carpal tunnel syndrome involving days away from work (BLS, 2007). Injury and illness data for 2007 as reported by the BLS (2008) demonstrated similar trends as 2006, however showed an overall decrease in most categories. Examples of frequently reported upper extremity MSDs include carpal tunnel syndrome, tendonitis of the hand and wrist, epicondylitis, rotator cuff tendonitis, shoulder impingement, and neck strain. Previous research has generally identified associations between occupational risk factors (force, repetition, vibration, and awkward postures) and MSDs of the upper extremities (NIOSH, 1997 and NRC/IOM, 2003). Frequent or sustained work with elevated arms has shown an increased risk for neck/shoulder disorders (Ariens et al., 2000; NIOSH, 1997). Flexion and extension of the wrist has shown association with tenosynovitis of the tendons in the wrist and carpal tunnel syndrome (Tichauer, 1966; Armstrong and Chaffin, 1979; Armstrong et al., 1984). Ulnar and radial deviation has shown association with DeQuervains, a tendon disorder at the base of the thumb (Tichauer, 1966; Hoffman, 1981).

Low-back pain (LBP) is common in both the general and working populations. Back disorders are multifactorial in origin and may be associated with both occupational and non-occupational risk factors (NIOSH, 1997). Awkward postures of the back include non-neutral trunk postures related to bending (flexion) or twisting (rotation) in extreme positions or at extreme angles (NIOSH, 1997). Most studies focus on substantial changes from a neutral trunk position, while relating risk to both the speed of change and degree of deviation. In most studies, awkward trunk postures are measured concurrently with other risk factors for work-related back disorders including heavy physical work, lifting, heavy forces, and whole body vibration (NIOSH, 1997). Injury and illness data for 2007 as reported by the BLS (2008) documented 235,960 cases involving injuries to the back.

Ergonomics Exposure Assessment Methods for Evaluating Posture

Many ergonomic exposure assessment tools have been developed to assess exposure to awkward postures (Karhu et al., 1977; Keyserling, 1986; McAtamney and Corlett, 1993; Hignett and McAtamney, 1995). Exposure assessment tools have ranged from indirect methods that provide qualitative estimates to direct methods that quantify physiologic signals. Awkward postures have been assessed by analyzing the frequency of extreme joint motion, duration in a specific posture, and magnitude of joint angle.

Typical tools used to measure these variables include questionnaires, direct observation, video observation, electrogoniometers, inclinometers, accelerometers, and three-dimensional kinematic programs.

Indirect Measurement Tools

Indirect measurement tools designed to assess exposure to awkward postures include questionnaires, direct observation methods, and video observation systems. Due to their low cost and relatively quick analysis time, observational methods have been commonly used to assess awkward postures in occupational settings. Observational methods are typically easy to use, do not interfere with job processes, and do not require expensive equipment. However, few studies have identified the sources of error associated with estimating angular deviations (Genaidy et al., 1993). Observational tools range from simple checklists and diagrams to computer-based programs (Priel, 1974; Karhu et al., 1977; Keyserling, 1986; van der Beek et al., 1992; McAtamney and Corlett, 1993; Fransson-Hall et al., 1995; Wiktorin et al., 1995; Hignett and McAtamney, 1995; Ergonomics Analysis and Design Research Consortium, 2003).

Many observation-based assessment tools have been designed via a paper/pen coding format. These tools are typically inexpensive to use and postural assessments are made without disturbance to those being observed (Li and Buckle, 1999). Burdorf and van der Beek (1999) described the most significant disadvantage of these systems is recording procedures lacks precision, and therefore, the reliability of these tools has varied. The limitations of paper/pen methods have restricted them to relatively static jobs (Li and Buckle, 1999). Examples of pen/paper assessment tools include the Posturegram (Priel, 1974), Posture Targeting (Corlett et al., 1979), Rapid Upper Limb Assessment

(RULA) (McAtamney and Corlett, 1993), and Rapid Entire Body Assessment (REBA) (Hignett and McAtamney (1995).

Priel (1974) created an observational analysis tool called the Posturegram. This method utilized a form that first establishes the starting position of limbs on the body. One can then record where the joints are located with respect to the starting position on the form by using reference planes around the body such as x, y, and z (Priel, 1974). This method is more applicable for assessing single postures instead of dynamic movements (Li and Buckle, 1999). In 1979, Corlett et al. created Posture Targetting as a technique to record postures. This method uses diagrams that partition angular movements of specific parts of the body using concentric circles (Corlett, 1979). When a posture departs from neutral, one marks the specific body part involved in its applicable concentric circle. This procedure is best suited for observing static postures (Li and Buckle, 1999). Using test-retest correlations, researchers found that even after modest training, observers could obtain high consistency when recording static postures (Corlett et al, 1979).

McAtamney and Corlett (1993) created RULA to investigate work-related upper extremity disorders. This method uses diagrams of body postures and three scoring tables to aid in evaluation of risk factors such as repetition, static muscle work, force, posture, and the length of time without a break. Only paper and pen are necessary to complete this analysis. This method is more applicable for sedentary jobs (McAtamney and Corlett, 1993). Tests of reliability indicated high consistency (McAtamney and Corlett, 1993). Discrepancies in posture estimation occurred at the border between two posture categories (McAtamney and Corlett, 1993). The REBA was created by Hignett and McAtamney (1995). While based on RULA, this tool assesses dynamic or static postures. This tool divides the body into segments and codes them individually using a scoring system. The scores are combined into an overall REBA activity score that refers to an action level of urgency (Hignett and McAtamney, 1995).

Other indirect measures of posture have been developed through computer-aided observational systems. These methods usually record working postures on videotape, while analysis takes place at a later time. Typically, a time-sampling or simulated real-time method aids in assessment of the video (Li and Buckle, 1999). These computer-aided techniques include: Ovako Working Analysis System (OWAS) (Karhu et al, 1977), VIRA (Kilbom et al, 1986), Task Recording and Analysis on Computer (TRAC) (van der Beek et al, 1992), HAnds Relative to BOdy (HARBO) (Wiktorin et al, 1995), Portable Ergonomic Observation (PEO) (Fransson-Hall et al, 1995), Keyserling's observation technique (1986), and Multimedia Video Task Analysis (MVTA) (Ergonomics Analysis and Design Research Consortium, 2003). Multimedia Video Task Analysis (MVTA) is a relatively new computer-based tool to assist in the observational assessment of posture. MVTA is a video-based exposure assessment program that can be used to automate time and motion studies and ergonomic analyses from video. In regards to posture, MVTA allows for the evaluation of any pre-determined posture over a continuous period of time.

Advantages of video-based systems include the ability to observe posture in real-time, forward to backward motion, and slow motion, if necessary. Observer bias, or the possible effects of an observer's presence, is avoided since body movements can be recorded by camera (Li and Buckle, 1999). Slowing or stopping the video allows for observations that are more detailed. High portability, reasonably low equipment costs, and the generation of permanent records of job tasks are other advantages. Disadvantages can include long and detailed observer training as well as long lengthy analysis time if the parameters are extremely detailed (Li and Buckle, 1999). In addition, camera setups may be limited in their ability to capture dynamic jobs; therefore, observers are looking at postures from camera angles that may not be adequate for posture estimation. In some cases, the criteria for posture classification are not well defined (Li and Buckle, 1999). This means that raters have the flexibility to use their own interpretations of how to measure or assess a posture. Lack of well defined rules for posture classification can lead to rater inconsistency possibly resulting in decreased reliability.

Direct Measurement Tools

Direct measurement tools designed to measure exposure to awkward postures include: direct measurements using protractors, motion capture systems, electromagnetic sensor systems, ultrasound emitters, electrogoniometry, and inclinometry. Direct measurement systems vary in complexity and accuracy and currently no system exists that can measure all desired postures in a field setting. Of particular interest for the proposed study are the capabilities of inclinometry devices.

Inclinometers are instruments used for measuring angles of slope (tilt), inclination, or elevation of an object with respect to gravity. Sensing technologies commonly found in inclinometers includes accelerometers, liquid capacitive and micro-electro-mechanical systems. Of interest for the proposed study, is the operation of the accelerometer sensing technology. Accelerometers measure the acceleration they experience relative to freefall. These devices detect magnitude and direction of the acceleration as a vector quantity, which can be used to sense orientation, vibration, and shock. Inclinometers have been increasingly used to study postures experienced during work tasks (Hansson et al., 2001; Bernmark and Wiktorin, 2002; Hansson et al., 2006). Advantages of these tools include cost effectiveness, ease of use, (Li and Buckle, 1999) and most recently, adequate battery life to allow for multi-shift measurement. One limitation of inclinometers is the inability to measure rotation. Therefore, in body segments, particularly the upper extremities which rotate around the long axis simultaneously with other movements, data must be interpreted with caution (Bernmark and Wiktorin, 2002). In addition, these tools have been historically limited to clinical settings or research settings in a laboratory, not dynamic work situations where the device has to be constantly readjusted or interferes with the worker's task (Li and Buckle, 1999).

Inclinometers/Accelerometers

The orientation of a body segment has three degrees of freedom, meaning it can be characterized by rotation around three orthogonal axes (Hansson et al., 2001). If the data are sampled at a sufficiently high rate (Nyquist theorem), all kinematic information can be derived from the acquired data. If the line of gravity is used as a reference, two of the three degrees of freedom in orientation can be measured by inclinometers (Hansson et al., 2001). Inclinometers measure the spherical coordinates of acceleration (ρ) acting on the body segment of interest, inclination (Θ) in relation to the vertical line, and direction (φ) of the inclination (Bernmark and Wiktorin, 2002). Inclinometer has three accelerometers mounted perpendicular to each other in the x,y,z directions.

Accelerometers with a direct current (DC) response are capable of measuring static acceleration and can be used for detecting orientation relative to the line of gravity (Hansson et al., 2001). The angle is measured relative to the orientation of the acceleration vector, which, in the presence of dynamic acceleration, can deviate from the line of gravity. Thus, during movements that are not constant in speed and direction, an angular error may be introduced. For recordings during these conditions, the magnitude of acceleration will deviate from 1G and the magnitude of the dynamic accelerations can be calculated as the difference between this magnitude and 1G (gravity) force. Thus, the max possible angular error can be estimated. The fundamental limitation of inclinometry, the rotation around the line of gravity cannot be assessed, in combination with the true 3D orientation of the body segments has to be considered when the data are interpreted. Regarding the upper arm, flexion/extension cannot be separated from abduction/adduction. Positioning of the transducer close to the shoulder is preferable, as this reduces errors caused by centrifugal acceleration (Hansson et al., 2001).

Forsman et al. (2002) utilized inclinometers to record angles, relative to gravity, of the head, upper back, and upper arms for two tasks in an automotive facility. To obtain head and upper arm measurements, one inclinometer was placed on the forehead and another was fixed to plastic a plate that was placed along the upper arm, with the lateral edge along the line from the lateral-posterior corner of the acromion to the lateral epicondyle, and the upper edge at the insertion of the deltoid muscle. The reference position for the head (0° flexion) was defined as the position obtained when the subject was standing upright, looking at a mark at eye level. The reference position for the upper arm (0° elevation) was recorded with the subject sitting, the arm hanging perpendicular with a 2 kg weight in hand. Samples were obtained at 20 Hz and the authors calculated 10^{th}, 50^{th}, and 90^{th} percentiles of the angle for the inclinometer. The method of zeroing the inclinometer and the use of reference positions as described in Forsman et al. (2002) has been utilized in other studies using tri-axial accelerometers as inclinometers (Akesson et al., 1997; Bernmark and Wiktorin, 2002). Bernmark and Wiktorin (2002) compared inclinometer readings with a MacReflex motion capture system. Six subjects performed simulated tasks while the inclinometer (20 Hz sampling rate) and motion capture system simultaneously recorded arm elevation. To measure arm elevation, the inclinometer was mounted caudal to the insertion of the deltoid muscle. The reference position, inclination in relation to the vertical line, for the upper arm was defined as the position obtained when the subject was sitting with a slight bend to the right with 2 kg weight in hand. Direction was defined by having the subject stand and hold their arm at 90° abduction. Results of the inclinometer were compared to the motion capture system using the Pearson product moment correlation coefficient as a linear relationship could be assumed. Arm elevation without influence of dynamic acceleration had almost perfect correlation. Even with dynamic acceleration, correlation was still good. When the arm changed direction, elevation as estimated by the inclinometer was higher than motion capture system. The authors attributed this discrepancy to marker placement of the motion capture system (Bernmark and Wiktorin, 2002).

RELIABILITY OF POSTURAL OBSERVATION IN ERGONOMICS RESEARCH

When measuring posture in a research study, the measured value consists of the true value plus some amount of measurement error. This measurement error is either a systematic error or random error. Systematic and random error affects the reliability of a measurement method. Random error fluctuates randomly possibly leading to an overestimation or underestimation of the measured value (CDC, 2001). Reliability represents the degree to which the measured values for a certain concept are consistent

(CDC, 2001). Deshon (2002) defined reliability as an assessment of the amount of error variance present in an observation. Intra-rater reliability estimates the consistency of a single rater's judgments over time or different occasions (Chen, 2004). Inter-rater reliability estimates the consistency of judgments among different raters. Inconsistent judgments threaten the validity of any conclusions made from those judgments.

Too often, exposure assessment tools have been utilized in studies to demonstrate associations or attempt to define causal relationships between risk factor and musculoskeletal outcome without any consideration for reliability and validity (Burt and Punnett, 1999). Field studies have been limited in number due to the difficulties with collecting quality postural data outside the laboratory (Keyserling, 1986). An important step in developing improved exposure assessment tools for MSDs is the evaluation of intra- and inter-rater reliability. When reporting research results, the need to assess reliability of the measurements is recognized and measures of reliability are expected to be reported. The lack of prospective studies and validated exposure assessment tools has made it difficult to determine a causal relationship between risk factor and outcome (Burt and Punnett, 1999). Epidemiological studies that attempt to determine this causal relationship without the demonstration of reliability and validity of the exposure assessment tool used to predict musculoskeletal outcome may lead to inaccurate conclusions between exposure and disease outcome (Burt and Punnett, 1999).

The following discussion summarizes major findings of reliability investigations for observation-based ergonomic exposure assessment methods used to evaluate posture. In regards to pen/paper based assessment tools, researchers have found that even after modest training, observers could obtain high consistency when recording static postures

(Corlett et al., 1979). Tests of reliability for Posture Targeting, RULA, and REBA have found high consistency among raters. Discrepancies in posture estimation commonly occurred at the border between two posture categories (McAtamney and Corlett, 1993). While several video-based observation methods have been evaluated in regards to reliability, few studies have assessed the reliability of the more recent sophisticated computer-based assessment systems. Most studies of reliability have evaluated direct observation or simple video techniques (Douwes and Dul, 1991; van der Beek et al., 1992; de Bruijn et al., 1998; Burt and Punnett, 1999; Pan et al., 1999; Ketola et al., 2001).

INTER-RATER RELIABILITY STUDIES

Using a method similar to OWAS, Keyserling (1986) found reliability to be strongest for shoulder flexion/abduction greater than 90˚. Low reliability was attributed to the analyst's inability to define precise boundaries between different postures of the trunk and shoulder (Keyserling, 1986). Keyserling (1986) reported that shoulder reliability was highest during extreme flexion/abduction events and concluded that reliability should increase with adequate training and improved decision criteria. These findings are similar to those of various studies of this nature (Stetson et al., 1991; Burt and Punnett, 1999; Lowe, 2004a). Like Keyserling (1986), Ketola et al. (2001) evaluated elevation of the upper arm for postures greater than 90˚. Ketola et al. (2001) reported good to moderate reliability based on percent agreement of 0.47 (left arm) and 0.71 (right arm) and kappa coefficients of 0.39 (left arm) and 0.68 (right arm). Ketola et al. (2001) reported relatively low kappa coefficients for the left and right wrists as 0.34 and 0.41, respectively.

Decreased reliability due to smaller joint movements has been discussed as a contributing factor in other studies of this nature (Keyserling, 1986; Stetson et al., 1991; Burt and Punnett, 1999). Stetson et al. (1991) estimated inter-rater reliability of two analysts observing various degrees of wrist flexion and extension. Based on a multiple regression analysis, Stetson et al. (1991) concluded that differences between the two raters were attributable to poor video clarity and difficulty determining angular values of postures. Stetson et al. (1991) also reported that inter-rater reliability was higher for extreme wrist deviations and lower for smaller deviations. In the present study, it was not completely clear if discrepancies in wrist posture ratings were truly attributable to the wrist's inherent smaller deviations or if other related factors account for the variance. Burt and Punnett (1999) found the highest percent agreement in wrist flexion greater than 30°. Burt and Punnett (1999) concluded that unclear definitions of postures, inadequate rater training, and difficulty observing slight body movements compared to gross body movements were possible variables that accounted for disagreement between raters. Burt and Punnett (1999) also concluded that precise estimates of joint deviation in degrees of excursion from neutral postures were more difficult than estimating postures using anatomical referencing, similar to research performed by Wiktorin et al. (1995) using the HARBO (HAnds Relative to the Body) system. The HARBO study reported high inter-rater reliability in regards to positions of the hands relative to the body, with intraclass correlation coefficients >0.90 and a Pearson product moment correlation coefficient of 0.97.

Several other studies have reported similar reliability coefficients and findings as discussed above. In a study by Pan et al. (1999), researchers evaluated inter-rater reliability of arm/shoulder postures using the PATH method. A kappa coefficient of 0.50 averaged over the four arm posture categories was reported. Lowe (2004b) reported intraclass correlation coefficients (ICCs) for wrist flexion and extension of 0.20 and 0.39, respectively, for a six-category wrist posture classification. Low reliability may be attributed to the combination of all wrist postures in the Lowe (2004b) study. Lowe (2004b) concluded that as the complexity of the observations increases based on the number and width of posture categories, the agreement between observers decreases. Stevens et al. (2004) reported an ICC of 0.66 for hand/wrist posture using the Strain Index (Moore and Garg, 1995). All postures were analyzed simultaneously for the reliability analysis in the Stevens et al. (2004) study, therefore making it difficult to determine the respective contribution of each posture when calculating the ICC.

A recent study (Bao et al., 2009) of inter-rater reliability of posture observations evaluated the reliability of a posture observation method that was used in a large epidemiological field study of upper extremity musculoskeletal disorders. Seven experienced raters observed posture of the upper extremities of four different jobs from video. Four jobs that differed greatly in the variations of work postures were used in the study: laundry handler, lumber handler, electronics assembler, and a pharmacist. Jobs were video-recorded from two angles for a duration of 15 minutes. Posture angles were estimated from 37-38 randomly selected frames (30 frames per second frame rate) in each of the four video clips by each of the seven raters. Posture angles of six different body parts (elbow, forearm, upper arm, neck, trunk, and wrist) for various movements were estimated for each frame. Raters estimated the angles by clicking on an area of a posture diagram displayed on a computer screen. The raters could also choose a 'missing' data option if the rater considered the posture angle not visible.

For analysis, the posture data were grouped into categories via three different strategies: 10° intervals, 30° intervals, and a pre-defined categorization strategy that evaluated specific ranges of postures (Bao et al., 2009). The mean estimate of the 7 raters for each of the posture parameters was computed for each frame. From the means per frame, descriptive statistics were calculated by job. Percentage of time spent in the pre-defined categories were calculated and summarized as neutral and non-neutral. The ANOVA and ICC's were used to investigate inter-rater reliability. The ANOVA showed no significant differences between the raters in terms of the mean posture angles for 12 of the 20 different pre-defined posture categories. A significant difference was found for trunk flexion, upper arm flexion and extension, and left wrist flexion and extension. Intraclass correlation coefficients varied with the lowest ICC's see among left and right wrist ulnar/radial deviation, neck lateral flexion, and trunk twisting (ICC's $\leq$ 0.20). In general, posture parameters had better precision for neutral posture categories than non-neutral posture categories (Bao et al., 2009).

The larger posture intervals (30°) had better agreement than the 10° intervals, causing the authors to conclude that posture evaluation based on observation may not be able to use posture categories with a width less than 30° due to decreased reliability. The authors (Bao et al., 2009) also concluded that larger body parts were easier to observe and resulted in better reliability. Reliability was also affected by variability of postures, camera positions, video quality, and complicated work postures. The authors (Baoeet al.,

2009) recommended that targeted training aimed at avoiding common errors with an improved data entry system should improve inter-rater reliability.

INTRA-RATER RELIABILITY STUDIES

Using a method similar to OWAS, Keyserling (1986) utilized a trained analyst who viewed video-recorded jobs on different occasions separated by two months. Keyserling (1986) attributed differences in shoulder posture estimations across occasions to positions of the shoulder when it was near the intersection of two postures (Keyserling, 1986). For example, if one is to estimate shoulder flexion greater than or less than 45°, it is easier to rate a posture at 60° than a posture that is 47°, or near the cut-off point. There was an inability of the analyst to consistently use the same boundary when rating adjacent postures. de Bruijn et al. (1998) reported good intra-rater reliability and found percent agreement for head postures using the OWAS system of 88% and a kappa coefficient of 0.68. de Bruijn et al. (1998) utilized photographic slides that were observed for three seconds by the analysts. de Bruijn et al. (1998) also reported kappa coefficients above 0.80 for a combination of various shoulder postures. Douwes and Dul (1991) also evaluated the intra-rater reliability of posture observation using the OWAS system. Correlation coefficients were 0.97 or higher. The results of the de Bruijn et al. (1998) and Douwes and Dul (1991) OWAS studies were similar.

INTER-METHOD RELIABILITY OF EXPOSURE ASSESSMENT METHODS IN ERGONOMICS RESEARCH

Reliability does not imply validity. A reliable measure is one that measures something consistently, but not necessarily what it is supposed to be measuring. Validity is a broad term that encompasses several different forms. In regards to observational-based posture analysis measurement, internal and external validity should be demonstrated (Kilbom, 1994). Internal validity, the ability to measure what is intended, of observation methods has been estimated with reference to motion capture systems, electrogoniometers, and inclinometers (Kilbom, 1994). The majority of studies have focused on back postures with limited information on the upper extremities (Baty et al., 1986; Burdorf et al., 1992; de Looze et al., 1994; Keyserling, 1986). A few studies have validated neck postures with motion capture systems or observations made from still pictures as references (Leskinen et al., 1997; Fransson-Hall et al., 1995).

For observation of posture, external validity is defined as the analysis method's ability to distinguish physical exposure levels of posture associated with an increased risk in musculoskeletal disorders for a given anatomical area (Lowe, 2004a). Examination of external validity requires an epidemiological investigation, whereas the examination of internal validity requires a kinesiological/biomechanical investigation. Most studies that have investigated the internal validity of observation of posture have been conducted in a laboratory setting.

A number of published validity studies have utilized electrogoniometers or inclinometers as the reference comparison method for observation of neck, shoulder, and wrist postures. However, one can argue that since these devices are not considered a 'true reference' method, comparisons of observation measures to those obtained through electrogoniometry or inclinometry should be considered inter-method reliability studies. This is primarily due to the significant levels of measurement error found in these devices (Spielholz et al., 2001). Inter-method reliability measures the ability of 'different instruments which measure the same underlying exposure to yield similar results on the same subjects' (Armstrong et al., 1994). The following discussion summarizes major findings of inter-method reliability investigations for observation-based ergonomic exposure assessment methods used to evaluate posture. Other studies that have investigated comparison of methods did not utilize video analysis techniques. These studies typically were comparing self-assessed exposures or exposure estimations obtained from questionnaires to direct technical measures of posture or physiological metrics (Balogh et al., 2004; Forsman et al., 2002; Hansson et al., 2001).

Juul-Kristensen et al. (2001) compared postures of repetitive work in poultry processing using PRIM (Project on Research and Intervention in Monotonous Work) and direct technical measurements (inclinometers and electrogoniometers). Researchers video-taped 21 women for approximately 20 minutes each. Paired t-tests were used to estimate the significance of the difference between the dominant and non-dominant arm and hand. Non-paired t-tests were utilized to estimate the significance of the differences between tasks. Duration for arm positions recorded in the range of 30°-60° were 3-11% for observation and 39-45% for the inclinometer. The smallest difference of 0.5-1% was found for > 60° arm elevation. The authors attributed differences in the two methods to different definitions for the reference position and how the postures were assessed. A lack of precise definitions of neutral postures of the cervical spine have previously been reported as one of the explanations for the large difference between duration of neck flexion recorded in observation methods versus direct technical measurements (Leskinen et al., 1997; Fransson-Hall et al., 1995).

Spielholz et al. (2001) evaluated self-report, observation, and direct measurement methods for three nursery tasks. Each task was performed for 30 minutes with cycles ranging from 10 seconds to 15 minutes. Video footage was analyzed via a VHS cassette and coding sheet by one analyst. Direct measures were obtained by a Noraxon 8-channel telemetric system and two electrogoniometers. The authors evaluated the percent of the work cycle in which the wrist was >30° flexion/extension, >10° radial deviation, and >15° ulnar deviation. Wrist flexion > 30° was selected as extreme posture based on increases in intra-carpal tunnel pressure (Stetson et al., 1991). The other extreme postures were selected based on previously developed tools (Armstrong et al., 1982; McAtamney and Corlett, 1993) or using half the range of motion for that specific movement. The authors selected 30° extension to maintain consistency with flexion. The authors used Pearson product moment correlation coefficients computed agreement measures for location (difference between means) and scale (differences in variability). Video observation and direct measurements had the highest correlation for extreme posture duration.

Lowe (2004b) expanded his original validity research (2004a) of the shoulder and elbow to the wrist/forearm. Lowe (2004b) utilized tri-axial electrogoniometers (30 Hz) and a hand-held miniDV camera (30 frames/second) to record postures experienced by the wrist/forearm during four simulated repetitive tasks. Five to twelve cycles (cycle times 8-56 seconds) of each task were video-taped synchronously with the electrogoniometers. Two categorical methods were used to scale wrist/forearm posture. Angular boundaries (>20° flexion/extension, >40° pronation/supination, and >10° radial/ulnar deviation) were consistent with other studies (McAtamney and Corlett, 1993; Jull-Kristensen, 2001; Ketola, 2001; Spielholz, 2001). Accuracy was calculated as the difference between the analyst's estimations of the variable and the measured variable as averaged over all of the individual work cycles. Lowe (2004b) recommended that future studies should evaluate the accuracy of posture analysis for more sophisticated observation techniques, such as the computer-based systems.

Burdorf (1995) performed inter-method reliability studies to estimate the systematic bias (accuracy) and random measurement error (precision) of various methods to assess postural load on the back. Burdorf (1995) explains that often a perfect measurement instrument is not available or it is infeasible so that alternative methods are used to ascertain validity of a measurement technique. Reproducibility studies are then performed to obtain indirect information on error distributions. In these types of studies, two or more separate assessments of exposure are performed on the same individuals by different instruments. The inter-method reliability coefficient of continuous variables is typically estimated by the Pearson correlation coefficient whereas the Spearman rank correlation coefficient is used for categorical variables (Burdorf, 1995). Based on previous inter-method reliability research, Burdorf (1995) found that postural load due to trunk posture is best assessed by direct observation or inclinometer measurements. Comparisons of direct observation and inclinometer measurements have shown reliability coefficients of 0.60 for the amount of time recorded in trunk flexion greater than 20º (Baty et al., 1986 and Burdorf et al., 1992).

Burdorf et al. (1992) evaluated an observation method and continuous measurement technique used to record bending of the trunk. Fourteen sedentary workers and 16 workers performing dynamic lifting tasks in two facilities were used in the study. Observations were made using OWAS while each worker performed their routine tasks during a 60 minute period. Observations focused on forward and backward trunk motion greater than 20°. The angle of inclination was defined as the angle between the straight line through the pelvis and shoulders and the vertical (Burdorf et al., 1992). Observations were made at the workplace every 20 seconds resulting in 180 observations per worker. Direct measures were obtained using the Portable Posture Registration Set, which consisted of an inclinometer, pendulum potentiometer, and portable recorder. The position of the trunk was measured by placing the inclinometer at the L2-L3 level of the spine.

The average percent time spent with the trunk in a bent position (>20°) was calculated for both the observation and direct measurement techniques. This was further categorized into a low and high category of time spent in a bent posture with 30% used as the cutoff point. Agreement was estimated using the kappa coefficient and Spearman rank correlation coefficient. Spearman correlation coefficients were reported as 0.62 for the sedentary workers and 0.57 for the dynamic workers. The kappa statistic was estimated at 0.43 for all workers, 0.36 for sedentary workers, and 0.51 for the dynamic workers. For the dynamic workers, the mean proportion of time with bent trunk posture was the same for observation and direct measurement. However, observation gave a significantly higher estimate of the time spent with bent trunk posture than the direct measures for the sedentary workers. While significant correlations were obtained for the summary data, large differences were found between data for individual subjects. The authors (Burdorf et al., 1992) attributed the differences between observation and direct measures to several factors: lack of precision if movements are concentrated at the border of the desired posture and different definitions of the angles of the trunk bending existed.

The occurrence of systematic bias in many of the referenced studies is problematic, since it may lead to an overestimation or underestimation of the risk per unit of exposure (Burdorf, 1995). Systematic bias can result from a dependency on the relative bias of the exposure magnitudes assessed and at a discrepancy in the definition of the angular position of the trunk (Baty et al., 1986 and Burdorf et al., 1992). Intra-method reliability studies focusing on changes in postural load over time are few (Burdorf, 1995). The author (Burdorf, 1995) emphasized that a core element for optimizing measurement strategies is an unbiased estimate of the intra-method reliability coefficient or variance ratio. The reliability coefficient depends on the exposure distributions among the analyzed workers or tasks and this dependability hampers application from one population to another (Burdorf, 1995).

EXPOSURE VARIATION AND MEASUREMENT STRATEGY IN ERGONOMICS RESEARCH

Exposure Variation

In epidemiologic research, variation in exposure has two fundamental dimensions: person and time (Loomis and Kromhout, 2004). While the concept that exposure varies across groups is fundamental to epidemiologic research, exposure variation within groups is also important. This within-group variability has implications, including potential for exposure measurement error and misclassification (Loomis and Kromhout, 2004). The evaluation of the dimensions of and determinants of exposure variability can be used to plan exposure measurements, assign estimates of exposure, or predict and control future exposures. While a considerable number of studies have evaluated physical workplace exposures, most studies have only evaluated short periods of exposure as a representation of total exposure. While much research has been devoted to improvements of direct technical measurements in regards to reliability and validity, few studies have investigated within and between subject variability for ergonomics exposure assessment tools (Ortiz et al., 1997; Allread et al., 2000; Anton et al., 2003; Mathiassen et al., 2003; Dahlberg et al., 2004; Moller et al., 2004; Svendsen et al., 2005; Hansson et al., 2006).

Within and between worker variability is critical when assessing the statistical precision of exposure estimates and the power of studies assessing exposures, particularly when trying to associate exposures with a health outcome (Mathiassen et al., 2003). In addition, exposure variability provides information regarding the variation of work tasks. Some research has recommended more variation for repetitive work over prolonged periods of time (Kilbom, 1994). To enhance the discussion of exposure variation, some additional measurement strategy research from the occupational hygiene field has been added to this section.

Mathiassen et al. (2003) evaluated within and between subjects exposure variability for workers performing a strictly controlled task with a hand-held tool. Muscle activity in the shoulders and lower arms was quantified via electromyography. Head and upper arm inclination was quantified using triaxial accelerometers. Seven subjects performed three simulated tasks using two different tools. Each task was repeated immediately after the first performance. For each task, the median values of inclinations as obtained by the accelerometers and the normalized electromyography amplitudes were obtained. Seven exposure parameters (task, tool, subject, all interactions, and residual) were evaluated using the ANOVA and subsequent coefficient of variation estimates to evaluate within and between subject variability. Mathiassen et al. (2003) were unable to evaluate variability between days since all measurements were obtained in the same day. Variability between subjects was large across all combinations of tasks and tools. Exposure variability between subjects was larger for electromyography amplitudes on the left side of the body, indicating that more measurements would be needed to achieve a precise group mean and to secure acceptable power. For the posture data, the coefficient of variation values were smaller than those found in previous research, 0.15-0.22 as compared to 0.5-0.7 found by Aaras et al. (1988). Mathiassen et al. (2003) concluded that inclination differences could reach significance in studies of a feasible size, ranging from 6-15 samples, at 95% confidence and β=0.80.

Additional studies that have evaluated within and between worker variability to the extent of the Mathiassen et al. (2003) study are few, particularly those that have evaluated between-days variability. Allread et al. (2000) determined the amount of data needed to ensure sufficient accuracy in estimating mean trunk motions by employing within and between worker variability analyses. This study (Allread et al., 2000) found that the majority of variability in mean trunk motion was due to the design of work tasks and variations due to repeated cycles of a task or to employees were minor. Hansson et al. (2006) evaluated six women who performed three simulated assembly tasks, all repeated on three different days. Inclinometry was used to evaluate postures of the head, upper back, and upper arms. Within subjects variability was found to be 3.4˚ and between-subjects at 4.0˚. Variability was found to be dependent on the percentage of time spent in a particular posture category.

Hansson et al. (2006) evaluated six women who performed three standardized assembly tasks in a laboratory setting, all repeated on three different days separated by at least seven days. Tri-axial inclinometers were applied to the head, upper back, and upper arms. When inclinometers are used, variability is inherently introduced due to the imperfect reproducibility of the reference positions. For a particular subject, there will be between-days variability. Also, different individuals will not perform the same task in the same manner, thus causing between-subjects variability. Cycle times for the tasks performed ranged from 24 to 58 seconds with a total task time of 20 minutes (Hansson et al., 2006).

Between-days (within subject) and between-subjects (within tasks) variance components were calculated for the proportion of time spent in posture categories of the head, upper back, neck, and upper arms. For the head, upper back, and neck, three categories (<-15º, >15º, >45º) were used for flexion/extension, and one (<-15º or >15º) or lateral flexion. For upper arm elevation, categories >30º and >60º were used. For proportion of time spent in specific posture categories, the variability depended on the percentage of time spent in the particular posture category. Variance components were derived using a restricted maximum likelihood algorithm in a linear random effects model. The relative variability was large with an average of 103% for between days and 56% for between subjects. While the three tasks differed in postures, the variability was only to a minor extent influenced by the tasks performed (Hansson et al., 2006). Regarding the choice of posture categories, the authors argued that results can be generalized to arbitrary sector limits since the variation is primarily related to the fraction of time spent in a particular category (Hansson et al., 2006).

Moller et al. (2004) attempted to quantify exposure similarity of working postures within and between individuals performing electronics assembly work, using simple variance measures. Five subjects performed six different tasks for one hour each over different working days. Inclination of the head and arms was monitored using triaxial accelerometers. Exposure level and frequency were evaluated using exposure variation analysis (EVA). For the EVA, postures were distributed into 10º increments starting at 0º and ending at 180º. For postures, the exposure parameters was defined as the percentage of time spent at an inclination (head) or elevation (arms) angle larger than 30º. The parameter describing frequency was defined as the percentage of cycle time spent in sequences shorter than one second within the same exposure category. ANOVAs were performed for each task and for all tasks combined. An intraclass correlation coefficient was used to express the relative sizes of within- and between-subject-days variance components.

Between-cycle variances were similar between the tasks. The variance between-cycles within days and subjects was consistently smaller than the variance between subject-days for most posture levels. The variance between-cycles within days and subjects was, in general, much larger consistently than the variance between subject-days for arm posture frequency, while similar for head posture frequency. The authors (Moller et al., 2004) concluded that their findings were similar to previous studies evaluating monotonous and repetitive tasks. They recommended that data from less controlled and more varied tasks and jobs are needed to fully appreciate and understand exposure variation.

Anton et al. (2003) assessed exposure to physical risk factors during variable non-cyclic work using electromyography. The authors argued that while direct measurement devices are able to quantify exposure more accurately and precisely, they generate considerable amounts of data. There are few efficient methods of reducing these data that are understandable by researchers and non-researchers (Anton et al., 2003). Exposure variation analysis (EVA) is one data reduction method that has been used for electromyography. When used for electromyography data reduction, EVA describes the intensity of muscular activity used during a period of work, as well as the duration at each intensity level. This study evaluated electromyographic activity of the finger flexors for 48 construction workers over a one hour work sample.

The data were analyzed using EVA. Eight intensity levels and seven duration levels were used. Clustered exposure variation analysis (CEVA), a modification of EVA, was used to further reduce the data. The intensity and duration levels were combined to create three intensity levels and two duration levels. Univariate analyses were conducted for the CEVA exposure category variables. A two-way mixed effects repeated measures ANOVA was used to evaluate the percentage of sampled work time in each CEVA category. The authors (Anton et al., 2003) concluded that EVA and more particularly CEVA, seem to be useful methods for contrasting non-cyclic work, typical of understudied industries. The authors (Anton et al., 2003) then further explained the differences between the EVA and CEVA. Analysis of EVA categories does not typically involve the intensity-duration interaction as an important variable. CEVA allows for this interaction. CEVA also differs from EVA in that it is a summary measure that is not intended to describe the entire exposure pattern. In addition, not all data are typically used to create the CEVA categories. This allows greater contrast between the groups. A multivariate ANOVA is typically used to evaluate EVA data, while a mixed repeated measures ANOVA is used for CEVA. The authors (Anton et al., 2003) concluded that

CEVA is a statistical method that easily evaluates and accurately quantifies exposure to forceful exertion and shows promise as an assessment method for other physical risk factors, such as awkward postures.

Mathiassen (2006) took a different approach to investigating exposure variability and proposed a framework for investigating and evaluating aspects of exposure variation based on explicit definitions of variation as “the change in exposure across time” and diversity as “the extent that exposure entities differ”. On the basis of literature review, Mathiassen (2006) argues the validity of the conviction that more “variation” is an effective remedy for improved musculoskeletal health. The author explains that variation focuses on an exposure time-line in terms of how much and how fast exposure changes and whether it exhibits patterns of similarity or regularly occurring events.

Mathiassen (2006) explains that while variation refers to features of an individual’s exposure over time, it does not consider exposure similarities or differences between tasks, jobs, occupations, or within days or cycles. These exposure events are better explained by the term diversity. Mathiassen (2006) provides several examples. Two tasks may be diverse, meaning they differ in mean exposure at the group level. Two work cycles performed by an individual may be similar with respect to cycle time, but may be diverse in terms of the proportion of muscular rest. Another term that Mathiassen (2006) describes is exposure variability. He suggests that this term is restricted to quantitative measures of dispersion with a basis in descriptive statistics such as mean, standard deviation, and error. Exposure variability is used generally to describe dispersion within days within subject, between days within subject, between subjects, between tasks, and between jobs. Mathiassen describes that variation has been measured and evaluated using counting, estimating frequency above or below certain thresholds, the amplitude probability distribution function (APDF), and the exposure variation analysis (EVA).

Peretz et al. (1997) evaluated the variability of exposure over time to lead, benzene, and dust within and between workers. The aims of the study were to estimate the magnitude of exposure variability over time including the variances between and within-worker, to explore the causes for between-worker variance, and to model variables affecting the within-worker variance. Fifty-four workers in six factories were recruited. Ten hygiene surveys performed at random intervals of three to seven weeks were completed at each facility using the 54 workers. Each worker had a minimum of six measurements. The ANOVA was used to estimate variance components in a three-way nested random effects model for the effects of air contaminant, factory nested within air contaminant, worker nested within factory and air contaminant, and repetition nested within worker, factor, and air contaminant.

Results demonstrated differences in median exposures among the workers within each factory (Peretz et al., 1997). Analysis of the variance components showed a variance of 51% for within-worker and 49% due to variances between workers, factories, and air contaminants. Further investigation into the causes of variance between workers found that the mobility of the exposure source accounted for the most variance followed by the factory and the environment. Based on the high exposure variance for both within- and between-worker, the authors concluded that single measure of hygiene exposure is insufficient and recommended repeated measures over a year with intervals of several weeks (Peretz et al., 1997).

In summary, few studies have provided a comprehensive evaluation of within and between worker variability for postural exposures. Data on within-day variability for postures during repeated tasks are very rare in the ergonomics literature. Only one study was found that evaluated full-shift exposures (Svendsen et al., 2005). In addition, most studies were executed by simulating tasks in a laboratory setting, not in real-world work environments.

Measurement Strategy

While various ergonomic measurement methods and tools exist, few of those methods and tools explicitly state when and how to sample exposures (Trask et al., 2008). Much ergonomic research focuses on capturing short segments of exposure information, due to time, cost, and measurement restrictions, and extrapolating this information to predict a full shift's exposure. Continuous full-shift exposure measurements remain to be uncommon. Two studies are discussed in this section. The first study by Trask et al. (2008) is most applicable to this dissertation research, while the Mathiassen et al. (2003) study provides supplemental information to measurement strategies.

Trask et al. (2008) compared several low back electromyography exposure metrics measured over an 8-hr work shift with the same metrics sampled over shorter durations in order to investigate adequate sampling durations. Full shift electromyography measurements were made for 35 workers with each worker sampled for two shifts. Summary statistics were calculated for each individual's work shift exposure data. Mean and 50th percentiles were used as measures of central tendency, while standard deviation and percentiles from amplitude probability distribution functions were used to represent the range of exposure. Comparison of different sampling durations was performed using the full-shift measures. This was accomplished by *a posterior* resampling within single, full work shifts. Data were resampled at the following shorter durations: 4-hr, 2-hr, 1-hr, 10-min, and 2-min. A randomly selected start time was used for each duration. This resulted in six sets of data for each worker which were then used in the summary statistical analyses. Pearson's correlation coefficients were used to determine the strength of the relationship between the full-shift and resampled shorter durations. Repeated measures ANOVA was used to test the significance between the full-shift and resampled shorter durations. Error and percent difference were used to determine the level of deviation from the full-shift data to the resampled shorter durations. Agreement between the full-shift data and resampled shorter durations were also assessed by calculating bias and limits of agreement.

The authors (Trask et al., 2008) found that the shorter sampling durations tended to overestimate the full-shift exposure for percentiles below the median and underestimate the full-shift exposure for percentile above the median. The ANOVA demonstrated significant differences between the full-shift and 2-min measures. Full-shift measures were significantly correlated (0.344-0.969) to the shorter sampling durations, except for the 2-min measure. Correlations were highest for the 4-hr and 2-hr measures and decreased with shorter sampling durations. The authors (Trask et al., 2008) concluded that full-shift electromyography sampling is feasible, but that shorter duration measurements are attractive because a greater number of assessments could be made during the same day. The 4-hr sampling duration most closely matched the full-shift measures. The authors (Trask et al., 2008) indicated that this was the first study of its kind to compare exposure metrics estimated at different sampling durations for electromyography. The issue of measuring exposure within a single day was also discussed and the authors explained that while obtaining a full-shift measure of exposure is better than shorter durations, it is not a perfect measure of typical exposure. The authors (Trask et al., 2008) concluded that sampling at 2-hr and 4-hr durations provides reasonable estimates of full-shift exposures, but that sampling at durations of 1-hr or less may produce large errors in exposure metrics, particularly for peak exposures.

Mathiassen et al. (2003) compared the efficiency of eight, one day exposure assessment strategies to determine the relationship between the number of data collected per subject and the precision of mean exposure estimates. The authors (Mathiassen et al., 2003) argued that epidemiologic studies rarely operate at the task level and that evaluating exposure by job disregards the fact that individuals with the same job title may have very different exposures due to different tasks.

Full-shift electromyography was performed on the right upper trapezius muscle of 24 cleaners and 23 office workers. The full-shift data were processed into one minute quanta to obtain gap time and jerk time. Gap time and jerk time are parameters that represent the level and frequency dimensions of muscle activation. On-site observations were made to document eight task categories of the jobs. Eight different sampling strategies were applied to the full-shift exposure data that had been split into one quanta. Strategies varied from consecutive sampling, to fixed-interval sampling, to random sampling while regarding or disregarding the task activity. Results indicated that a fixed-interval sampling strategy that disregarded tasks doubled efficiency as compared with random sampling. Proportions of gap time and jerk time clearly differed at the job and task levels of both the cleaners and office workers.

One-day-only (or less) sampling methods to assess exposure to ergonomic risk factors are common in both ergonomic intervention studies and epidemiologic investigations (Mathiassen et al., 2003). Typically, the goal is to capture an exposure estimate that is representative of a longer period of time. Mathiassen et al. (2003) focused on the within-day component of the variance of the mean exposure estimate. Based on previous research, the authors (Mathiassen et al., 2003) presumed that the between-days variance would be small thus allowing the within-in day variance to be a good approximation of the uncertainty of the mean exposure across days. The authors found that fixed-interval sampling was superior in efficiency as compared to random sampling. Consecutive and random sampling were equally efficient, only if the data were not autocorrelated. Autocorrelation describes correlation between values of a process at different points in time, as a function of the two times or the time difference. Autocorrelation occurs when residual error terms from observations of the same variable at different times are correlated or related. Consecutive sampling can lead to biased exposure estimates if those exposures change systematically over a shift. Factors such as fatigue or change in work pace could contribute to these biased exposure estimates (Mathiassen et al., 2003).

The authors (Mathiassen et al., 2003) concluded that while many ergonomic studies have applied task-based exposure assessment, few have investigated the classification of tasks. When classifying tasks, a reliable procedure for obtaining the amount of time spent performing the specific tasks is important for the success of a task-based exposure assessment. Exposure assessment estimates, whether they are made randomly, at intervals, or consecutively, can become uncertain if errors in determining task proportions exist. The authors (Mathiassen et al., 2003) also stressed the importance of obtaining 'objective' determinants of task proportions instead of using self-reports. When performing task-based exposure assessment, optimal task classification should contain a few, clearly identifiable tasks that must show clear exposure contrasts.

STATISTICAL METHODS TO ASSESS RELIABILITY AND EXPOSURE VARIABILITY

Rater Reliability

Statistical methods used to assess reliability of observation of posture often include proportion of agreement, the kappa statistic, the Pearson product moment correlation coefficient, the Spearman rank correlation coefficient, and various forms of the intraclass correlation coefficient. The use of these statistics depends on the data collected and the research questions. These commonly used reliability statistics are very useful and some distinguish between true and error variance (Fleiss and Cohen, 1973; Shrout and Fleiss, 1979). However, a statistic utilized in the social sciences known as Generalizability Theory, combines the utility of the intraclass correlation coefficient with a primary emphasis on examining the magnitudes of error from all possible sources.

"Generalizability (G) theory is a statistical theory about the dependability of behavioral measurements" (Shavelson and Webb, 1991). In G-theory, the generalizability of a measure depends on the research goals. G-Theory has foundations in classical test theory and is based on the analysis of variance (Deshon, 2002). G-theory extends the use of the intraclass correlation coefficient by allowing one to estimate the variance components and decompose the error into its constituents, thus allowing multiple sources of error to be estimated separately in a single analysis (Deshon, 2002;

Lievens et al, 2004; Shavelson and Webb, 1991). Classical reliability theory distinguishes only between true and error variance, whereas generalizability theory allows simultaneous source variance estimation (Lievens et al, 2004; VanLeeuwen, 1997). Classical reliability statistics, such as the kappa coefficient, do not provide information on whether or not inferences obtained can be safely generalized across all variables in the study (Deshon, 2002). G-theory focuses on the breadth to which results are applicable. The broadest inference made from results consists of treating all variables in the model as random and evaluating them simultaneously and interactively such as a random-effects model (Deshon, 2002). Determining the sources of measurement error allows researchers to reduce errors in future measurements (Deshon, 2002).

Using G-theory, one can make relative or absolute decisions. G-theory provides a summary coefficient or G-coefficient that is analogous to the reliability coefficient obtained using classical test theory (Shavelson and Webb, 1991). The G-coefficient is an intraclass coefficient that accounts for the design of the study to partition measurement errors (Masse and Heesch, 2002). G-theory also enables researchers to determine how many sources are needed to obtain reliable outcomes (Shavelson and Webb, 1991). G-theory considers two types of studies: generalizability (G) studies and decision (D) studies. G-studies are used to estimate as many sources of error as possible, while D-studies are used to obtain measurements for a particular study design or purpose (VanLeeuwen, 1997). Information from G-studies is used to design D-studies to obtain a measurement outcome having the desired reliability level (VanLeeuwen, 1997). Researchers can predict reliability estimates under different measurement conditions such as raters and occasions (Deshon, 2002) using D-studies. One can obtain a D-coefficient for each proposed study design. This is extremely important in determining study design for future research.

While Generalizability Theory is a useful statistical tool to evaluate rater reliability, it should be noted that the more commonly used statistical methods to assess rater reliability within ergonomics research are the proportion of agreement, the kappa statistic, and the intraclass correlation coefficient. The simplest and most commonly used index of agreement for two raters with k categories is the overall proportion of agreement. Proportion of agreement is calculated using the sum of the frequencies along the main diagonals of a contingency table (Fleiss and Cohen, 1973). One problem associated with proportion of agreement statistics is that a high proportion of agreement may be interpreted as observers having the same difficulty identifying a particular posture (Burt and Punnett, 1999). Also, proportion of agreement does not take chance into account, meaning that no information is provided on chance agreement of the raters (Fleiss and Cohen, 1973).

Fleiss and Cohen (1973) defined kappa as "the proportion of agreement corrected for chance, scaled to vary from –1 to +1 so that a negative value indicates poorer than chance agreement, zero indicates exactly chance agreement, and a positive value indicates better than chance agreement." The kappa coefficient is commonly used to assess rater agreement. Kappa is a useful measure to determine inter-rater reliability for categorical scales, although it may vary greatly depending on the prevalence of the characteristic observed (Fleiss and Cohen, 1973). Some researchers have concluded that the interpretation of kappa should not draw inferences beyond what the data justify (Kraemer and Bloch, 1988). A large problem associated with the kappa statistic is attempting to compare its magnitude with that of the intraclass correlation coefficient which is used with quantitative data and is interpretable as a proportion of variance (Fleiss and Cohen, 1973).

Fleiss and Cohen (1973) described that the reliability of quantitative data is better measured by the intraclass correlation coefficient, which they considered a special case of weighted kappa. Intraclass correlation coefficients assess agreement of quantitative measures (Muller and Buttner, 1994) and are estimated using variance components from analysis of variance models. Since assessment of reliability is variance dependent, the intraclass correlation coefficient may be low even when proportion of agreement is high (Bartko, 1994). Maclure and Willett (1987) concluded that the intraclass correlation coefficient is superior to kappa when analyzing ordinal data. However, they stated that choosing logical standard weights that adjust for magnitude of agreement/disagreement makes the weighted kappa statistic equivalent to using the intraclass correlation coefficient. Weights represent the relative seriousness of each kind of disagreement (Fleiss and Cohen, 1973). Shrout and Fleiss (1979) described that while the intraclass correlation coefficient is a useful tool to assess reliability, there are numerous versions that may produce very different results and interpretations.

Inter-Method Reliability

Statistical methods used to assess validity of observation of posture are similar or identical to those used to assess reliability and include the Spearman rank correlation coefficient, the Pearson product moment correlation coefficient, the kappa statistic, and the use of the analysis of variance (ANOVA). In general, ergonomic studies attempting to investigate validity compare one exposure assessment tool to a ‘reference’ or to a tool that is considered more accurate than the tool being investigated. In the case of comparing a tool to a ‘reference’ or ‘gold standard’, one would consider this to be an investigation into the validity of that tool. When comparing a tool to one that is considered more accurate, one would consider this to be an investigation in the inter-method reliability or cross-validation of that tool (Spielholz et al., 2001). This demonstrates the similarity in statistical approach to reliability and validity studies. Ultimately, the choice of statistic relies on the study approach and the type of data collected.

Exposure Variation and Measurement Strategy

Within- and between-worker exposure variability is typically assessed using the ANOVA. ANOVA models are a useful tool for quantitatively describing variability in exposure by partitioning the variability into variance components (Loomis and Kromhout, 2004).

Methods typically used to quantify aspects of exposure variation: the amplitude probability distribution function (APDF) and the contraction frequency analysis (CFA). However, these methods omit important aspects of the exposure variation (Mathiassen and Winkel, 1991). Neither the APDF nor CFA consider the length of the analysis period or reflect changes in the distribution of the variable interest (posture) along a real-time scale (Mathiassen and Winkel, 1991). In regards to assessing variation in exposures over a period of time, a statistical method known as Exposure Variation Analysis (EVA) has been used in ergonomics research that apply electromyography (EMG) and electrogoniometry. Mathiassen and Winkel (1991) developed EVA as a data reduction method to assist with effectively quantifying EMG or electrogoniometry data. Exposure

Variation Analysis allows one to quantify variation in physical workloads in relation to intensity and duration. Exposure Variation Analysis describes the intensity of the measured variable during a period of work, as well as the duration at each intensity level (Anton et al., 2003). This analysis allows one to measure multiple exposure dimensions simultaneously (Anton et al., 2003). Typically, researchers evaluate each EVA category using the application of the ANOVA. Summary measures like EVA reduce detailed full-shift data into a simpler array that is easy for researchers and non-researchers to comprehend (Anton et al., 2003).

SECTION THREE

STUDY I: RATER RELIABILITY

ABSTRACT

The purpose of this study was to determine the inter- and intra-rater reliability of assessing upper extremity postures of workers performing manufacturing tasks Assessment of neck, shoulder, and wrist postures of 20 manufacturing employees was conducted by two raters observing digital video files using Multimedia Video Task Analysis (MVTA). Generalizability theory was used to estimate the inter- and intra-rater reliability. The results demonstrated good to excellent inter-rater reliability for neck and shoulder postures and fair to excellent inter-rater reliability for wrist postures. Intra-rater posture assessment demonstrated good to excellent reliability for both raters in all postures of the neck, shoulder, and wrist. This study demonstrated that posture assessment of manufacturing workers using MVTA is a reliable method.

The manuscript for this study was published in a peer-reviewed journal. A copy of this manuscript is found in Appendix A. Journal citation and publisher copyright information is as follows:

> Reliability of assessing upper limb postures among workers performing manufacturing tasks. Dartt, A., Rosecrance, J., Gerr, F., Anton, D., and Merlino, L. *Applied Ergonomics*, 40: 371-378. © 2009, Elsevier.

SECTION FOUR

STUDY II: RATER RELIABILITY EXPANDED

ABSTRACT

Rater reliability is a critical component when assessing postures of workers from video observation. The purpose of this study was to evaluate inter- and intra-rater reliability of the upper limbs of both cyclic and non-cyclic manufacturing tasks. Assessment of neck, shoulder, and wrist postures of 40 manufacturing workers was conducted by four raters observing previously recorded digital video files. Generalizability theory, a form of the intraclass correlation coefficient, and the Pearson Product Moment correlation coefficient were used to estimate inter- and intra-rater reliability. Reliability coefficients ranged from poor to excellent depending on the anatomical area and posture evaluated. Evaluation of rater differences revealed that the quality of the video picture, the nature of the tasks assessed, and factors related to the analysis procedures influenced reliability measures. Based on the study results, a series of recommendations for improving reliability during video observation were identified. The primary recommendations included: 1) use of a systematic training program with precise definitions of postures, detailed posture estimation guidelines and decision criteria and extensive feedback during training; 2) pilot video capture and use of a two-camera set-up; 3) application of multiple statistical methods; 4) the use of a missing data category; and 5) early and ongoing evaluation of rater reliability throughout research studies.

INTRODUCTION

Rater reliability is a critical component when quantifying postures from video observation. While rater reliability does not ensure validity, when it is not properly established, the data and interpretations of the data cannot be considered valid. The purpose of this study was to expand the scope of a previous study (Dartt et al., 2009) completed by the researchers and further investigate the inter- and intra-rater reliability of assessing upper extremity postures of workers performing manufacturing tasks. This study was performed using video recordings of manufacturing tasks obtained by Heartland Education and Research Center investigators from a larger prospect cohort study. This study was funded as a pilot research project by the Rocky Mountain Center for Occupational and Environmental Health (University of Utah) in 2008.

The present study was performed in a similar manner as the previous study (Dartt et al., 2009), but expanded the previous study in several areas. The present study utilized four raters who analyzed postures of the upper limbs of workers for 40 video-recorded cyclic and non-cyclic manufacturing. Therefore, the present study expanded the number of raters, the number of tasks analyzed, and the type of work evaluated (cyclic and non-cyclic tasks). The postures analyzed were the same as those analyzed in the previous study (Dartt et al., 2009). Rationale for posture selection included biomechanical and pathophysiological factors as well as previous research studies (Armstrong et al., 1982; Stetson et al., 1991; McAtamney and Corlett, 1993; Juul-Kristensen et al., 2001; Spielholz et al., 2001; Lowe, 2004a; Lowe, 2004b). If a physiological reason was not available, posture categories were selected based on previously developed tools, to maintain bi-directional consistency, or by using half the range of motion for that specific body segment.

The primary objectives of the present study were to:

- Assess inter- and intra-rater reliability of four raters that were evaluating the duration of time spent in postures of the upper extremities of workers performing cyclic and non-cyclic manufacturing tasks using video observation via MVTA.
- Apply a novel statistic (Generalizability Theory) to a rater reliability study for ergonomics research and compare this statistic to commonly applied reliability statistics in the ergonomics literature.

The hypotheses for the present study were:

- Intra-rater reliability for shoulder postures will be good to excellent ($\rho > 0.75$).
- Intra-rater reliability for neck and wrist postures will be at least fair to good ($0.50 < \rho < 0.75$).
- Inter-rater reliability for shoulder postures will be good to excellent ($\rho > 0.75$).
- Inter-rater reliability for neck and wrist postures will be at least fair to good ($0.50 < \rho < 0.75$).
- Rater reliability coefficients will be higher for analyses of cyclic tasks as compared to analyses of non-cyclic tasks.

METHODS

Description and Training of Raters

Four raters were recruited within the research institution. Before any video analyses were conducted, raters attended a 20-hour training program. The training program was based on recommendations from a previous study on reliability of observation posture assessment (Dartt et al., 2009). Training consisted of three steps. During Step 1, the raters and an experienced trainer observed and discussed posture analyses of 6-10 working tasks. Detailed posture estimation guidelines, created previously, were used in this step of the training (Appendix B). This first step of the training was approximately 3-4 hours. During Step 2, raters analyzed four tasks and compared (frame·by·frame) results to that of the trainer's results. Any discrepancies in posture selections were discussed. The second step of the training was approximately 12-15 hours. During Step 3, raters began to analyze tasks independently with a weekly review by the trainer.

Video Recorded Tasks

Investigators from the University of Iowa Occupational and Environmental Health Department had previously recorded over 700 manufacturing tasks using two camera views in a larger prospective cohort study. The Iowa investigators digitally synchronized the two camera views and saved to file. Video files were sent to the Colorado State University investigators for posture assessment.

For each worker recruited into the larger prospective cohort study, a 45 minute sample of each task they performed was video recorded. Video samples were reviewed by the investigators. For cyclic tasks, three cycles were extracted from a 30 minute video record of the tasks (one cycle from each of the following time ranges: 0-5 minutes, 15-20 minutes, and 25-30 minutes). For non-cyclic tasks, 5 minutes of time were extracted from the middle 20 minutes of the recorded non-cyclic tasks. For those 20 minutes, a 30 second time period was analyzed at the beginning of 2-minute intervals, resulting in a total analysis time of 5 minutes.

Upon meeting eligibility for the study, videos were randomly selected from the total list of videos via a random number selection process. Videos considered for the present study had to meet several levels of eligibility. First, any videos analyzed in the previous study (Dartt et al, 2009) or as part of the larger prospective cohort study could not be used in the present study. Second, any videos to be considered had to be in possession of the CSU investigators. Third, at least 25% of the videos to be analyzed had to contain non-cyclic tasks. This percentage was based on the proportion of non-cyclic tasks in the total study population.

Video Analysis Using MVTA

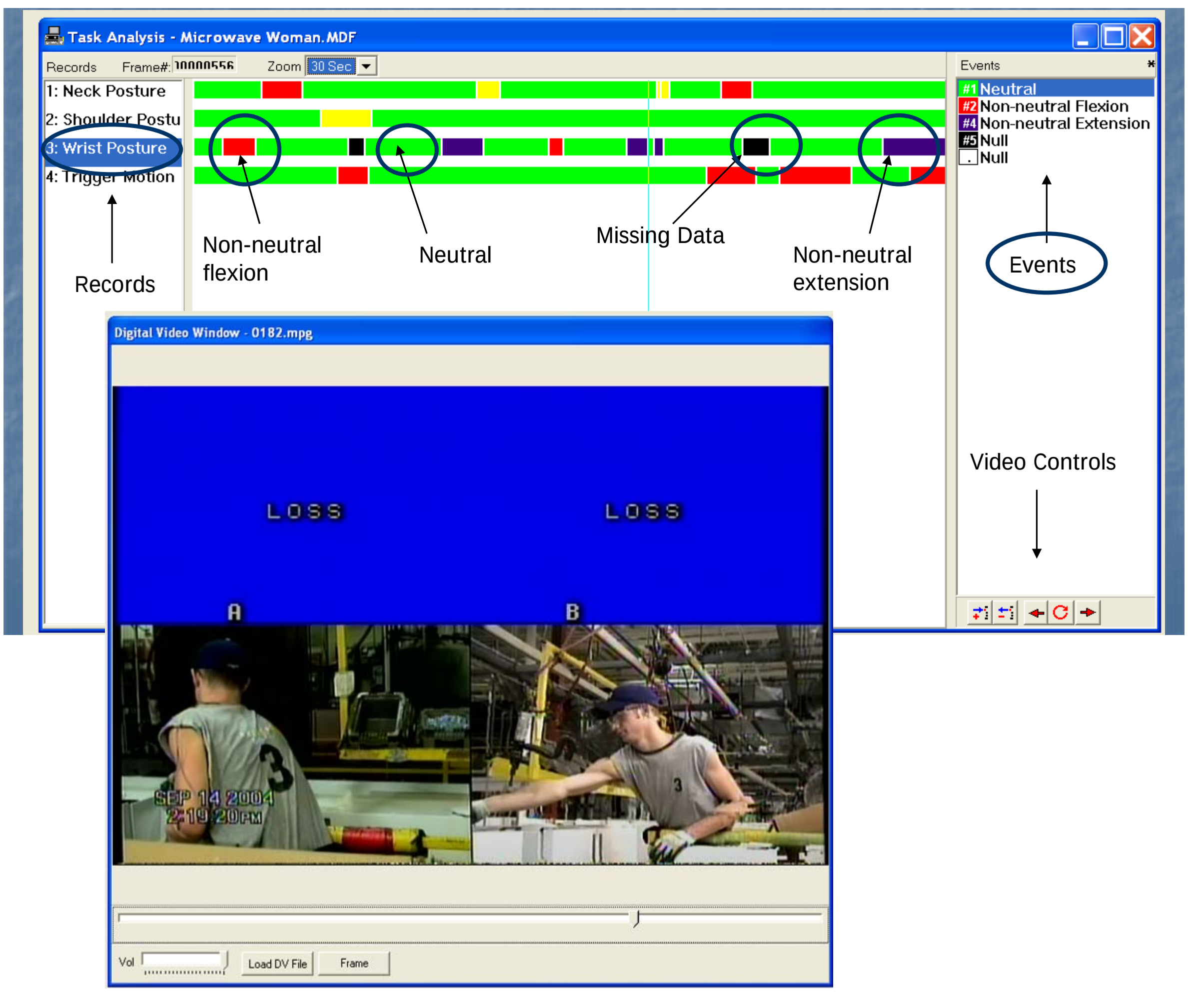
Analysis took place in the Ergonomics Research Lab at Colorado State University located in Fort Collins, Colorado. A computer setup with two side-by-side monitors was used for video analysis. Multimedia Video Task Analysis (MVTA), created by the Ergonomics Research Consortium at the University of Wisconsin-Madison, had previously been installed on the computer used. Videos were sent by University of Iowa researchers in mpeg format stored on digital video discs. They were downloaded onto the computer. Once the videos were accessible on the computer, MVTA was utilized to perform the other tasks needed for analysis.

Multimedia Video Task Analysis is a video-based exposure assessment program that can be used to automate time and motion studies and ergonomic analyses from video (Ergonomics Analysis and Research Consortium, 2003). This program uses a computer controlled videocassette recorder that allows interactive study of video footage enabling researchers to time log data such as posture events. Video is reviewed at any speed and in any sequence (real time, slow/fast motion, or frame by frame).

To analyze postures, a video window (Figure 4.1) and task analysis window were necessary. The video window was a viewing area for a video as well as VCR controls. Angle estimation was aided by known angles printed on transparencies to the angular degrees specified by the study parameters. These transparent angles were placed on the computer monitor screen to aid in angle estimation when views of the body part of interest were perpendicular to the viewing line of sight. The task analysis window displayed the raters' logged postures on a timeline (Figure 4.1). To log data, the raters utilized two categories:

FIGURE 4.1

MVTA Task Analysis and Video Window



records and events. The records were the specific anatomical areas of interest: neck, shoulder, or wrist. These are displayed on the left side of Figure 4.1. Each record had a list of events. Events consisted of the different posture categories for the neck, shoulder, and wrists. Events for the wrist are displayed on the right side of Figure 4.1. Refer to Table 4.1 for posture categories of the neck, shoulder, and wrist. Each different posture for each body part had a corresponding keystroke and color to mark the beginning and end of that event. The middle section of Figure 4.1 illustrates a posture observation log and corresponding events. An additional event category was created for each body part that could not be adequately observed. This event category was labeled as "missing data."

TABLE 4.1

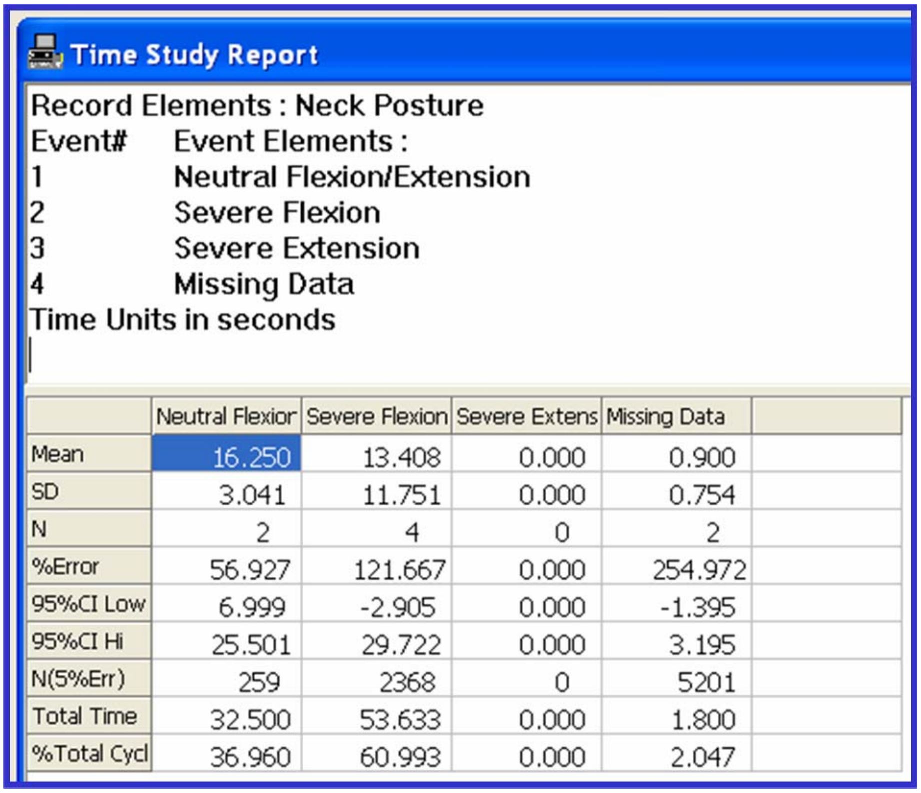
Postures of the Neck, Shoulder, and Wrist

Neck	Shoulder	Wrist
Extension >20˚	Neutral (0˚flex/abd-60˚ flex/abd)	Extension >30˚
Neutral (20˚ ext-45˚ flex)	Mild Flexion/Abduction 60˚-90˚	Neutral (30˚ ext-30˚flex)
Flexion >45˚	Severe Flexion/Abduction >90˚	Flexion >30˚

The primary variable of interest for this study was the duration of task time spent in the different posture categories of the neck, shoulder, and wrist. After completion of each subject analysis, MVTA was used to generate a time study report (Figure 4.2). These reports provided the percentage of total cycle time spent in each posture category for all three body parts. These percentages were then used in the statistical analyses. Figure 4.2 represents an example time study report for the neck. The last row of the table shows the percent of the total time analyzed for each neck posture.

FIGURE 4.2

Example of a Time Study Report Generated by MVTA



Time Study Report

Record Elements : Neck Posture

Event#	Event Elements :
1	Neutral Flexion/Extension
2	Severe Flexion
3	Severe Extension
4	Missing Data

Time Units in seconds

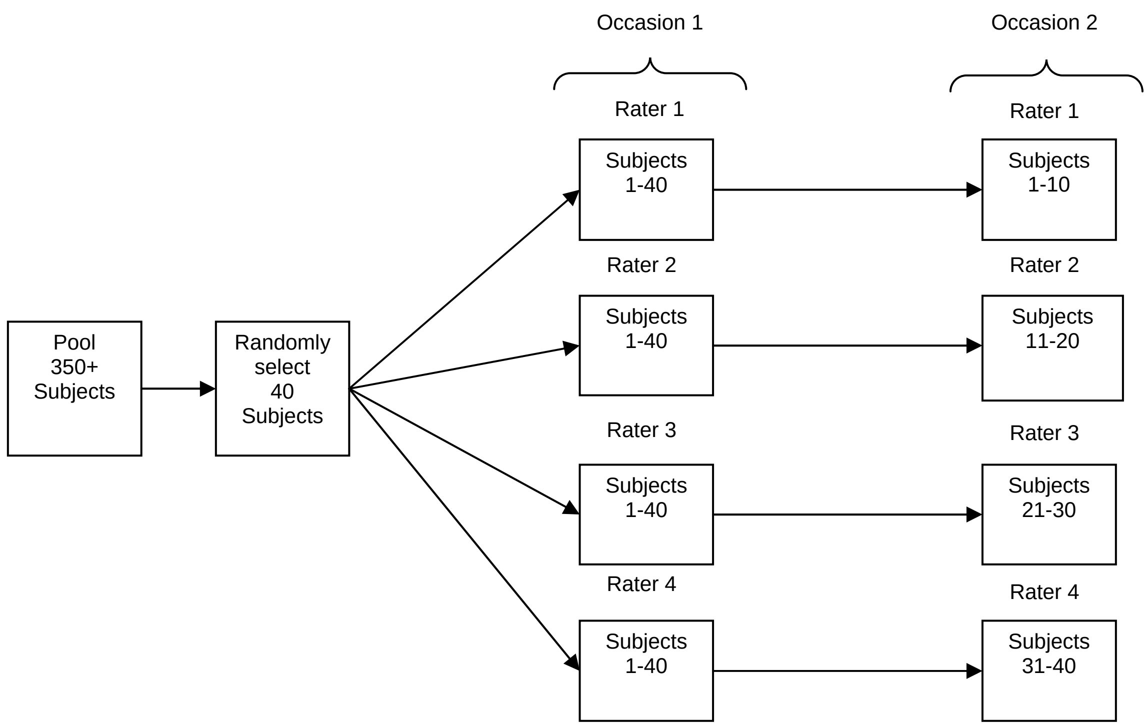
	Neutral Flexior	Severe Flexion	Severe Extens	Missing Data	
Mean	16.250	13.408	0.000	0.900	
SD	3.041	11.751	0.000	0.754	
N	2	4	0	2	
%Error	56.927	121.667	0.000	254.972	
95%CI Low	6.999	-2.905	0.000	-1.395	
95%CI Hi	25.501	29.722	0.000	3.195	
N(5%Err)	259	2368	0	5201	
Total Time	32.500	53.633	0.000	1.800	
%Total Cycl	36.960	60.993	0.000	2.047	

Data Collection Procedures

Each of the four raters assessed forty identical work tasks selected at random from a sample of 700. Using MVTA, raters assessed postures of the neck, shoulder, and wrist (Table 4.1). Figure 4.3 provides a diagram of the research design. In addition to the postures of categories of interest, raters could assign a 'missing data' category. This category was used when the anatomical areas were obstructed from camera view. Raters maintained a log of each assessment they completed explaining their reason for assigning 'missing data'. Four weeks following analyses, raters repeated the MVTA analyses on 10 of the 40 previously analyzed tasks. Raters maintained a log of each task analysis and indicated the time required to complete each assessment and an explanation of their reasons for assigning ‘missing data.’ Multimedia Video Task Analysis was used to generate reports that detailed the percent time spent in each posture as well as the ‘missing data’ category. MVTA was used to generate reports that detailed the percent time spent in each posture as well as the ‘missing data’ category.

FIGURE 4.3

Research Design



Sampling Duration and Sample Size

Sampling duration as described above was based on the nature of the tasks, total sample size, and the detail of the analysis. MVTA allows for continual analysis of posture based on the video collection frame rate (30 frames/second). Therefore, a posture was rated for every frame over the entire analysis period. A task with a mean analysis of time of 2.9 minutes would result in 5220 data points. While analysis duration may appear to be short, the frame·by·frame analysis capability results in a large number of data points.

Sampling duration was also based upon previous research of reliability for other posture assessment tools (Keyserling, 1986; Burt and Punnett, 1999; Juul-Kristensen et al., 2001; Spielholz et al., 2001; Lowe, 2004a; Lowe, 2004b). Sampling duration for those studies referenced in the literature review ranged from 2 to 5 minutes. Analysis durations for several of these studies were difficult to estimate, because unlike the continuous analysis of postures of the present study, most studies analyzed a specific posture for a pre-determined number of frames. An example of this method is explained by Bao et al. (2009) where raters estimated posture angles from 37-38 randomly selected frames in four different 15 minute video clips.

In regards to sample size, Walter et al. (1998) set an optimal sample size for the intraclass correlation coefficient (ICC) based on the desired power, magnitude of the predicted ICC, and the lower confidence limit. Based on these parameters, Walter et al. (1998) concluded that if the customary 0.95 confidence level and 0.80 power level were used and there were two ratings per subject, the necessary sample size to prove the estimated ICC was different from zero would range from 5 when the estimated ICC was 0.9 to 616 when the estimated ICC was 0.1. A similar study performed by Bonett (2002) concluded that optimum sample size is a function of the size of the ICC, number of ratings per subject, desired significance level (α), and the desired width (w) of the confidence interval. The smallest sample size suggested was 15. Shoukri et al. (2004) examined sample size requirements for reliability studies. A summary table utilized $n=2$, $\alpha=0.05$, and $\beta=0.20$ to estimate the number of k subjects needed to reach the desired significance and power levels. With a minimal acceptable reliability coefficient of 0.6 and a desired reliability coefficient of 0.80, 39 subjects would be necessary to achieve the desired significance and power. Based on the reliability coefficients found in the previous study (Dartt et al, 2009) and the parameters of the present study, it was concluded that 40 tasks had adequate power (β=0.20) to permit accurate estimation of the reliability coefficients. This sample size assessment utilized the ICC as a referent. The previous study (Dartt et al., 2009) and the present study utilized Generalizability Theory for the statistical analysis. The advantage of Generalizability Theory is the identification of sources of measurement error. It also allows one to estimate a desired level of generalizability by altering the number of observations and/or raters.

Data Analysis

Generalizability theory (G-theory) was used to determine the intra- and inter-rater reliability of posture observations for the neck, shoulder, and wrist using MVTA. This theory was used to examine the generalizability of inferences concerning rater observations of posture across subjects and occasions. Generalizability theory is a comprehensive measure of reliability commonly used in the social sciences (reference), but very applicable in health and safety research. The statistical model that drives G-theory is the analysis of variance (ANOVA) (Burns, 1998). Subjects as the object of measurement are a common source of variation in most G-theory models. Other variables, or facets, in the model such as occasions or raters can be added (Burns, 1998). G-theory facets are analogous to factors or variables in an ANOVA. G-theory provides a summary coefficient expressed as rho (ρ) that is analogous to the intraclass correlation coefficient (Shavelson and Webb, 1991).

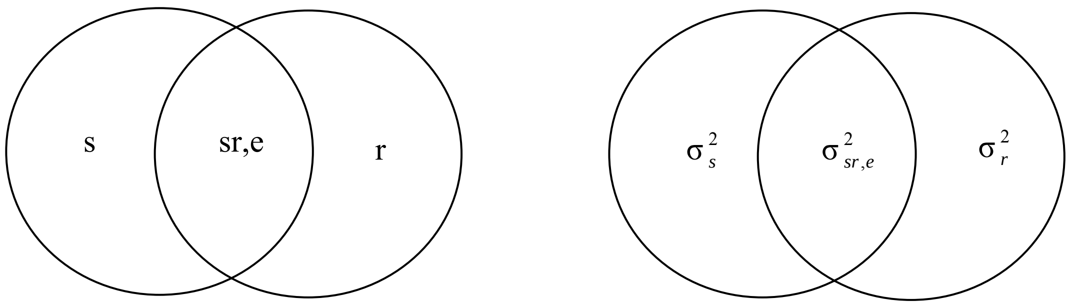
The primary variable of interest in determining reliability was the duration of total task time spent in a specific posture category (Table 4.1) as well as the missing data category for cyclic versus non-cyclic tasks. Analysis of variance was used to obtain variance estimates for the variables of interest. The variance estimates were then used to calculate absolute reliability coefficients for each posture assessed. Intra-rater reliability was determined using a repeated measures design. To facilitate comparison with other research studies, Pearson product moment correlation coefficients were also estimated.

Inter-Rater Reliability

Data obtained by the four raters were used to investigate inter-rater reliability. Data obtained by the four raters during the first set of 39 analyses (Occasion 1 – Figure 4.3) were evaluated for cyclic and non-cyclic tasks combined as well as separately. Variance estimates were obtained for a subject x rater fully-crossed random effects model. The combined model consisted of four raters and 39 subjects, the cyclic task model consisted of four raters and 29 subjects, and the non-cyclic model consisted of four raters and 10 subjects. Variance estimates were obtained using SPSS 17.0 statistical package. Figure 4.4 illustrates the sources of variability and variance components, where 's' represents subject, 'r' represents rater, and 'e' represents residual error. Figure 4.4 is modeled after example figures in Generalizability Theory: A Primer (Shavelson and Web, 1991). Subjects and raters were treated as random variables. Variance estimates were obtained for each posture category (Table 4.1) of the neck, shoulder, and wrist as separate analyses. Posture analyses were completed separately since the posture categories were not independent of each other. Explanations of the variance components for the analyses performed at Occasion 1 are provided in Table 4.2.

FIGURE 4.4

Venn Diagrams for the Subject(s) x Rater(r) Design



(a) Sources of Variability (b) Variance Components

In the residual (e) variance component, the subject(s) x rater(r) interaction is confounded with unmeasured or unsystematic variability. The subject(s) x rater(r) component reflects whether the postures of the subjects differed across raters. The residual (e) component reflects unsystematic or random error sources. It also includes systematic error from variables not explicitly included or controlled for in the study.

TABLE 4.2

Variance Components

Variable	**Symbol**	**Description**
Subject variance	σ_s^2	Variance in posture ratings across subjects or how much subjects differ from one another in posture ratings
Rater variance	σ_r^2	Variance in posture ratings across raters or how much raters differ from each other in posture ratings
Subject x Rater variance, residual	$\sigma_{sr,e}^2$	Variance of posture ratings due to the subject x rater interaction and variance that cannot be explained by raters

Variance estimates were then used to compute G-theory reliability coefficients (Table 4.3). The equation presented in Table 4.3 was used to compute all Generalizability coefficients, differing only by the data analyzed: cyclic and non-cyclic tasks combined, cyclic tasks only, and non-cyclic tasks only. Reliability coefficients were calculated for interpretation of absolute decisions. This type of generalizability indexes the absolute level of the posture assessment for each subject with no comparative reference to the posture assessments of the other subjects (VanLeeuwen, 1997).

TABLE 4.3

Formula Used to Estimate the Reliability Coefficients at Occasion 1

Study Design		
Rater	**Occasion**	**Coefficient estimation formulas**
4	1	$\dfrac{\sigma_s^2}{\left\{\sigma_s^2 + \sigma_r^2/4 + \sigma_{sr,e}^2/4\right\}}$

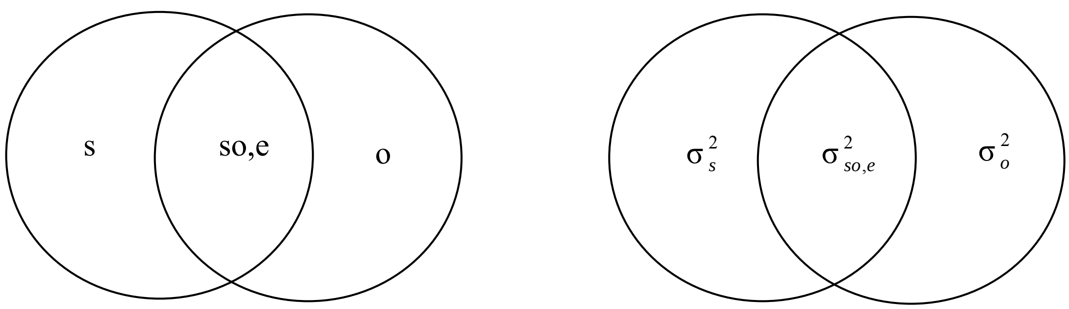
Intra-Rater Reliability

Data obtained by the four raters on the two separate occasions were used to investigate intra-rater reliability. A test-retest design was used to evaluate the intra-rater reliability of the four raters, each of whom rated 10 different subjects across the two occasions (Figure 4.3). Ratings by each of the raters at Occasions 1 and 2 were analyzed using the ANOVA. Variance estimates for Raters 1-4 were obtained for the following facets using the SPSS 17.0 statistical package for windows: subjects, occasions, subjects x occasions, and residual. Figure 4.5 illustrates the sources of variability and variance components, where ‘s’ represents subject, ‘o’ represents occasion, and ‘e’ represents residual error. Figure 4.5 was modeled after example figures in Generalizability Theory: A Primer (Shavelson and Web, 1991).

To compute the variance estimates, subjects and occasions were treated as random variables. Variance estimates were obtained for each posture category (Table 4.1) of the neck, shoulder, and wrist as separate analyses. Posture analyses were completed separately since the posture categories were not independent of each other. Explanations of the variance components for the analyses of intra-rater reliability are provided in Table 4.4.

FIGURE 4.5

Venn Diagrams for the Subject(s) x Occasion(o) Design



(b) Sources of Variability (b) Variance Components

In the residual (e) variance component, the subject(s) x occasion(o) interaction is confounded with unmeasured or unsystematic variability. The subject(s) x occasion(o) component reflects whether the postures of the subjects differed across occasions. The residual (e) component reflects unsystematic or random error sources. It also includes systematic error from variables not explicitly included or controlled for in the study.

TABLE 4.4

Variance Components

Variable	**Symbol**	**Description**
Subject variance	σ_s^2	Variance in posture ratings across subjects or how much subjects differ from one another in posture ratings
Occasion variance	σ_o^2	Variance in posture ratings across occasions or how much occasions differ from each other in posture ratings
Subject x Occasion variance, residual	$\sigma_{so,e}^2$	Variance of posture ratings due to the subject x occasion interaction and variance that cannot be explained by occasion

Variance estimates were then used to compute G-theory reliability coefficients for the four rates over the two occasions analyses (Table 4.5). Generalizability coefficients were computed using equations published in Shavelson and Webb (1991) as well as Deshon (2002). Generalizability coefficients to assess intra-rater reliability were computed for every posture per body part across subjects and occasions for Raters 1-4. The equation presented in Table 4.5 was used for all four raters, differing only by the rater and subjects analyzed by a particular rater. Reliability coefficients were calculated for interpretation of absolute decisions. This type of generalizability indexes the absolute level of the posture assessment for each subject with no comparative reference to the posture assessments of the other subjects (VanLeeuwen, 1997).

TABLE 4.5

Formula Used to Estimate the Intra-Rater Reliability Coefficients

Study Design		
Rater	**Occasion**	**Coefficient estimation formulas**
4	1	$\dfrac{\sigma_s^2}{\left\{\sigma_s^2 + \sigma_o^2/2 + \sigma_{so,e}^2/2\right\}}$

RESULTS

Of the 39 subjects analyzed, 50% were male, 90% were right-hand dominant, mean age was 45.2 (34-62), mean years worked at the appliance manufacturing facility was 17.7 (6-36), mean BMI was 27.9 (22.7-34.8), and all subjects had at least a high school education. Ten of the 39 analyses were considered non-cyclic tasks. Mean task time averaged over the 39 subjects for cyclic and non-cyclic tasks was 162 (SD=59) seconds and 272 (SD=102) seconds, respectively. The mean time to perform one analysis averaged over the 39 subjects, all raters, and both occasions was 68 (SD=22) minutes. The mean time to perform one analysis averaged over the 39 subjects and all raters for the first occasion was 69 (SD=22) minutes while it was 66 (SD=22) minutes for the second occasion. The mean analysis times for Raters 1 and 3 combined were, on average, 11.7 minutes shorter than the mean analysis times for Raters 2 and 4 combined.

Means of the raw data for each posture category and the missing data category for the neck, shoulder, and wrist are provided in Tables 4.6-4.8. The data provided in Table 4.6 represent the mean percent time spent in each posture category as analyzed by all raters across the 39 subjects across both occasions for both cyclic and non-cyclic tasks.

Table 4.7 represents the mean percent time spent in each posture category for cyclic tasks and Table 4.8 represents the mean percent time spent in each posture category for non-cyclic tasks. Figure 4.6 provides a column chart that describes the percent time spent in the four postures categories of the neck, shoulder, and wrist for all four raters across two occasions and 39 subjects.

TABLE 4.6

Average Percent Time Spent in Each Category for All Raters

Across Two Occasions and Thirty-Nine Subjects

	Average % Time (Range)			
	Neutral	**Flexion or Mild Flexion/Abduction**	**Extension or Severe Flexion/Abduction**	**Missing Data**
Neck	75.3 (28.2-100.0)	18.4 (0.0-71.6)	3.0 (0.0-31.0)	3.4 (0.0-40.2)
Shoulder	83.1 (46.9-100.0)	11.6 (0.0-43.8)	3.8 (0.0-40.4)	1.5 (0.0-18.2)
Wrist	63.0 (22.8-98.1)	4.3 (0.0-33.6)	11.2 (0.0-45.6)	21.5 (0.0-75.0)

TABLE 4.7

Average Percent Time Spent in Each Category

Across Two Occasions for Cyclic Tasks

	Average % Time (Range)			
	Neutral	**Flexion or Mild Flexion/Abduction**	**Extension or Severe Flexion/Abduction**	**Missing Data**
Neck	76.8 (28.2-100.0)	16.9 (0.0-71.6)	3.3 (0.0-31.0)	3.0 (0.0-40.2)
Shoulder	83.8 (51.0-100.0)	12.2 (0.0-43.8)	3.0 (0.0-40.4)	1.0 (0.0-18.1)
Wrist	62.0 (26.5-98.1)	4.3 (0.0-33.6)	12.3 (0.0-45.6)	21.4 (0.0-75.0)

TABLE 4.8

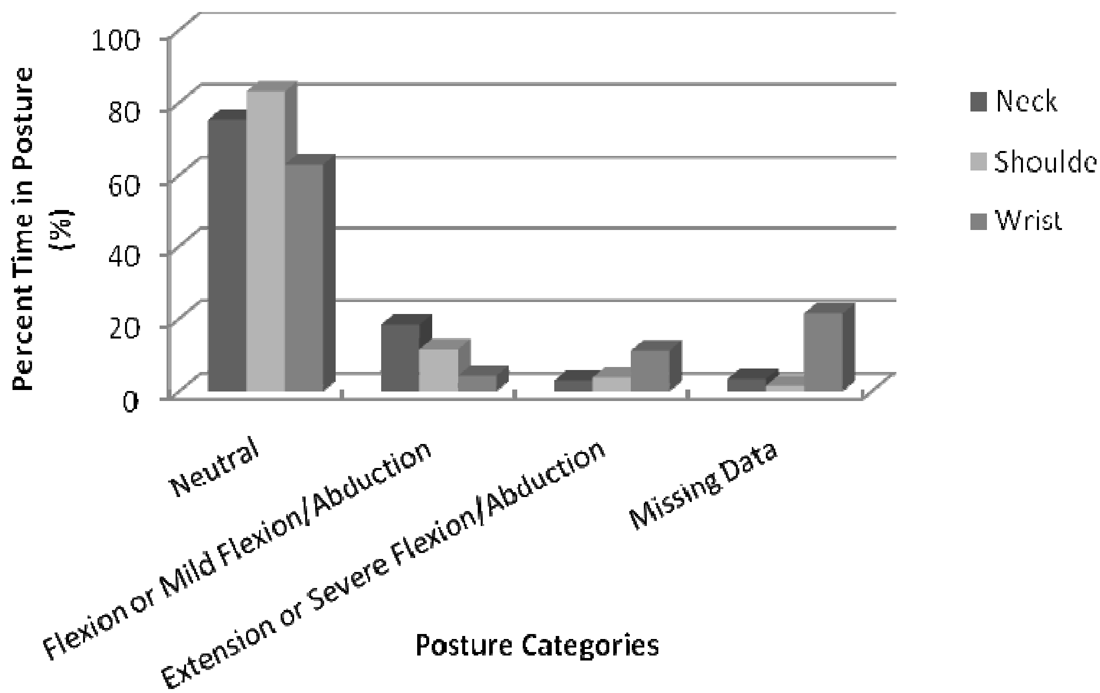
Average Percent Time Spent in Each Category

Across Two Occasions for Non-Cyclic Tasks

	Average % Time (Range)			
	Neutral	**Flexion or Mild Flexion/Abduction**	**Extension or Severe Flexion/Abduction**	**Missing Data**
Neck	70.8 (29.0-96.0)	22.5 (0.3-65.9)	1.9 (0.0-14.9)	4.7 (0.0-20.5)
Shoulder	81.4 (50.4-98.8)	9.9 (1.2-29.3)	5.9 (0.0-33.7)	2.8 (0.0-13.2)
Wrist	66.0 (33.0-91.2)	4.2 (0.0-17.8)	8.1 (0.7-25.3)	21.7 (2.9-51.6)

FIGURE 4.6

Average Percent Time Spent in Posture Categories of the Neck, Shoulder, and Wrist



Inter-Rater Reliability

Results of the analysis of variance computed for the inter-rater reliability assessment of all subjects analyzed at Occasion 1 are provided in Tables 4.9-4.11. Variance component estimations as well as the percent of total variance for each variance component are provided. Explanations for each variance component are provided in Table 4.2. Tables 4.9-4.11 provide variance estimates for cyclic and non-cyclic tasks combined, cyclic tasks only, and non-cyclic tasks only. The percent of total variance provides insight as to which variables accounted for the largest amounts of variability. For the purposes of observations of postures by raters, it is optimal for most of the variability to occur in the 'subjects' variance component. Variability within the other variance components leads to decreased reliability. Variability in the 'residual' variance component is undesirable because it is difficult to postulate the cause or causes of this variability.

TABLE 4.9

Inter-Rater Reliability Variance Estimates for Neck Postures

Posture	Task Type	Source of Variation	Estimated Variance Component	% Total Variance
Neutral	Cyclic & Non-Cyclic	subjects (s)	231.504	65.1
		raters (r)	48.624	13.7
		sr,e	75.256	21.2
	Cyclic	subjects (s)	188.648	59.0
		raters (r)	48.298	15.1
		sr,e	82.795	25.9
	Non-Cyclic	subjects (s)	343.381	76.9
		raters (r)	50.948	11.4
		sr,e	52.015	11.7
Flexion	Cyclic & Non-Cyclic	subjects (s)	287.722	83.9
		raters (r)	14.314	4.2
		sr,e	40.848	11.9
	Cyclic	subjects (s)	245.360	82.3
		raters (r)	10.312	3.5
		sr,e	42.247	14.2
	Non-Cyclic	subjects (s)	413.723	86.8
		raters (r)	28.044	5.9
		sr,e	34.664	7.3
Extension	Cyclic & Non-Cyclic	subjects (s)	10.039	44.3
		raters (r)	1.504	6.7
		sr,e	11.107	49.0
	Cyclic	subjects (s)	10.719	40.2
		raters (r)	2.030	7.6
		sr,e	13.925	52.2
	Non-Cyclic	subjects (s)	8.211	73.8
		raters (r)	0.280	2.5
		sr,e	2.632	23.7
Missing	Cyclic & Non-Cyclic	subjects (s)	11.257	23.7
		raters (r)	10.182	21.4
		sr,e	26.086	54.9
	Cyclic	subjects (s)	12.955	25.9
		raters (r)	9.308	18.5
		sr,e	27.829	55.6
	Non-Cyclic	subjects (s)	4.127	10.9
		raters (r)	10.385	27.4
		sr,e	23.365	61.7

TABLE 4.10

Inter-Rater Reliability Variance Estimates for Shoulder Postures

Posture	Task Type	Source of Variation	Estimated Variance Component	% Total Variance
Neutral	Cyclic & Non-Cyclic	subjects (s)	131.875	77.3
		raters (r)	13.425	7.9
		sr,e	25.231	14.8
	Cyclic	subjects (s)	114.020	71.7
		raters (r)	15.554	9.8
		sr,e	29.493	18.5
	Non-Cyclic	subjects (s)	198.925	90.8
		raters (r)	8.637	3.9
		sr,e	11.487	5.3
Mild Flexion\ Abduction	Cyclic & Non-Cyclic	subjects (s)	63.884	71.3
		raters (r)	8.016	8.9
		sr,e	17.757	19.8
	Cyclic	subjects (s)	76.609	71.3
		raters (r)	9.006	8.4
		sr,e	21.866	20.3
	Non-Cyclic	subjects (s)	26.225	70.5
		raters (r)	5.050	13.6
		sr,e	5.940	15.9
Severe Flexion\ Abduction	Cyclic & Non-Cyclic	subjects (s)	51.149	85.4
		raters (r)	0.302	0.5
		sr,e	8.415	14.1
	Cyclic	subjects (s)	27.809	70.4
		raters (r)	0.542	1.4
		sr,e	11.137	28.2
	Non-Cyclic	subjects (s)	122.776	99.9
		raters (r)	0.000	0.0
		sr,e	0.127	0.1
Missing	Cyclic & Non-Cyclic	subjects (s)	4.083	39.8
		raters (r)	0.619	6.0
		sr,e	5.569	54.2
	Cyclic	subjects (s)	2.762	28.6
		raters (r)	0.465	4.8
		sr,e	6.426	66.6
	Non-Cyclic	subjects (s)	5.756	58.1
		raters (r)	0.698	7.0
		sr,e	3.452	34.9

TABLE 4.11

Inter-Rater Reliability Variance Estimates for Wrist Postures

Posture	Task Type	Source of Variation	Estimated Variance Component	% Total Variance
Neutral	Cyclic & Non-Cyclic	subjects (s)	139.453	45.7
		raters (r)	115.805	37.9
		sr,e	50.065	16.4
	Cyclic	subjects (s)	140.800	42.2
		raters (r)	136.673	41.0
		sr,e	56.080	16.8
	Non-Cyclic	subjects (s)	134.161	60.4
		raters (r)	62.452	28.1
		sr,e	25.454	11.5
Flexion	Cyclic & Non-Cyclic	subjects (s)	15.830	56.5
		raters (r)	4.640	16.5
		sr,e	7.568	27.0
	Cyclic	subjects (s)	16.482	55.3
		raters (r)	4.653	15.6
		sr,e	8.648	29.1
	Non-Cyclic	subjects (s)	15.580	63.3
		raters (r)	4.134	16.8
		sr,e	4.902	19.9
Extension	Cyclic & Non-Cyclic	subjects (s)	64.890	57.1
		raters (r)	31.294	27.5
		sr,e	17.517	15.4
	Cyclic	subjects (s)	78.832	58.7
		raters (r)	35.287	26.3
		sr,e	20.206	15.0
	Non-Cyclic	subjects (s)	15.161	34.0
		raters (r)	20.022	44.9
		sr,e	9.410	21.1
Missing	Cyclic & Non-Cyclic	subjects (s)	128.463	64.0
		raters (r)	33.936	16.9
		sr,e	38.455	19.1
	Cyclic	subjects (s)	136.261	60.5
		raters (r)	41.285	18.3
		sr,e	47.789	21.2
	Non-Cyclic	subjects (s)	119.300	83.2
		raters (r)	16.203	11.3
		sr,e	7.810	5.5

Inter-rater reliability coefficients derived from G-theory for all posture categories of the neck, shoulder, and wrist for Occasion 1 (across four raters) are outlined in Table 4.12. Non-cyclic tasks had the highest reliability coefficients except for two categories: the neck missing data category and wrist extension greater than 30°. The difference in reliability coefficients varied depending on the anatomical area and specific posture category; however, the extension or severe flexion/abduction and missing data categories differed the most consistently across all three anatomical areas (Figures 4.7-4.9).

All Tasks

Inter-rater reliability coefficients for the neck posture categories ranged from 0.55 to 0.95 (Table 4.12). The neck posture category of flexion greater than 45° had the highest reliability among the neck posture categories evaluated. Inter-rater reliability coefficients for the shoulder posture categories ranged from 0.73 to 0.96 (Table 4.12). The shoulder posture category of flexion/abduction greater than 90° had the highest reliability among the shoulder posture categories evaluated. The lowest reliability value for both the neck and shoulder was observed in the missing data category. Inter-rater reliability coefficients for the wrist posture categories ranged from 0.77 to 0.88 (Table 4.12). The wrist posture categories of extension greater than 30° and flexion greater than 30° had the highest reliability among the wrist posture categories evaluated. The lowest reliability value for the wrist was obtained for the neutral posture category.

Cyclic Tasks

Inter-rater reliability coefficients for the neck posture categories ranged from 0.58 to 0.95 (Table 4.12). The neck posture category of flexion greater than 45° had the highest reliability among the neck posture categories evaluated. Inter-rater reliability coefficients for the shoulder posture categories ranged from 0.62 to 0.91 (Table 4.12). The shoulder posture categories of neutral and flexion/abduction of 60°-90° had the highest reliability among the shoulder posture categories evaluated. The lowest reliability value for both the neck and shoulder was observed in the missing data category. Inter-rater reliability coefficients for the wrist posture categories ranged from 0.75 to 0.86 (Table 4.12). The highest reliability coefficient for the wrist was found in the missing date category while the lowest reliability value for the wrist was found in the neutral posture category.

Non-Cyclic Tasks

Inter-rater reliability coefficients for the neck posture categories ranged from 0.33 to 0.96 (Table 4.12). The neck posture category of flexion greater than 45° had the highest reliability among the neck posture categories evaluated. Inter-rater reliability coefficients for the shoulder posture categories ranged from 0.85 to 1.00 (Table 4.12). The shoulder posture category of flexion/abduction greater than 90° had the highest reliability among the shoulder posture categories evaluated. The lowest reliability value for both the neck and shoulder was observed in the missing data category. Inter-rater reliability coefficients for the wrist posture categories ranged from 0.67 to 0.95 (Table 4.12). The highest reliability coefficient for the wrist was found in the missing date category while the lowest reliability value for the wrist was found in the wrist extension greater than 30° category.

TABLE 4.12

Inter-Rater Reliability Results Using Generalizability Theory

		G-Coefficients			
		Neutral	**Flexion or Mild Flexion/ Abduction**	**Extension or Severe Flexion/Abduction**	**Missing Data**
Neck	Cyclic & Non-Cyclic	0.88	0.95	0.76	0.55
	Cyclic	0.85	0.95	0.73	0.58
	Non-Cyclic	0.93	0.96	0.92	0.33
Shoulder	Cyclic & Non-Cyclic	0.93	0.91	0.96	0.73
	Cyclic	0.91	0.91	0.90	0.62
	Non-Cyclic	0.98	0.91	1.00	0.85
Wrist	Cyclic & Non-Cyclic	0.77	0.84	0.84	0.88
	Cyclic	0.75	0.83	0.85	0.86
	Non-Cyclic	0.86	0.87	0.67	0.95

FIGURE 4.7

Cyclic vs Non-Cyclic G-Coefficients of Neck Postures

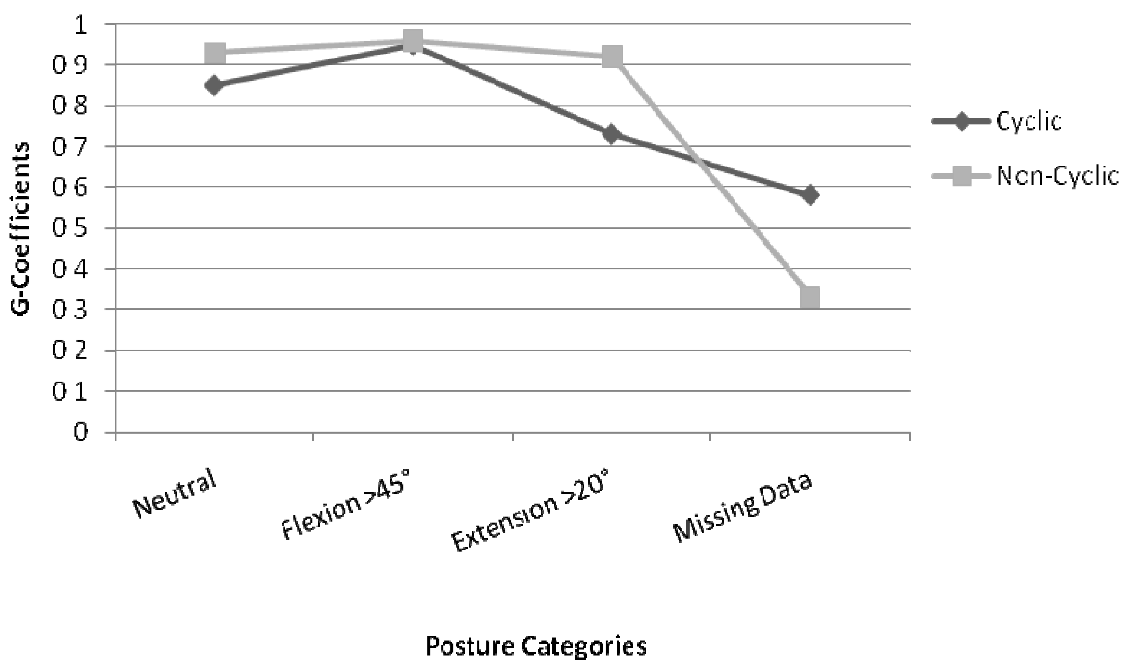


FIGURE 4.8

Cyclic vs Non-Cyclic G-Coefficients of Shoulder Postures

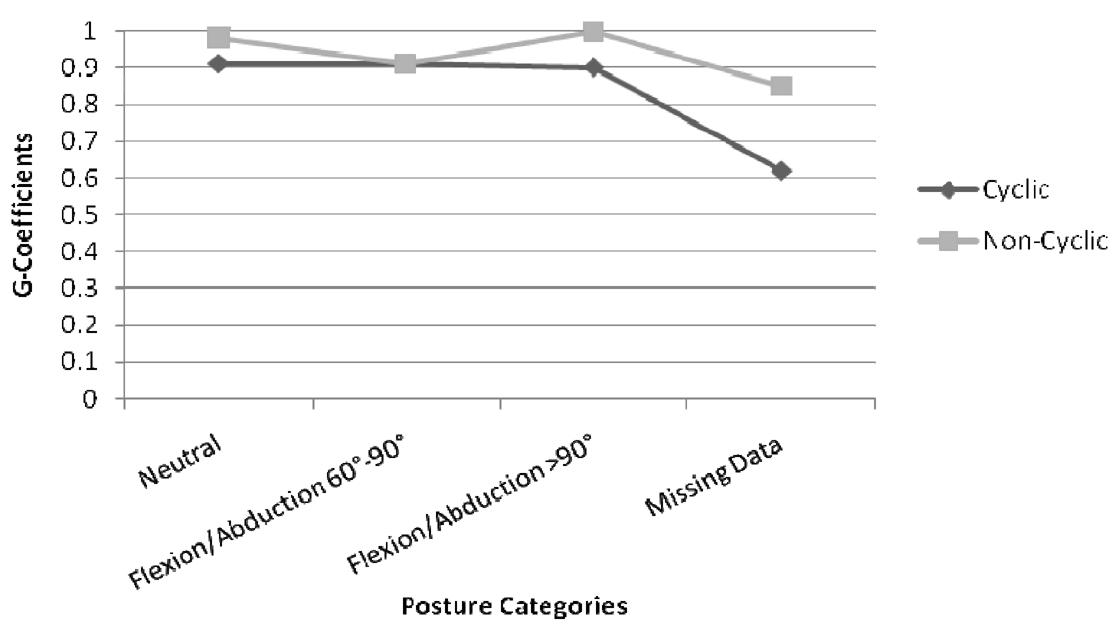
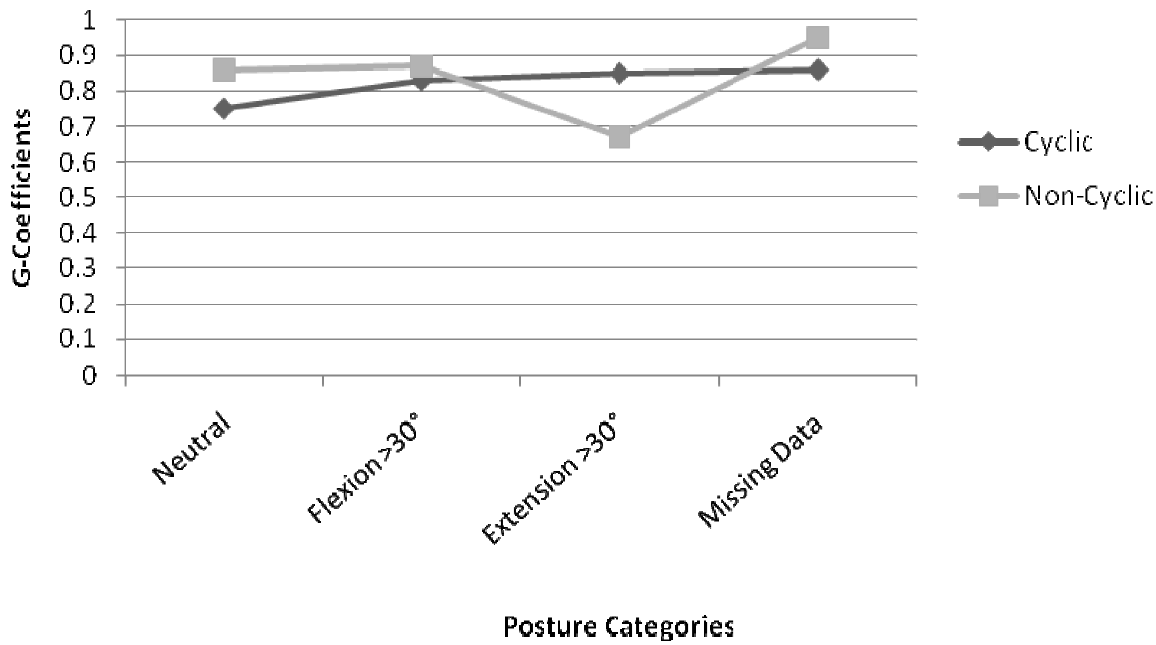


FIGURE 4.9

Cyclic vs Non-Cyclic G-Coefficients of Wrist Postures



Pearson product moment correlation coefficients are also presented as a classical test of reliability (Table 4.13). Correlations were calculated for each possible pair of raters (six rater pairs), and then the means of these correlations were computed across all pairs of raters for each posture category. In regards to specific rater pairs, seven correlation coefficients were not significant at the 0.05 level. The majority (191 correlations over six rater pairs, cyclic and non-cyclic tasks, all body parts, and all posture categories) of correlations were significant at the 0.01 level, while 12 correlation coefficients were significant at the 0.05 level. The following postures and rater pairs and non-significant correlations: neck extension >20° for Raters 1 and 4, shoulder missing data for Raters 2 and 3, and neck missing data for Raters 1 and 3, 1 and 4, 2 and 3, 2 and 4, and 3 and 4. Rater pairs differed in correlation when examining the mean correlation across all posture categories in a specific body part (Figures 4.10-4.12). Figures 4.10-

4.12 describe the Pearson correlation coefficients the neck, shoulder, and wrist for each rater pair. The means of the posture specific correlation coefficients was computed for each body part and are displayed over the columns on the figures. The shoulder (across all postures) had the highest overall Pearson correlation across all rater pairs, followed by the wrist, and then the neck (Figures 4.10-4.12). Rater pair (1,2) had the highest correlation coefficients across all body parts, while rater pair (1,4) had the lowest correlation coefficients across all body parts.

TABLE 4.13

Inter-Rater Reliability Results Using Pearson Product Moment Correlation

		r-Coefficients			
		Neutral	**Flexion or Mild Flexion/ Abduction**	**Extension or Severe Flexion/Abduction**	**Missing Data**
Neck	Cyclic & Non-Cyclic	0.78	0.89	0.58	0.54
	Cyclic	0.73	0.86	0.54	0.60
	Non-Cyclic	0.88	0.93	0.86	0.37
Shoulder	Cyclic & Non-Cyclic	0.86	0.81	0.88	0.63
	Cyclic	0.83	0.80	0.79	0.54
	Non-Cyclic	0.95	0.86	1.00	0.79
Wrist	Cyclic & Non-Cyclic	0.74	0.79	0.82	0.80
	Cyclic	0.72	0.78	0.83	0.79
	Non-Cyclic	0.84	0.86	0.74	0.94

FIGURE 4.10

Neck Posture Pearson Correlation Coefficients for Rater Pairs for All Tasks

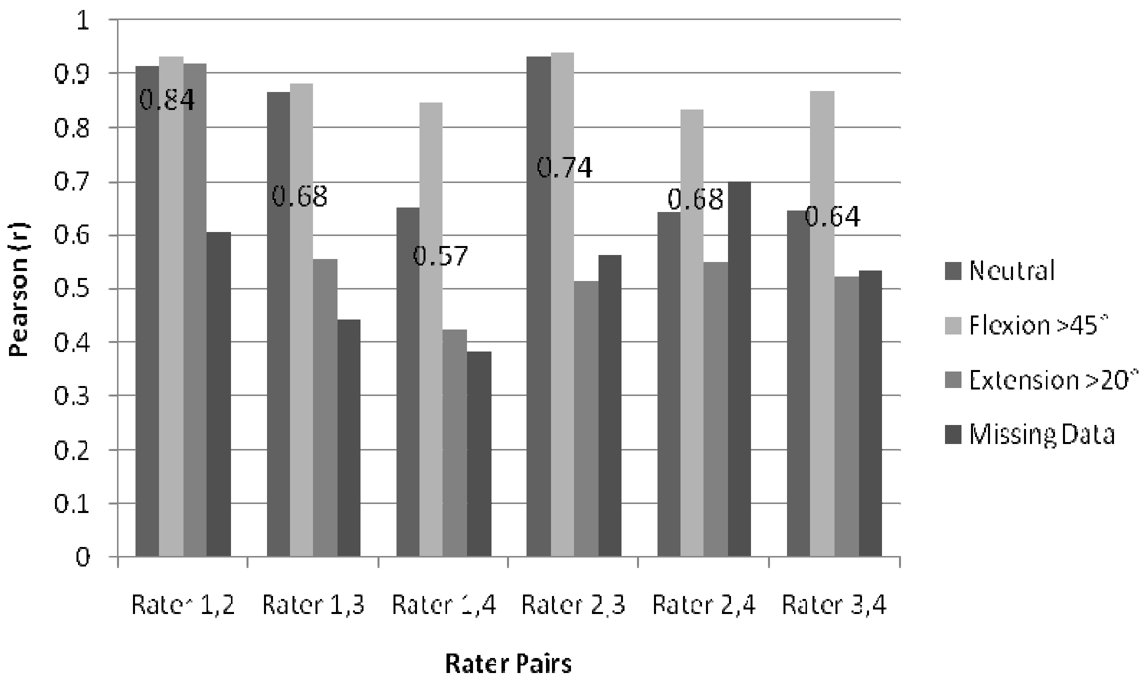


FIGURE 4.11

Shoulder Posture Pearson Correlation Coefficients for Rater Pairs for All Tasks

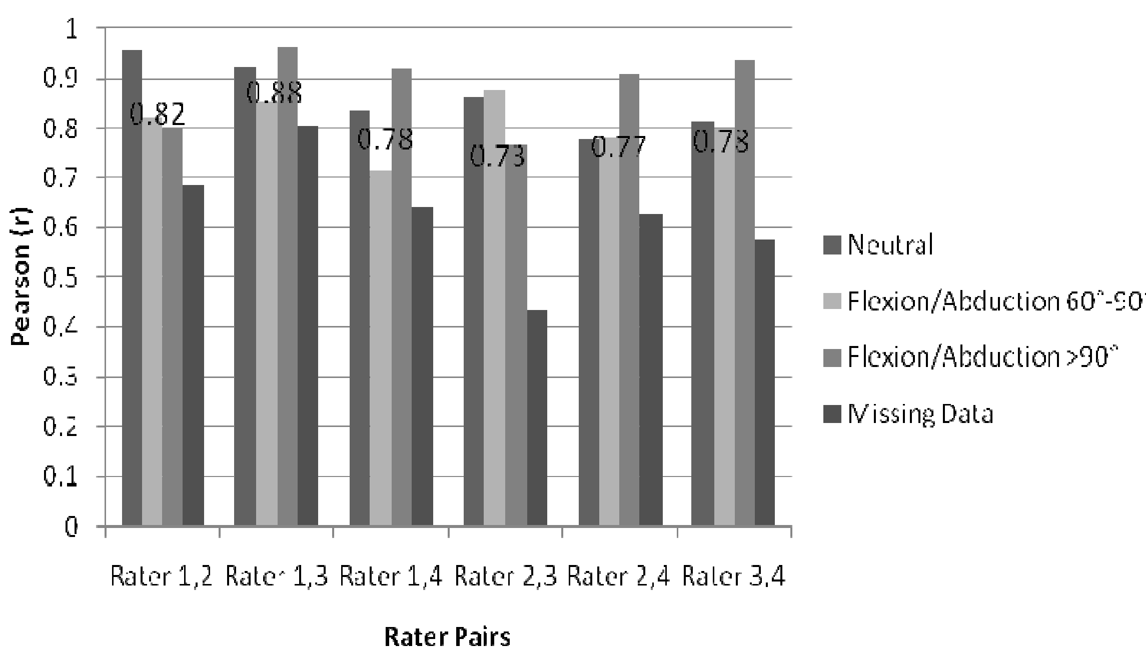
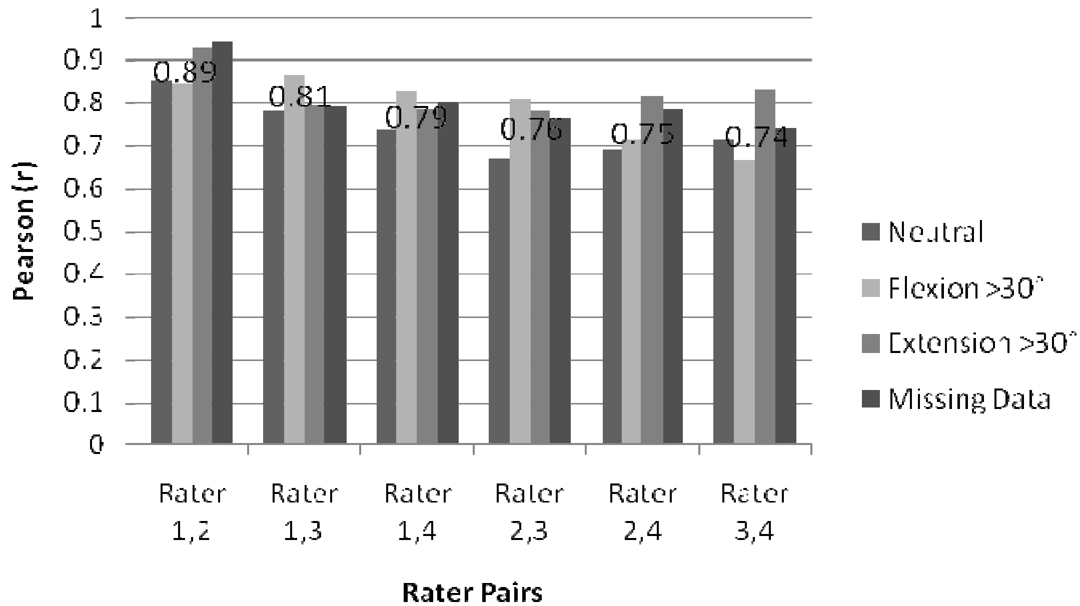


FIGURE 4.12

Wrist Posture Pearson Correlation Coefficients for Rater Pairs for All Tasks



Intra-Rater Reliability

Results of the analysis of variance computed for the intra-rater reliability assessment are provided in Tables 4.14-4.18. Tables 4.14-4.17 contain variance component estimations as well as the percent of total variance for Raters 1-4, while Table 4.17 contains variance component estimations and percent total variance for all raters combined. Explanations for each variance component are provided in Table 4.2. The percent of total variance provides insight as to which variables accounted for the largest amounts of variability. For the purposes of observations of postures by raters, it is optimal for most of the variability to occur in the 'subjects' variance component. Variability within the other variance components leads to decreased reliability. Variability in the 'residual' variance component is undesirable because it is difficult to postulate the cause or causes of this variability.

TABLE 4.14

Intra-Rater Reliability Variance Estimates for Rater 1

Posture	Task Type	Source of Variation	Estimated Variance Component	% Total Variance
Neck	Neutral	subjects (s)	364.635	80.7
		occasions (o)	34.622	7.7
		so,e	52.533	11.6
	Flexion	subjects (s)	406.619	82.2
		occasions (o)	24.281	4.9
		so,e	63.740	12.9
	Extension	subjects (s)	1.229	41.1
		occasions (o)	0.473	15.8
		so,e	1.291	43.1
	Missing	subjects (s)	2.799	67.1
		occasions (o)	0.000	0.0
		so,e	1.373	32.9
Shoulder	Neutral	subjects (s)	161.074	96.4
		occasions (o)	1.831	1.1
		so,e	4.270	2.5
	Flexion	subjects (s)	77.995	97.8
		occasions (o)	0.501	0.6
		so,e	1.260	1.6
	Extension	subjects (s)	33.154	97.2
		occasions (o)	0.121	0.4
		so,e	0.823	2.4
	Missing	subjects (s)	4.038	92.6
		occasions (o)	0.039	0.9
		so,e	0.282	6.5
Wrist	Neutral	subjects (s)	151.556	71.7
		occasions (o)	9.420	4.5
		so,e	50.341	23.8
	Flexion	subjects (s)	10.767	73.8
		occasions (o)	24.468	9.6
		so,e	39.406	16.6
	Extension	subjects (s)	104.615	62.1
		occasions (o)	24.468	14.5
		so,e	39.406	23.4
	Missing	subjects (s)	200.269	84.4
		occasions (o)	4.631	1.9
		so,e	32.487	13.7

TABLE 4.15

Intra-Rater Reliability Variance Estimates for Rater 2

Posture	Task Type	Source of Variation	Estimated Variance Component	% Total Variance
Neck	Neutral	subjects (s)	294.350	96.4
		occasions (o)	0.000	0.0
		so,e	10.928	3.6
	Flexion	subjects (s)	260.227	98.9
		occasions (o)	0.000	0.0
		so,e	2.817	1.1
	Extension	subjects (s)	6.475	69.1
		occasions (o)	0.000	0.0
		so,e	2.890	30.9
	Missing	subjects (s)	3.718	51.0
		occasions (o)	0.000	0.0
		so,e	3.567	49.0
Shoulder	Neutral	subjects (s)	141.870	94.2
		occasions (o)	0.139	0.1
		so,e	8.668	5.7
	Flexion	subjects (s)	93.973	89.4
		occasions (o)	0.934	0.9
		so,e	10.253	9.7
	Extension	subjects (s)	93.631	97.2
		occasions (o)	0.155	0.2
		so,e	2.531	2.6
	Missing	subjects (s)	1.589	85.9
		occasions (o)	0.049	2.6
		so,e	0.213	11.5
Wrist	Neutral	subjects (s)	112.234	64.9
		occasions (o)	11.011	6.3
		so,e	49.771	28.8
	Flexion	subjects (s)	59.194	89.6
		occasions (o)	1.338	2.0
		so,e	5.580	8.4
	Extension	subjects (s)	25.777	41.3
		occasions (o)	10.800	17.3
		so,e	25.852	41.4
	Missing	subjects (s)	52.266	80.7
		occasions (o)	0.000	0.0
		so,e	12.479	19.3

TABLE 4.16

Intra-Rater Reliability Variance Estimates for Rater 3

Posture	Task Type	Source of Variation	Estimated Variance Component	% Total Variance
Neck	Neutral	subjects (s)	185.590	90.5
		occasions (o)	7.995	3.9
		so,e	11.511	5.6
	Flexion	subjects (s)	239.895	92.7
		occasions (o)	9.866	3.8
		so,e	9.156	3.5
	Extension	subjects (s)	22.271	98.7
		occasions (o)	0.000	0.0
		so,e	0.299	1.3
	Missing	subjects (s)	0.012	6.9
		occasions (o)	0.034	19.5
		so,e	0.128	73.6
Shoulder	Neutral	subjects (s)	60.495	91.4
		occasions (o)	0.365	0.6
		so,e	5.297	8.0
	Flexion	subjects (s)	49.404	84.9
		occasions (o)	0.000	0.0
		so,e	8.781	15.1
	Extension	subjects (s)	5.102	62.2
		occasions (o)	0.240	2.9
		so,e	2.860	34.9
	Missing	subjects (s)	0.477	70.0
		occasions (o)	0.048	7.1
		so,e	0.156	22.9
Wrist	Neutral	subjects (s)	155.208	77.3
		occasions (o)	18.348	9.1
		so,e	27.221	13.6
	Flexion	subjects (s)	3.541	98.2
		occasions (o)	0.000	0.0
		so,e	0.064	1.8
	Extension	subjects (s)	40.742	69.4
		occasions (o)	3.380	5.8
		so,e	14.559	24.8
	Missing	subjects (s)	110.031	86.2
		occasions (o)	4.248	3.3
		so,e	13.413	10.5

TABLE 4.17

Intra-Rater Reliability Variance Estimates for Rater 4

Posture	Task Type	Source of Variation	Estimated Variance Component	% Total Variance
Neck	Neutral	subjects (s)	37.565	17.5
		occasions (o)	0.000	0.0
		so,e	177.678	82.5
	Flexion	subjects (s)	171.225	63.1
		occasions (o)	0.000	0.0
		so,e	99.984	36.9
	Extension	subjects (s)	60.605	92.2
		occasions (o)	0.249	0.4
		so,e	4.882	7.4
	Missing	subjects (s)	105.832	86.2
		occasions (o)	0.000	0.0
		so,e	16.925	13.8
Shoulder	Neutral	subjects (s)	146.610	67.3
		occasions (o)	0.000	0.0
		so,e	71.136	32.7
	Flexion	subjects (s)	108.229	80.5
		occasions (o)	0.000	0.0
		so,e	26.163	19.5
	Extension	subjects (s)	49.912	99.2
		occasions (o)	0.000	0.0
		so,e	0.384	0.8
	Missing	subjects (s)	4.006	17.6
		occasions (o)	0.000	0.0
		so,e	18.792	82.4
Wrist	Neutral	subjects (s)	123.814	59.8
		occasions (o)	0.000	0.0
		so,e	83.380	40.2
	Flexion	subjects (s)	3.790	17.7
		occasions (o)	1.981	9.3
		so,e	15.612	73.0
	Extension	subjects (s)	29.614	36.2
		occasions (o)	0.000	0.0
		so,e	52.121	63.8
	Missing	subjects (s)	274.543	80.2
		occasions (o)	0.000	0.0
		so,e	67.977	19.8

TABLE 4.18

Intra-Rater Reliability Variance Estimates for All Raters Combined

Posture	Task Type	Source of Variation	Estimated Variance Component	% Total Variance
Neck	Neutral	subjects (s)	254.796	82.5
		occasions (o)	8.789	2.8
		so,e	45.446	14.7
	Flexion	subjects (s)	238.373	82.1
		occasions (o)	12.113	4.2
		so,e	39.747	13.7
	Extension	subjects (s)	10.142	81.3
		occasions (o)	0.000	0.0
		so,e	2.339	18.7
	Missing	subjects (s)	3.382	58.6
		occasions (o)	0.000	0.0
		so,e	2.389	41.4
Shoulder	Neutral	subjects (s)	162.393	94.8
		occasions (o)	0.000	0.0
		so,e	8.916	5.2
	Flexion	subjects (s)	72.892	87.9
		occasions (o)	0.000	0.0
		so,e	9.999	12.1
	Extension	subjects (s)	42.623	92.8
		occasions (o)	0.000	0.0
		so,e	3.302	7.2
	Missing	subjects (s)	2.732	89.4
		occasions (o)	0.000	0.0
		so,e	0.323	10.6
Wrist	Neutral	subjects (s)	84.118	52.8
		occasions (o)	1.661	1.0
		so,e	73.613	46.2
	Flexion	subjects (s)	5.111	52.8
		occasions (o)	1.352	14.0
		so,e	3.209	33.2
	Extension	subjects (s)	17.628	25.4
		occasions (o)	2.585	3.7
		so,e	49.119	70.9
	Missing	subjects (s)	128.685	82.6
		occasions (o)	0.828	0.5
		so,e	26.245	16.9

Intra-rater reliability coefficients computed using G-theory for the neck, shoulder, and wrist posture categories of Raters 1-4 across two occasions are outlined in Table 4.19. Table 4.19 also contains intra-rater reliability coefficients for each posture category when data from all four raters was combined. Again, each rater re-analyzed 10 different subjects of the original 39 subjects on Occasion 2. Rater 2 analyzed nine subjects since there were only 39 subjects to choose from for re-analysis. When reliability coefficients were calculated across all four raters, the shoulder had the highest intra-rater reliability coefficients while the wrist had the lowest intra-rater reliability coefficients.

Intra-rater reliability coefficients using the Pearson Product Moment correlation coefficient are provided in Table 4.20. Pearson correlation coefficients were presented to foster comparability to other studies; however, Generalizability Theory provides a more sophisticated analysis of reliability and provides insight into the sources of variation. The Pearson correlation coefficients should be treated with caution, since the sample size for each rater was <15. Pearson's correlation coefficient is not recommended for sample sizes <15 where there are two ratings (Walter et al., 1998). For sample sizes <15, Pearson's correlation can overestimate the correlation, therefore an intraclass correlation coefficient is recommended (Walter et al., 1998). In the current study, the Pearson correlation coefficients were computed based on 9-10 samples and therefore may be overestimating the correlation.

TABLE 4.19

Intra-Rater Reliability Results Using Generalizability Theory

		G-Coefficients			
	Rater	**Neutral**	**Flexion or Mild Flexion/Abduction**	**Extension or Severe Flexion/Abduction**	**Missing Data**
Neck	**1**	0.89	0.90	0.58	0.80
	2	0.98	0.99	0.82	0.68
	3	0.95	0.96	0.99	0.13
	4	0.30	0.77	0.96	0.93
	All Raters	**0.90**	**0.90**	**0.90**	**0.74**
Shoulder	**1**	0.98	0.99	0.96	0.96
	2	0.97	0.94	0.99	0.92
	3	0.96	0.92	0.77	0.82
	4	0.80	0.89	1.00	0.30
	All Raters	**0.97**	**0.94**	**0.96**	**0.94**
Wrist	**1**	0.84	0.85	0.77	0.92
	2	0.79	0.94	0.58	0.89
	3	0.87	0.99	0.82	0.93
	3	0.75	0.30	0.53	0.89
	All Raters	**0.69**	**0.69**	**0.41**	**0.90**

Intra-rater reliability coefficients for the neck posture categories ranged from 0.13-0.99 (Table 4.19). When calculated across all four raters, neutral neck posture, neck flexion greater than 45°, and neck extension greater than 20° all had the same and highest reliability coefficient (0.90). Intra-rater reliability coefficients for the shoulder posture categories ranged from 0.30-1.00 (Table 4.19). When calculated across all four raters, shoulder flexion/abduction from 60°-90° had the highest reliability coefficient with shoulder posture greater than 90° only lower by 0.01 (Table 4.19). When calculated across all four raters, the missing data category had the lowest intra-rater reliability coefficient for both the neck and shoulder. Intra-rater reliability coefficients for the wrist posture categories ranged from 0.53-0.99 (Table 4.19). When calculated across all four raters, the missing data category had the highest reliability coefficient for the wrist while wrist extension greater than 30° had the lowest reliability coefficient.

Reliability varied across raters and body parts (Figures 4.13-4.16). As a reminder, each rater analyzed 9-10 different subjects, therefore there may be factors specifically related to the subjects each rater analyzed that may have influenced reliability. For example, lower variance in a specific posture could have lead to lower reliability coefficients or specific subjects may have been more difficult to analyze due inadequate camera angles or the nature of the task being performed. While recognizing these possible discrepancies between subjects, some general comparisons will be made. For postures of the neck and shoulder (Figures 4.13-4.14), Rater 2 had the highest overall reliability (across all four posture categories of the neck and shoulder), while Rater 4 had the lowest overall reliability. For postures of the neck (Figure 4.13), Raters 1, 2, and 3 had similar reliability across all four posture categories, while Rater 4 had much different ratings for neutral neck posture and the missing data category. For postures of the shoulder (Figure 4.14), Raters 1, 2, and 3 had similar reliability across all four posture categories, while Rater 4 had a much different rating for the missing data category. For posture of the wrist (Figure 4.15), Rater 4 had the highest overall reliability (across all four posture categories, while Rater 3 had the lowest overall reliability. Reliability coefficients for the wrist varied more greatly across posture categories than for the neck and shoulder.

FIGURE 4.13

Comparison of Rater Reliability Coefficients for the Neck

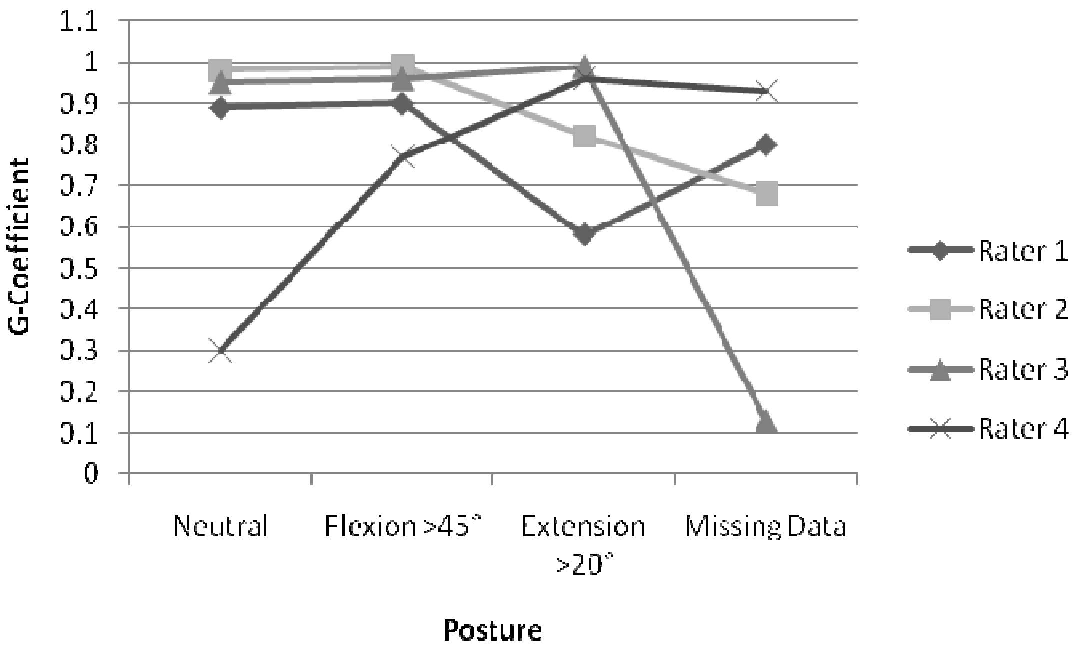


FIGURE 4.14

Comparison of Rater Reliability Coefficients for the Shoulder

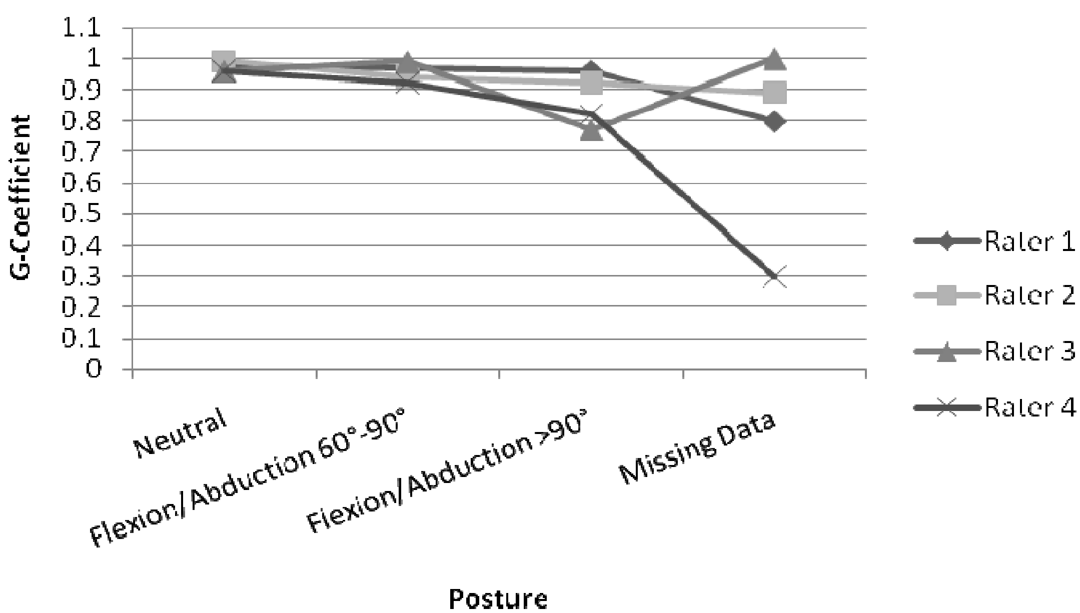
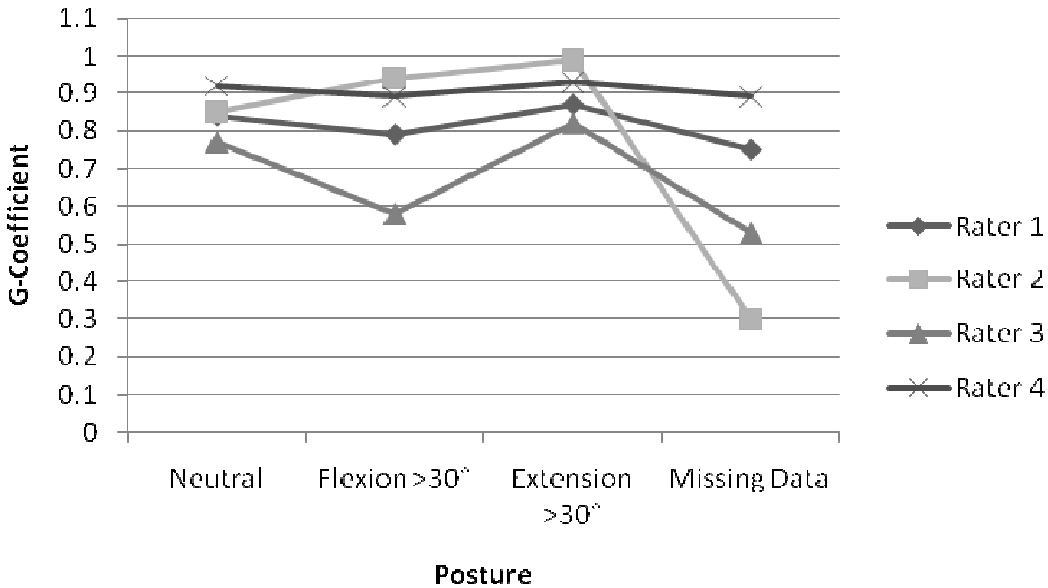


FIGURE 4.15

Comparison of Rater Reliability Coefficients for the Wrist



Pearson product moment correlation coefficients are also presented as a classical test of reliability (Table 4.20). These coefficients represent intra-rater reliability for each rater and the specific subjects each rater analyzed across both occasions. Pearson correlations are either 1.00 if there is an increasing linear relationship, -1.00 if there is a decreasing linear relationship, or some value between 1.00 and -1.00. The correlation values indicate the degree of linear dependence between two variables. While the Pearson coefficients are presented to foster comparability to other studies, computation of reliability coefficients using ANOVA provides a more sophisticated analysis of reliability and provides insight into the sources of variance. In addition, as explained previously, Pearson correlation coefficients are not recommended for sample sizes <15 and therefore the results should be interpreted cautiously.

TABLE 4.20

Intra-Rater Reliability Results Using Pearson Product Moment Correlation

		r-Coefficients			
	Rater	**Neutral**	**Flexion or Mild Flexion/Abduction**	**Extension or Severe Flexion/Abduction**	**Missing Data**
Neck	**1**	0.89	0.88	0.52*	0.64
	2	0.96	0.99	0.82	0.49*
	3	0.94	0.96	0.99	0.28
	4	0.13*	0.73	0.99	0.88
	All Raters	**0.76**	**0.82**	**0.92**	**0.89**
Shoulder	**1**	0.98	0.99	0.98	0.94
	2	0.94	0.89	1.00	0.92
	3	0.94	0.91	0.67	0.97
	4	0.70	0.81	1.00	0.27*
	All Raters	**0.87**	**0.89**	**0.97**	**0.45**
Wrist	**1**	0.77	0.82	0.76	0.87
	2	0.74	0.97	0.53*	0.83
	3	0.87	0.98	0.75	0.89
	4	0.58*	0.26*	0.35*	0.86
	All Raters	**0.74**	**0.86**	**0.58**	**0.83**

*Indicates significant difference based on an $\alpha = 0.05$.

In regards to all raters, eight correlation coefficients were not significant at the 0.05 level. Significance means there is evidence to reject the null hypothesis ($r = 0$) or that there is a statistically significant relationship between the specific pair of raters evaluated. Rater 1 had one non-significant correlation for neck extension >20°. Rater 2 had two non-significant correlations for neck missing data and wrist extension >30°. Rater 3 had no non-significant correlations. Rater 4 had the highest number (five) of non-significant correlations: neutral neck posture, shoulder missing data, neutral wrist posture, wrist flexion >30°, and wrist extension >30°.

Review of the results for the 39 analyses performed by the four raters and rater logs documented for every analysis were used to determine factors that contributed to rater disagreement. Factors differed by body part, but in general, raters contributed the following factors to difficulty in estimating posture: distances estimated at greater than 10 feet between the subject and the camera(s), camera views that were not perpendicular to the plane of joint motion, camera views blocked by equipment or personnel, use of personal protective equipment, and tasks requiring workers to reach or lean into appliances. In many cases, the raters noted these factors as part of their decision criteria on whether to rate the posture or assign the missing data category.

In an attempt to perform a more detailed investigation into the causes of disagreement, ratings by each rater for all posture categories for the neck, shoulder, and wrist were compared. For each posture across the 39 analyses, the mean percent time as rated by each rater was calculated along with standard deviation. An average of the 39 standard deviations for each posture category was then calculated. Standard deviations that fell above the mean were used to target specific subjects. Based on this approach, certain body parts, postures, and subjects had higher standard deviations due to differences between the four raters. The majority of standard deviations greater than the mean were found in the wrist, followed by the neck, and then the shoulder. This finding is not apparent from the reliability coefficients, since the neck had the lowest reliability across all posture categories.

Upon further analysis of the wrist, the majority of standard deviations greater than the average specific to the wrist were in the missing data category, followed by extension >30° category, the neutral posture category, and lastly the flexion >30° category. Further analysis of the missing data category revealed that Rater 4 assigned the missing data category 11% more than the average of the four raters and that Rater 3 assigned the missing data category 9% less than the average of the four raters. Upon review of the rater's logs for the subjects associated with the data points in the missing data category, Rater 4 documented inadequate camera views and obstruction of the wrist more often than the other raters. Rater 4 also documented that she felt hesitant to choose between assigning a posture versus the missing data category based on the camera views available. Of the four raters, Rater 4 was the least experienced in rating postures using MVTA. Other factors noted by the raters for difficulty in estimating wrist posture included far camera distances, poor lighting, and heavy gloves. Based on review of the rater logs, Raters 1, 2, and 3 seemed more likely to assign a posture than missing data even when the camera angles were less than perfect or some obstruction was present.

Upon further analysis of the neck, the majority of the standard deviations above the average occurred in the flexion >45° category, followed by the missing data category, the neutral category, and finally the extension >20° category. Further analysis of the flexion >45° category revealed that Rater 3 assigned the flexion >45° category 9% more than the average of the four raters and that Rater 1 assigned the flexion >45° category 8% less than the average of the four raters. Of the four raters, Rater 1 was the most experienced in rating postures using MVTA. Upon review of the rater logs for the subjects associated with the data points in the flexion >45° category, Rater 3 documented inadequate camera views, particularly those posterior to the subject, more often than the other raters.

In addition to evaluating the specific data points by body part and posture, standard deviations greater than the average were also evaluated by subject. Five subjects accounted for a large portion (1/3) of the standard deviations higher than the mean. All five subjects had performed cyclic tasks. Two of the five subjects accounted for nearly half of this portion. Most of the differences in raters for these subjects were found in the missing data category. Upon reviewing the two subjects and the rater logs, raters indicated difficulty estimating postures of the neck, shoulder, and wrist because the subjects had to bend over inside the appliances to perform their tasks. The need to bend over to access inside an appliance was limited to these two subjects. Rater 4 consistently assigned the missing data category for a higher percentage of time than the other raters and typically assigned lower percentages of time to the other posture categories. Rater 4 was the least inexperienced of the raters in analyzing postures using MVTA.

DISCUSSION

While the present study estimated both inter- and intra-rater reliability coefficients of posture observations of the upper limb using a video analysis tool, the discussion will not focus on comparing reliability coefficients to previous studies. In most cases, comparisons are difficult due to different body parts and postures assessed difference in exposure to the postures of interest, and differences in the statistical methods used. Instead, the discussion, using the results of the present study, will focus on statistical methods used to evaluate reliability, evaluation of factors related to decreased reliability, rater reliability as it relates to cyclic and non-cyclic tasks, techniques to help increase rater reliability, and the application of reliability in ergonomics research.

Statistical methods used to assess reliability of posture variables often include percentage of agreement, Cohen's Kappa statistic, and various forms of the intraclass correlation coefficient. In addition, the Pearson Product Moment correlation coefficient (Dartt et al., 2009) and Spearman Rank correlation coefficient (Lowe, 2004(a); Lowe, 2004(b)) are used as measures of reliability depending on whether the data are rank or interval in nature. Percentage of agreement and the kappa statistic can be used to evaluate reliability for categorical data. Bao et al. (2009) reported percentage of agreement statistics for a study evaluating inter-rater reliability of four raters using three different posture classification strategies to rate four different job tasks. Many researchers have used the kappa statistic in posture reliability studies (Lowe, 2004(a); Lowe, 2004(b); Ketola et al., 2001; Paquet et al., 2001; Burt and Punnett, 1999; Pan et al., 1999; de Bruijn et al., 1998) to expand the capability of simple percentage of agreement. Cohen's Kappa can be used to assess inter-rater reliability when there are only two raters; therefore, this statistic is not applicable to the current study. In addition, the kappa statistic is influenced by posture distributions and may therefore over or underestimate reliability.

The intraclass correlation coefficient (ICC) and the ANOVA have also been used in posture reliability studies (Bao et al., 2009; Dartt et al., 2009; Lowe, 2004(a); Burt and Punnett, 1999; Genaidy et al., 1993; Stetson et al., 1991). The ICC is used to evaluate continuous data and is conceptualized as the ratio of between-groups variance to total variance. One concern with the ICC is that provided there is the same magnitude of variations among raters' estimates, a posture with smaller variation will result in lower ICCs compared with postures with larger variations (Bao et al., 2009). In the current study, this same phenomenon tended to occur. Neck extension greater than 20º, neck missing data, and shoulder missing data, which all had smaller variance, had lower reliability coefficients (0.76, 0.55, and 0.73, respectively) than the other postures (Table 4.12 and Figures 4.7-4.9). However, one should not assume that lower variance will always lead to lower reliability coefficients. For example, neck extension greater than 20° had very low overall variance compared to the other body parts and postures, with missing data for the shoulder as the only exception (Table 4.9). The phenomenon of lower reliability because of lower variance did not hold true when evaluating non-cyclic tasks for neck extension greater than 20°. The reliability coefficient was high (0.92), even though the variance was low relative to the other postures evaluated (Tables 4.9 and 4.12). This demonstrates the importance of evaluating the specific variance components rather than evaluating variance in general.

Some additional statistical methods not mentioned previously used to investigate rater agreement have included: paired t-tests of mean differences (Burt and Punnett, 1999), standard deviation as a measure of precision (Bao et al., 2009), and an application of the intraclass correlation coefficient known as Generalizability Theory (Dartt et al., 2009). Based on review of the various statistical methodologies, the present study utilized Generalizability theory, a form of ICC, to estimate rater reliability. Generalizability theory extends the use of reliability measures, such as percentage of agreement, kappa, and correlation, by considering multiple sources of error simultaneously and allows a more accurate assessment of the measurement situation (VanLeeuwen, 1997). Therefore, one can estimate the magnitude of each source of error separately in a single analysis and use this information to optimize the reliability of the measurement (Shavelson and Webb, 1991). Generalizability theory distinguishes between relative and absolute decisions, in the same manner as computing classical intraclass correlation coefficients (Shavelson and Webb, 1991). Absolute agreement measures whether raters assign the same absolute score and is used when systematic variability due to raters is relevant. Relative agreement, or consistency, considers raters consistent as long as their relative ratings are similar. In addition to providing an intraclass correlation coefficient, or Generalizability coefficient, Generalizability theory utilizes the variance components estimated to design a more efficient and effective measurement procedure to be used in the future. For example, a researcher can estimate the number of raters necessary to achieve a particular level of reliability based on pilot reliability assessment.

Findings on rater reliability may differ greatly depending on the statistic used. This discrepancy had been reported in other research that has investigated rater reliability for observation of posture (Bao et al., 2009; Burt and Punnett, 1999). Burt and Punnett (1999) found substantial differences between percentage of agreement and the kappa statistic. For example, percentage of agreement (96%) was high for left wrist extensions, while the kappa statistic (0.55) was considered low. The authors (Burt and Punnett, 1999) contributed this difference to the fact that percentage of agreement does not account for chance agreement and that chance agreement is high for a rarely observed event. In addition, the kappa statistic can vary depending on the prevalence of the posture being observed and will ultimately be lower for rarely observed postures. Boa et al. (2009) noted similar issues when using the ICC. Similar to the kappa statistic, the ICC is dependent on the overall variation in the factors being studied, making these coefficients difficult to compare from study to study. Bao et al. (2009) noted high percentage of agreement for trunk flexion and extension (91% overall) and high precision in between-rater standard deviation, but the ICC (0.49) was considered moderate due to the smaller variations in trunk flexion and extension.

The present study found similar discrepancies when comparing results found using Generalizability Theory (a form of the ICC) and the Pearson correlation coefficient. In most cases, the reliability coefficients obtained using Generalizability Theory and the Pearson correlation coefficients were similar (Tables 4.12 and 4.13). In all cases, with the exception of neck missing data for cyclic tasks, neck missing data for non-cyclic tasks, and wrist extension >30° for non-cyclic tasks, the Pearson correlation coefficients were lower than the G-coefficients. Postures with a 0.10 difference or greater between the G-coefficients and Pearson coefficients were: neck – neutral for all tasks and cyclic tasks and extension >20° for all tasks and cyclic tasks; shoulder – flexion/abduction 60°-90° for all tasks and cyclic tasks, flexion/abduction >90° for cyclic tasks, and missing data for all tasks. The three Pearson correlation coefficients that were higher than the G-coefficients could be explained by the lower variance in the neck missing data category and the wrist extension >30° category. Lower variance in a specific posture category can cause a lower G-coefficient. Sample size could explain why most of the Pearson correlation coefficients were higher than the G-coefficients, particularly for the non-cyclic tasks since sample size was <10. However, the largest discrepancies were found in posture categories that had adequate sample size (29-39) and higher variances, relative to the total variance. This difference could be attributable to the structure of the statistic. Calculation of the G-coefficient (ICC) centers and scales the data using a pooled mean and standard deviation, whereas the Pearson correlation centers and scales each variable by its own mean and standard deviation. In addition, Pearson's could not simultaneously compare more than two of the raters, so separate analyses had to be completed. On the other hand, the ANOVA used to compute the G-coefficients could simultaneously analyze all four raters.

Inter-Rater Reliability

Reliability of posture observations between different raters can be a complicated issue, particularly when those raters are analyzing postures to be later used to investigate a causal relationship between exposure to awkward postures and health outcomes. The current study was part of a larger five-year prospective cohort study investigating the relationship between ergonomic risk factors and musculoskeletal outcome. A previous reliability study (Dartt et al., 2009) evaluated two of the raters from the larger study who evaluated upper limb postures of workers performing manufacturing tasks. In addition, the previous study (Dartt et al., 2009) attempted to expand previous reliability research by evaluating a computer-based exposure assessment tool (MVTA) and continuous video footage. Most previous rater reliability research has evaluated direct observation or photographs (Ketola et al., 2001; Paquet et al., 2001; Burt and Punnett, 1999; Pan et al., 1999; Stetson et al., 1991; Keyserling, 1986) and evaluation of reliability for video-based observation methods have been limited to frame-rate methods (Bao et al., 2009). Posture assessment through observation, direct or via video, is common in ergonomics research studies because it allows for an efficient exposure assessment of large populations of individuals without the setup time or interference of direct reading instrumentation.

The present study determined that inter-rater reliability was fair to excellent for observational measurements of neck, shoulder, and wrist postures among workers performing appliance manufacturing tasks (Table 4.12). Analysis of cyclic tasks resulted in inter-rater reliability coefficients that were also fair to excellent for observational measurements of neck, shoulder, and wrist postures, but lower than when evaluating all tasks combined (Table 4.12). Analysis of non-cyclic tasks resulted in inter-rater reliability coefficients that were also fair to excellent for observational estimates of neck, shoulder, and wrist postures, but poor for one posture category, neck missing data (Table 4.12).

Evaluation of reliability for cyclic and non-cyclic tasks demonstrated higher overall reliability coefficients for the non-cyclic tasks for the neck, shoulder, and wrist (Tables 4.12 and 4.13). In general, the non-cyclic tasks had less overall variation in postures as compared to the cyclic tasks, as determined by the mean and range of postures for each posture category (Tables 4.6-4.8). Therefore, the conclusion that the non-cyclic tasks had larger reliability coefficients based on larger variation in postures does not hold true. Examination of the variance components from the ANOVA (Tables 4.9-4.11) demonstrated higher subject variation for specific body parts and postures within the non-cyclic tasks. A larger variation within the subjects component for the non-cyclic tasks can account for the higher reliability coefficient for certain body parts and postures. Based on the variance components, the cyclic tasks (n=29) were more similar between subjects in regards to the postures analyzed than the non-cyclic tasks (n=10). Since non-cyclic tasks tend to be more variable in nature than cyclic tasks, this finding is not unreasonable.

Previous inter-rater reliability research for observation of posture (Bao et al., 2009; Ketola et al., 2001; Burt and Punnett, 1999; Stetson et al., 1991) has covered a wide range of jobs and tasks; however, there has not been a specific breakdown of analyses by whether the task was cyclic or non-cyclic. Ketola et al. (2001) evaluated rater reliability of two raters assessing various postures of five cyclic job tasks in the meat packing industry, Burt and Punnett (1999) evaluated rater reliability of two raters assessing various postures of 75 jobs in a stamping plant, and Bao et al. (2009) evaluated rater reliability of seven raters assessing various postures of four jobs from four different job locations. Based on the description of the studies, both Burt and Punnet (1999) and Bao et al. (2009) evaluated non-cyclic tasks or non-cyclic portions of the jobs evaluated. Based on the results of the present study, the generalizability of inter-rater reliability coefficients can be dependent on whether the tasks evaluated are cyclic or non-cyclic. Knowledge of the work environment and the nature of the tasks can be critical when analysis of risk factors involves rater judgment or estimation. Based on the results of the present study, a work environment containing only cyclic tasks similar in posture percentages and subject variance may warrant more raters than a work environment containing primarily non-cyclic tasks with large subject variation to achieve the same level of reliability.

Previous reliability studies that have assessed posture ratings have provided factors and guidance as to the causes of low reliability or disagreement between raters. These have included: the postures being investigated, posture variation, rater training and experience, quality of the video image, and posture definitions (Bao et al., 2009; Paquet et al., 2001; Burt and Punnett, 1999; Pan et al., 1999; Stetson et al., 1991; Keyserling, 1986). In a previous study that evaluated training effects on reliability, the researchers concluded that rater reliability was proportional to the degree of difficulty in rating the video-taped task (Denis et al, 2002). The authors also found that reliability decreased as raters expressed less confidence in their answers, thus leading to the potential for mental fatigue. While the degree of difficulty in rating postures or mental fatigue was not measured in this study, the missing data category could be an indirect measure of the degree of difficulty in rating postures from video. One could conclude that a higher percentage of time assigned to the missing data category could be related to more times a rater had to make a decision between assigning a posture or assigning missing data. This constant decision making could have led to increased mental fatigue and potentially decreased reliability.

Based on the results of the present study, certain posture categories were easier to observe than others. The four raters logged a higher percent of time in the missing data category for the wrist than for the neck and shoulder. This indicates that the wrist, a smaller body part than the neck and shoulder, caused difficultly among the raters in analyzing and assigning the specified posture categories. Previous research studies have also noted this finding (Bao et al., 2009; Burt and Punnett, 1999; Stetson et al., 1991; Keyserling, 1986). Bao et al. (2009) found that postures of larger body parts, trunk and arms, were easier to observe than smaller body parts, such as the wrists. This was reflected in the larger ICCs and the larger amount of missing data for the smaller body parts. While the present study had more missing data for the wrist, the reliability coefficients did not reflect this difficulty (Table 4.12). Reliability coefficients for the wrist were greater than 0.7 for all posture categories and the reliability coefficients for the wrist missing data category were higher than missing data for the neck or shoulder. Exceptions similar to this were documented in previous research. Bao et al. (2009) found that the ICC for trunk flexion and extension was low (0.18) even though the trunk is considered a large body part. The authors (Bao et al., 2009) contributed the low ICC to the smaller amount of posture variation found in the trunk. This corresponds to the present study. Neck and shoulder reliability coefficients for the missing data category were likely lower than the wrist missing data category due to the smaller amounts of variation (Tables 4.9-4.11).

The authors of the present study attributed the lower reliability coefficients associated with the missing data categories to the rater's judgment of whether to confidently estimate a posture or assign it as missing data. It was not unusual for one rater to assign missing data while the other rater estimated a posture, based on review of the analyses. It was likely that the threshold raters used to make this determination was based on time constraints, pressure to estimate a posture rather than have incomplete or missing data, and visual and/or mental fatigue associated with several hours of data analysis. While the Bao et al. (2009) study used missing data as a category that rater's could choose from, reliability coefficients were not reported. Raters were encouraged to record postures whenever possible. However, in many cases, two cameras could not capture the views necessary to estimate posture of the neck, shoulder, or wrist. It was clear to the raters to record missing data when the entire body part was not visible. However, decisions on whether one can or cannot estimate neck, shoulder, or wrist postures from poor camera views were more subjective. The specific amount of deviation in degrees from a perpendicular view of the neck where a rater was allowed or not allowed to estimate posture was not specified in the study. The authors of the present study feel that the use of missing data variables and associated reliability implications will become increasingly important for studies that attempt to quantify exposure through video observation.

Based on the review of the 39 subjects and rater logs, there were other factors related to lower reliability within the neck posture categories, in addition to smaller variation. In general, raters documented the following as factors making it difficult to judge neck posture: camera angles not perpendicular to the subject's body, particularly camera views posterior to the subject, the neck partially blocked by equipment, and work requiring the worker to bend over. Analysis of specific subjects and rater logs based on the standard deviation of the percent time assigned to all postures as described in the results section, demonstrated discrepancies in the assignment of the neck posture categories. Since, as a whole, the posture categories were dependent, meaning if a rater assigned the neutral posture, he/she could not assign missing data, consistent discrepancies in rater judgments can lead to large differences. For example, Rater 3 assigned the neck flexion >45 category 9% more than the average and Rater 1 assigned the same category 8% less than the average. This resulted in a large difference between Rater 1 and 3. Based on this finding, Raters 1 and 3 were applying the posture definitions and rules differently. Rater 3 documented difficulty in estimating neck posture due to the camera angles and posterior location to the subjects. Rater 1, the more experienced rater, did not note this difficulty for the same subjects.

In regards to the neck missing data category, Rater 4 assigned missing data 16% more than the average while Rater 1 assigned missing data 8 % less than the average. As described above, it was apparent that Rater 1 and Rater 4 were applying the posture definitions and rules differently. Rater 4 was the least experienced of the raters, which could contribute to this discrepancy. Evaluation of rater pairs using the Pearson correlation coefficient also demonstrated discrepancies between raters for neck postures (Figure 4.10). The missing data category and extension >20° had the lowest correlation coefficients for the neck for all rater pairs demonstrating consistent disagreement in ratings for these two categories across all rater pairs. Rater pair (1,2) had the highest overall correlation coefficient for neck postures while rater pair (1,4) had the lowest overall correlation coefficient for neck postures and the only significantly different correlation coefficient in the neck missing data category. Raters 1 and 2 had the most experience and tenure analyzing postures, while Rater 4 had the least experience and tenure analyzing postures, but no different than Rater 3. This indicates that there may have been an experience effect; however, this is not completely understood since Rater 3 had similar experience, but higher correlation coefficients.

Reliability coefficients within the wrist posture categories were considered fair to excellent (Table 4.12). However, based on the review of the 39 subjects and rater logs, the wrist had the highest number of standard deviations above the average. Based on this finding, specific subjects and their associated rater logs were reviewed to achieve a better understanding of rater disagreement. The majority of discrepancies were found in the missing data category and extension >30°category. Review revealed that Rater 4 assigned the missing data category 11% more than the average, while Rater 3 assigned missing data 9% less than the average. Based on the review of the rater logs, Rater 4 documented inadequate camera views and obstruction of the wrist more often than the other rates. This indicates that Rater 4 was more likely to assign missing data than the other raters for the same point in time due to the likelihood of using the missing data category rather than assigning one of the posture categories. Evaluation of rater pairs using the Pearson correlation coefficient demonstrated no significance difference for any of the rater pairs; however, the more experienced raters (Raters 1 and 2) had the highest correlation coefficient (Figure 4.12). In general, raters documented the following factors as making it difficult to judge wrist posture: camera distances estimated at greater than 10 feet, poor lighting, personal protective equipment such as heavy gloves, camera angles, and obstruction due to equipment.

Reliability coefficients within in the shoulder posture categories were considered fair to excellent (Table 4.12) and the shoulder had the highest overall reliability as compared with the neck and wrist postures. Evaluation of rater pairs using the Pearson correlation coefficient demonstrated similar results for all rater pairs. Based on review of the 39 subjects and rater logs, the raters documented much fewer instances where they had trouble judging the postures for the shoulder. The two most common factors were a posterior camera view and the arm blocked or partially blocked by equipment. Shoulder flexion/abduction greater than 90° had the highest reliability coefficient reported in the present study. This was the same as the findings in the previous reliability study (Dartt et al., 2009) and expected based on previous research with similar findings (Ketola et al., 2001; Keyserling, 1986).

Intra-Rater Reliability

Intra-rater reliability, while not as commonly reported in the ergonomics literature, also plays an important role in posture observation. While research studies should verify adequate inter-rater reliability, they should also verify adequate intra-rater reliability to ensure stability and reliability for rater's judgments over time. Ultimately, low intra-rater reliability could lead to low inter-rater reliability. The present study determined that intra-rater reliability was poor to excellent for observational measurements of neck, shoulder, and wrist postures among workers performing appliance manufacturing tasks (Table 4.19). Lower reliability coefficients reported in some of the posture categories suggest that raters may have changed their decision criteria from Occasion 1 to Occasion 2.

Intra-rater reliability coefficients varied by rater, the posture categories, and body parts analyzed. Rater 4 had the lowest overall reliability for the neck and shoulder, but had the highest reliability for the wrist. Evaluation using the Pearson correlation coefficient demonstrated that Rater 4 had significant differences in ratings from Occasion 1 to Occasion 2 for neutral neck posture, missing data for the shoulder, neutral wrist posture, wrist flexion >30°, and wrist extension >30° (Table 4.20). Further review of the analyses performed by Rater 4 revealed that five of the 10 subjects analyzed had greater than a 14% difference from Occasion 1 to Occasion 2 when assigning neutral neck posture. One particular subject had a 43% difference when assigning neutral neck posture with Occasion 2 significantly lower than Occasion 1. It was determined that from Occasion 1 to Occasion 2, that the time previously logged in the neutral category was logged in the flexion >45° category, indicating the rater modified their decision criteria for the neck from Occasion 1 to Occasion 2. Review of the other four subjects revealed that again, the rater modified their decision criteria from Occasion 1 to Occasion 2 assigning more time to the neutral category and less time to the flexion, extension, and missing categories.

These findings were similar to those found by Keyserling (1986). Keyserling (1986) attributed differences in shoulder posture estimations of two raters from video across occasions to positions of the shoulder when it was near the intersection of two postures (Keyserling, 1986). For example, if one is to estimate shoulder flexion greater than or less than 45°, it is easier to rate a posture at 60° than a posture that is 47°, or near the cut-off point. There was an inability of the analyst to consistently use the same boundary when rating adjacent postures. The present study supports this finding as discussed previously. Raters lacked some consistency in assigning the missing data category for the neck, shoulder, and wrist. It seemed that the raters assigned the missing data category more or less often across occasions, meaning that the boundary between assigning a posture or assigning missing data changed slightly over the two occasions.

Based on review of the rater logs for the posture categories demonstrating significant differences between Occasion 1 and Occasion 2, raters typically recorded inadequate camera angles and obstructed view of the body part for one of the analyses, but not the other. This again supports the notion that how stringently raters applied the decision criteria changed from one occasion to the next. This may imply an experience or learning curve effect, meaning as the raters analyzed more subjects and experienced different postures and work situations, they may have applied the posture estimation rules and decision criteria differently. This experience or learning effect was not quantified in this study. The rater logs attempted to capture factors not measured by the ANOVA that ultimately could have contributed to the residual variance; however, these logs were not systematic enough to include in the statistical analyses.

Suggestions for Improving Reliability in Posture Observation

Factors found in the present study that affected rater reliability were multiple and varied as described in the previous sections. These factors were related to the: 1) video viewing quality (inadequate camera angles, body parts blocked by equipment, poor lighting, personal protective equipment, camera distances estimated at >10 feet, work requiring the subjects to bend over), 2) nature of the task (variability within posture categories, time spent in the posture categories, the amount of variance between subjects, amount of missing data), and 3) factors related to the analyses (size of the body parts analyzed, postures analyzed, posture estimation guidelines, decision criteria, time constraints, pressure to assign a posture rather than missing data, experience of the raters, mental fatigue).

Previous reliability research that assessed posture ratings of the upper extremities (Bao et al., 2009; Burt and Punnett, 1999) listed unclear definitions of postures, inadequate rater training, difficulty observing slight body movements compared to gross body movements, variability of the posture parameters, camera positions, video quality, and complicated work postures as possible factors that accounted for disagreement between raters. While the present study had some inter- and intra-rater reliability coefficients below 0.50 for certain posture categories and body parts, reliability coefficients tended to be much higher than those found in previous studies. These differences between studies could be attributed to: different statistical analysis methods, angular differences in the posture categories, different variances, different posture estimation rules and decision criteria, and different rater training programs.

The authors of the present study recommend a systematic training program that consists of precise and detailed definitions of the posture categories, detailed posture estimation guidelines, detailed decision criteria, and extensive feedback by experienced ergonomists during the training period. The more detailed the posture estimation decision criteria, the fewer subjective judgments the raters will be required to make, ultimately increasing reliability. The detailed decision criteria used in the present study can be found in Appendix B. Training, as described in the methods section, should consist of several phases that includes demonstration by an experienced analyst, detailed review and feedback, and a quality control phase that continues throughout the data collection process.

Previous research (Burt and Punnett, 1999) found that precise estimates of joint deviation in degrees of excursion from neutral postures were more difficult than estimating postures using anatomical referencing. The authors of the present study recommend detailed definitions of the posture categories and posture estimation guidelines that include a combination of techniques to aid in posture estimations. Three techniques were used in the present study. The first technique consisted of training the raters to recognize specific degrees of excursion from neutral for the neck, shoulder, and wrist. The second technique included comparing known angles drawn on transparencies to anatomical positions seen on the computer monitor. The third technique involved referencing anatomical landmarks. For example, raters were trained to recognize 90° shoulder flexion/abduction, utilize a 90° reference angle, and to record greater than 90° shoulder flexion/abduction each time the elbow rose above the shoulder. In regards to the neck, raters were trained to identify >20° extension and >45° flexion. Based upon our experiences, determining the degree of neck flexion or extension was often difficult due to high body mass index and other factors such as long hair and clothing that obstructed the view of the neck. Because of these challenges, the authors defined neck flexion and extension using anatomical referencing. The base of the nostrils and tragus of the ear were used to aid in posture estimation (Norkin and White, 1987). The base of the nostrils and the tragus of the ear fall on an approximate parallel line from one another. Neck position was defined as a line drawn through the tragus of the ear and the base of the nostrils relative to the trunk. In regards to the wrist, raters were trained to identify >30˚ flexion and extension and could use known angles drawn on transparencies to assist in posture estimation. The metacarpals and forearm were utilized as anatomical references when assessing wrist posture. Based on the results obtained in the present study, the authors recommend using a combination of degree estimation techniques and anatomical referencing when determining postures from video recorded work tasks.

Since observation of postures from video is affected by video viewing quality (inadequate camera angles, body parts blocked by equipment, poor lighting, personal protective equipment, camera distances estimated at >10 feet, work requiring the subjects to bend over), a comprehensive knowledge of the work environment and pilot video capture is recommended before the start of data collection. Knowledge of the work area layout, equipment, structures, lighting, and personal protective equipment is essential in obtaining the best video quality possible. The use of two cameras aids in capturing views perpendicular to the body part of interest and is recommended. Since farther camera distances are needed to capture larger body movements such as the arms, trunk, and neck, it is recommended that at least one camera provide a closer view of the wrists when necessary. Pilot video capture of a representative number of tasks should be analyzed by at least two raters prior to data collection. These pilot analyses can assist with improving video collection, assist in formulating posture estimation guidelines and decision criteria, and can also be used as a preliminary evaluation of rater reliability.

In regards to the nature of the task (variability within posture categories, time spent in the posture categories, the amount of variance between subjects, amount of missing data), the tasks analyzed will stipulate posture variability and task variance. While it's ideal to evaluate as large a range of postures as possible, it is not always feasible. When evaluating hundreds of employees in a single facility, it is likely the majority of postures will be moderate with fewer at the extremes. Therefore, it is important to apply multiple statistical techniques that allow for further investigation into the reliability. Reporting of variance estimates in the ergonomics literature is recommended to foster comparisons between research studies.

Another important factor is the use of a missing data category. Few ergonomics studies (Bao et al., 2009; Dartt et al., 2009) have reported the use of a missing data category with assessing posture through observation. The use of a factor like this becomes increasingly important for video observation. When evaluating multiple body parts and posture categories from video, it is not probable that raters will have a view of the body parts 100% of the time, particularly for dynamic tasks that involved a lot of movement from the subjects. The missing data category eliminates the need for raters to "guess" at assigning postures when body parts are out of camera view, partially blocked, too far to estimate, etc. The authors recommended the use of a missing data category for observation of posture from video; however, video quality, posture estimation guidelines, and decision criteria should maximize the use of true posture categories and minimize the need for a missing data category since this category can have implications when assessing the relationship between awkward postures and musculoskeletal outcomes.

CONCLUSIONS

Ergonomic research studies that use raters to assess ergonomic risk factors, such as awkward postures, should evaluate inter- and intra-rater reliability as early as possible in the research process. If later reliability analyses reveal poor inter-rater reliability, conclusions made from the data collected by these raters could be flawed, thereby leading to potentially wrong associations or causations between risk factors and musculoskeletal outcomes. A causal effect between awkward postures and musculoskeletal effect has yet to be realized. Based on this, one could argue that there is no causal effect between awkward postures experienced during work and musculoskeletal outcome. This argument has been used to repeal regulation of ergonomic risk factors. The authors of the present study argue that the lack of causal effect could reside in the exposure assessment techniques and tools used to evaluate awkward postures in addition to the multi-factorial nature of most musculoskeletal outcomes.

The present study attempted to further evaluate inter- and intra-rater reliability, particularly the causes for lower reliability when assessing posture from video. While reliability coefficients were acceptable overall for this study, previous research as reported much lower reliability, in general. Generalization to other occupational tasks outside of manufacturing is limited due to the narrow scope of this project. The present study evaluated postures of the upper extremities in cyclic and non-cyclic work tasks, therefore generalization is limited to the body parts and postures analyzed. The present study also employed a statistic, Generalizability Theory, commonly used in the social sciences to evaluate observations of raters. While this statistic was used in its primary form, as an intraclass correlation coefficient, its power resides in the ability to predict the number of raters and occasions necessary to achieve a certain level of reliability. The prediction capability of Generalizability Theory assists with early evaluation of rater reliability.

Rater reliability should always be calculated and reported for ergonomics research that uses raters to assess a risk factor, such as awkward posture. An appropriate minimum acceptable level of reliability should be identified and rater reliability should be initially tested during pilot data collection. The pilot data evaluation and prediction capability of Generalizaiblity Theory can assist in determining the number of raters and occasions necessary to achieve the desired level of reliability. Rater reliability should then be formally assessed during the full data collection period. Final reports and peer-reviewed literature should contain the outcomes of the reliability study, the number of raters, the type of risk factor evaluated by the raters, the tool used to quantify the risk factor, the number of subjects/tasks evaluated, the training protocol, the statistical methods used, how disagreements in reliability were resolved, and how the reader can obtain more information on the reliability assessment.

Rater reliability is a critical component when quantifying awkward postures from video observation. While rater reliability does not ensure validity, when it is not properly established, the data and interpretations of the data cannot be considered valid. The authors of the present study recommend the following to improve rater reliability when assessing awkward posture from video and ultimately improve the quality of posture data obtained from video: a systematic training program, precise definition of postures, detailed posture estimation guidelines and decision criteria, extensive feedback during rater training, pilot video capture, at least a two camera setup, application of multiple statistical methods, the use of a missing data category, and early evaluation of rater reliability.

SECTION FIVE

STUDY III: INTER-METHOD RELIABILITY

ABSTRACT

Convergent validity is a critical component for ergonomics exposure assessment, particularly for research that attempts to establish a causal relationship between ergonomic risk factors and musculoskeletal outcomes. The purpose of this study was to determine the convergent validity of assessing upper arm and trunk postures of workers performing manufacturing tasks using a video observation technique and direct technical measurements by inclinometry. Assessment of left and right upper arm elevation and trunk inclination postures of 15 manufacturing workers performing tasks in three different work areas was conducted by video-based observational analysis synchronized with inclinometry. Generalizability theory, a form of the intraclass correlation coefficient, and the Pearson Product Moment correlation coefficient were used to evaluate the two methods. Correlation coefficients ranged from poor to excellent depending on the anatomical area and posture category evaluated. Further evaluation of the individual workers revealed possible sources of error that lead to decreased correlation. Based on the present study, video observation and inclinometry produced similar results. However, there were discrepancies depending on the anatomical area and posture categories evaluated.

INTRODUCTION

The purpose of this study was to evaluate the inter-method reliability of assessing upper limb postures of workers performing manufacturing tasks using a video-based observation technique and direct measurements. A reliable measure is one that measures something consistently. Validity is a broad term that encompasses several different forms. In regards to observational-based posture analysis measurement, internal and external validity should be demonstrated (Kilbom, 1994). Internal validity, the ability to measure what is intended, of observation methods has been estimated with reference to motion capture systems, electrogoniometers, and inclinometers (Kilbom, 1994). The majority of studies have focused on back postures with limited information on the upper extremities (Baty et al., 1986; Burdorf et al., 1992; de Looze et al., 1994; Keyserling, 1986). A few studies have validated neck postures with motion capture systems as compared to observations made from still pictures (Leskinen et al., 1997; Fransson-Hall et al., 1995).

For observation of posture, external validity is defined as the analysis method’s ability to distinguish physical exposure levels of posture associated with an increased risk in musculoskeletal disorders for a given anatomical area (Lowe, 2004a). Internal validity would then be the analysis method’s ability to measure what is intended or true. Most studies that have investigated the internal validity of observation of posture have been conducted in a laboratory setting. Examination of external validity requires an epidemiological investigation, whereas the examination of internal validity requires a kinesiological/biomechanical investigation.

A number of published validity studies have utilized electrogoniometers or inclinometers as the reference comparison method for observation of neck, shoulder, and wrist postures. However, one can argue that since these devices are not considered a 'true reference' method, comparisons of observation measures to those obtained through electrogoniometry or inclinometry should be considered inter-method reliability research. This is primarily due to the significant levels of measurement error found in these devices (Spielholz et al., 2001). Inter-method reliability measures the ability of 'different instruments which measure the same underlying exposure to yield similar results on the same subjects' (Armstrong et al., 1994).

Observational techniques have been historically used in much ergonomic research and remain to be extremely common in current ergonomic research studies. Recently, advances in technology have allowed for direct reading instruments to be used more extensively in ergonomics research. Cost, setup time, and memory capability are not as much of a factor as they once were. Since observation is historically and currently used and since direct reading instruments are being used more often , it is important to verify whether these exposure assessment methods are reliable when compared to one another. The lack of well-defined and adequately evaluated exposure assessment methods continues to be an issue in the field of ergonomics (Spielholz et al., 2001). Investigation into and understanding of the inter-method reliability between observation and direct reading instruments is extremely important when attempting to establish a causal relationship or even association between risk, such as awkward postures, and musculoskeletal outcomes.

The primary objective of the present study was to:

- Compare observational assessments of postures of the shoulders and trunk using a video-based observation tool to postures of the shoulder and trunk using inclinometry.

The hypotheses of the present study were:

1. There would be no statistically significant difference between measurements obtained from the two measurement tools.
2. Inter-method reliability coefficients comparing video observation versus direct technical measurements will be moderate (0.5-0.75).

METHODS

Research Design

Fifteen workers performing seven different tasks in a northern Colorado manufacturing facility were recruited for the present study. Postures of the shoulders and trunk (Table 5.1) were recorded simultaneously with video and inclinometry. Video recordings were time-synchronized with the inclinometers using a time stamp. Video recordings were obtained in a similar manner as in a previous study (Dartt et al., 2009). However, in the present study, only one camera was used to capture images of the shoulders and trunk instead of two cameras. Direct measures of shoulder and trunk postures were obtained using a Virtual Corset (MicroStrain, Inc; Williston, VT) inclinometer. Video footage and inclinometry were collected on each worker performing the same task during two different shifts. Upon completion of the data collection, subsequent video analyses was completed using Multimedia Video Task Analysis (MVTA) and subsequent inclinometer analysis was completed using LabView 8.6

software (National Instruments). Figure 5.1 provides a visual schematic of the research design. The simultaneously collected inclinometer and video analysis data were used to evaluate inter-method reliability.

TABLE 5.1

Postures of the Shoulders and Trunk Evaluated

Using Video Observation and Inclinometry

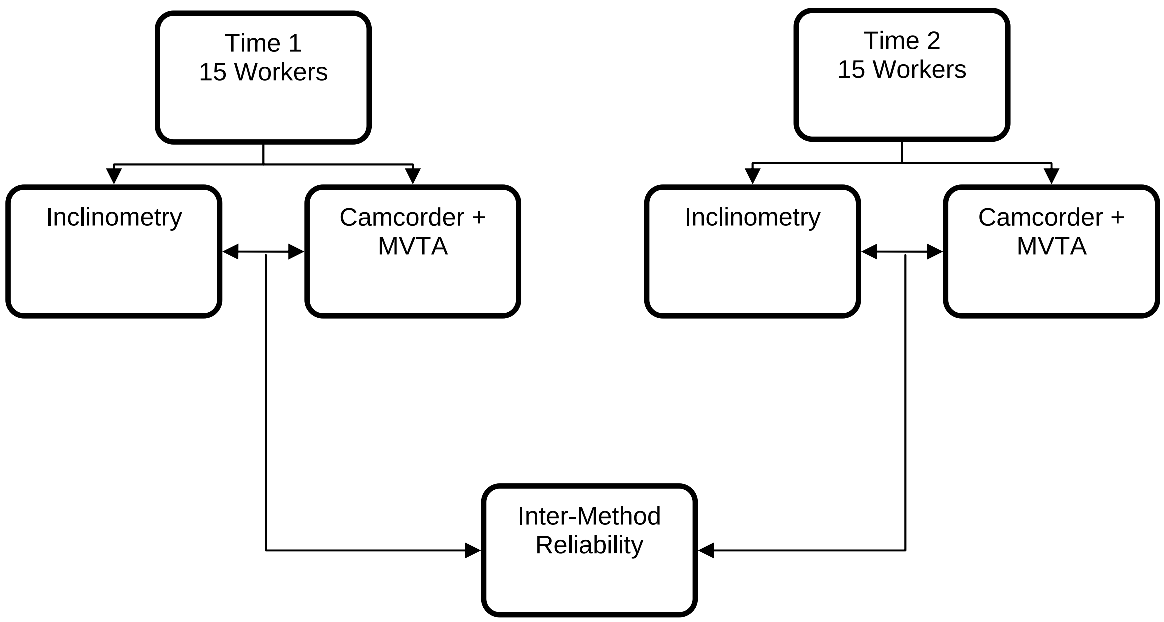
Posture Category	Shoulder (Flexion/Abduction)	Trunk (Flexion)
1	0°-45°	< 45°
2	45°-90°	≥ 45°
3	> 90°	-

Evaluation of posture for the present study was limited to the shoulders and trunk based on the limitations of the inclinometry device. Rationale for posture selection included biomechanical and pathophysiological factors as well as previous research studies (Armstrong et al., 1982; Stetson et al., 1991; McAtamney and Corlett, 1993; Juul-Kristensen et al., 2001; Spielholz et al., 2001; Lowe, 2004a; Lowe, 2004b). There is some disagreement in the literature regarding cut points for shoulder postures that create risk for musculoskeletal outcomes. The NIOSH review (1997) of 13 studies that have examined awkward postures and their relationship to shoulder musculoskeletal disorders concluded that there is evidence for a relationship between repeated or sustained shoulder postures with greater than 60° of flexion or abduction and shoulder musculoskeletal disorders. The NIOSH review (1997) considers shoulder flexion/adduction greater than 60° as awkward due to the greatest mechanical pressure on the supraspinatus tendon at arm elevations between 60°-120°. Much of the laboratory and clinical focus has been on upper arm elevation greater than 60°. A study that evaluated the intramuscular pressure of the infra- and supraspinatus muscles in relation to arm posture found that intramuscular pressure increases in association with upper arm elevation for both shoulder muscles (Palmerud et al., 2000). The elevation plane (flexion versus abduction) also influenced the intramuscular pressure; however, the elevation angle affected intramuscular pressure more. In addition, hand load greatly influenced the intramuscular pressure. A relatively new research study has shown an association and increased odds between upper arm flexion greater than or equal to 45° in combination with a pinch grip and rotator cuff syndrome (Silverstein et al., 2008). Shoulder postures were assessed by analysis of video footage using MVTA. A total of 733 subjects performing manufacturing tasks in 12 industries were evaluated. Based on the most recent findings established (Silverstein et al., 2008), a postural threshold of 45° was used when evaluating shoulder postures with MVTA and inclinometry (Table 5.1).

In regards to the trunk (low-back), the NIOSH review (1997) provided evidence that awkward postures of the trunk are associated with low-back disorders. The review of 12 studies that examined low-back postures evaluated a variety of different back postures simultaneously and did not differentiate between specific back postures as associated with low-back disorders. Determining posture cut points for the trunk is not as straightforward for those of the shoulder since so many other risk factors affect the outcome of low back disorders. Bloswick and Villnave (2000) proposed the use of three cut points based on the range of motion of the back as one bends forward. This model uses 90° as the extreme range of motion for forward bending (flexion). This 0°-90° range is quartered and results in 4 posture categories: 0°-23°, 23°-45°, 45°-67°, and 67°-90°. The present study combined these categories into two major categories: 0°-45° and 45°-90° (Table 5.1).

FIGURE 5.1

Research Design Schematic



Site and Subject Selection

The participant group was comprised of workers recruited at a single northern Colorado manufacturing facility. The participant group was expected to contain both male and female workers between the ages of 18-65, however only male participants were recruited due low numbers of female employees for the tasks evaluated. Workers with significant upper extremity pathology were not included in the study. During initial recruitment, subjects were asked if they had any history of upper extremity disorders.

Subjects from three work areas (bottling, kegging, and canning) of a brewery were recruited to participate in the study. Each work area consisted of several tasks, with each task typically performed by one worker. The three tasks were chosen for the study based on awkward body positions observed by the investigators. The bottling tasks were much more dynamic in nature than tasks performed in the kegging and canning areas.

The two primary tasks on the bottling line involved loading a variety of cardboard trays, carriers, and cases into a machine. The majority of the work was automated; however, the workers had to retrieve the cardboard materials from pallets and load them into the machine throughout their shifts. Loading the cardboard materials was not considered repetitive (1-2 stacks/minute); however, awkward postures were noted during troubleshooting and non-standard tasks. Workers did not rotate within a shift.

The kegging line was comprised of two major tasks: loading and offloading kegs. Workers handled two keg sizes, full-sized kegs and 1/6 kegs. When managing full-size kegs, one worker loaded the empty kegs (approximately 30 lbs) onto the line from pallets while the other worker used a lift-assist device to offload the filled kegs (approximately160 lbs). In contrast to managing 1/6 kegs, workers loaded the empty kegs (approximately 14 lbs) and offloaded the filled kegs (approximately 56 lbs) manually without mechanical assistance. Loading and offloading kegs was considered repetitive, due to upper arm movements based on the keg rate (3-4 full kegs/minute; 4-5 1/6 kegs/minute). In addition, loading kegs involved lifting a keg from floor to near waist or chest height using a lift assist device.

The canning line was comprised of three major tasks: loading cans, gluing cases, and loading cases into trays. Loading cans involved loading full 12-ounce aluminum cans into cardboard cases. Gluing cases involved the use of a gluing machine to seal the case. Once glued, cases were loaded into cardboard trays for palletization.

Exposure Assessment Tools

The current study employed two data collection tools to assess exposures to postures of the shoulders and trunk of 15 manufacturing workers. The first tool was a direct-reading instrument best described as an inclinometer. Inclinometers are instruments used for measuring angles of slope (tilt), inclination, or elevation of an object relative to gravity. The inclinometer (Figure 5.2) used in the present study was a pager-sized datalogging device called the Virtual Corset (MicroStrain, Inc.; Williston, VT). The Virtual Corset continuously collects postural data in two dimensions. This inclinometer can be mounted on the upper arms, sternum, or upper back of individuals and can continuously record a worker's postural exposures in two dimensions, relative to gravity. With programmable sampling rates, the Virtual Corset can collect continuous data over an entire work shift or over multiple days up to 80 hours (Microstrain, Inc.; Williston, VT). This measuring capability has not been practical previously due to limitations in memory and battery life.

The Virtual Corset inclinometer logs inclination in one degree increments over +/-180° in the flexion/extension axis and +/-70° in the lateral axis. When set in the 'Linear' mode, data are recorded in the actual angle of inclination in a continuous stream from beginning to the end of a session. The inclinometer can be programmed to record the X and Y axis separately or together (Figure 5.3). X is the angle developed by movement in the coronal or frontal plane of the body during lateral movement. Y is the angle developed by movement in the sagittal or lateral plane of the body during anterior or posterior movement.

FIGURE 5.2

MicroStrain Virtual Corset Inclinometer

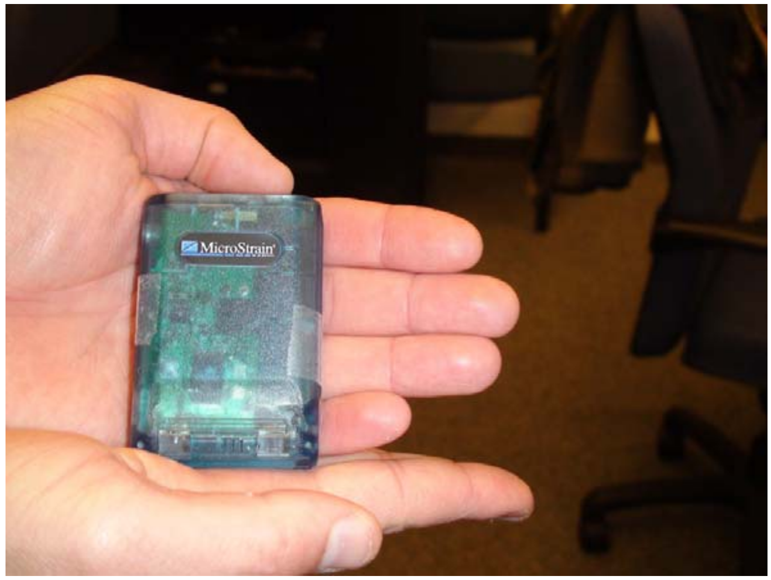
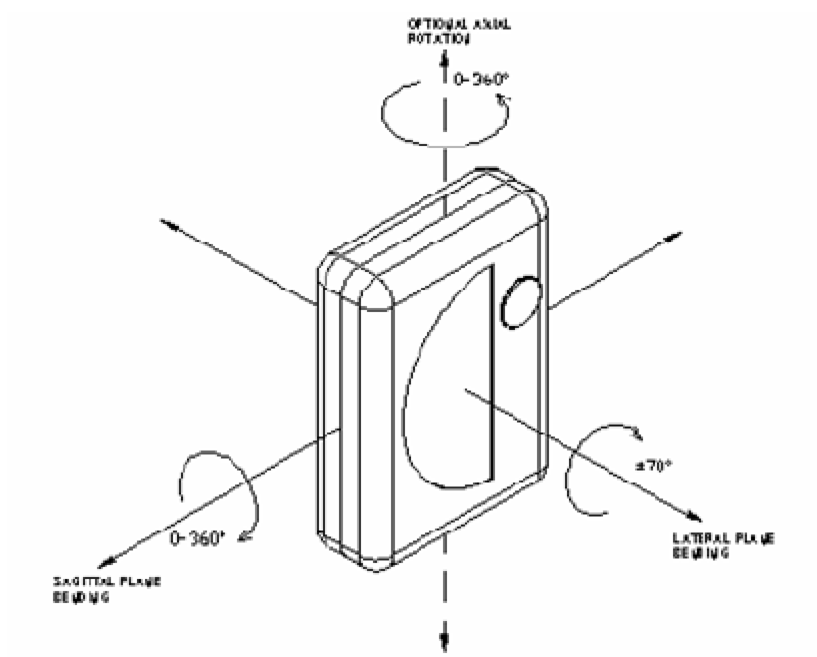


FIGURE 5.3

Diagram of the Virtual Corset Inclinometer



The second tool, or set of tools, used in this study were a Sony HD Handycam camcorder with subsequent video analysis obtained using Multimedia Video Task Analysis (MVTA). The camcorder device was used to capture exposures to postures of the shoulder and trunk indirectly via digital video recordings. The Sony Handycam records at 60 frames per second (fps) with the capability to record at 240 fps for slow motion capture. MVTA is a video-based exposure assessment program that can be used to automate time and motion studies and ergonomic analyses from video (Ergonomics Analysis and Research Consortium, 2003). MVTA uses a computer controlled video playback controller that allows interactive study of video footage enabling researchers to time log data such as posture events. Video is reviewed at any speed and in any sequence (real time, slow/fast motion, or frame by frame).

Data Collection

To assess exposures to postures of the shoulder and trunk, inclinometers were directly attached to the upper arms and trunk of each subject. One inclinometer was attached to each of the upper arms by mounting the inclinometer to the posterior aspect of the upper arm, mid-way between the shoulder and elbow. The reference position for the upper arm (0° elevation) was defined with the subject standing, the arm hanging relaxed to the side. A combination of athletic wraps and tapes (Figure 5.4) were used to secure the inclinometers to the upper arms. First, the upper arm was wrapped with pre-wrap to create a barrier between the skin and plastic housing of the inclinometer. Once the mounting position was determined, the inclinometer was secured with a combination of tape and wraps (Figure 5.5).

Another inclinometer was mounted tightly to the trunk at approximately the location of the T6 spinous process (Trask et al., 2006a; Trask et al., 2006b). To approximate the T6 spinous process, the researchers located the C7 spinous process for each subject and manually counted down the spinous processes until reaching the T6 spinous process. Once located, the inclinometer was centered over the spinous process and attached using a Tegaderm (3M: St. Paul, MN)(Figure 5.5). The Tegaderm created a seal around the inclinometer while maintaining flexibility during movements.

FIGURE 5.4

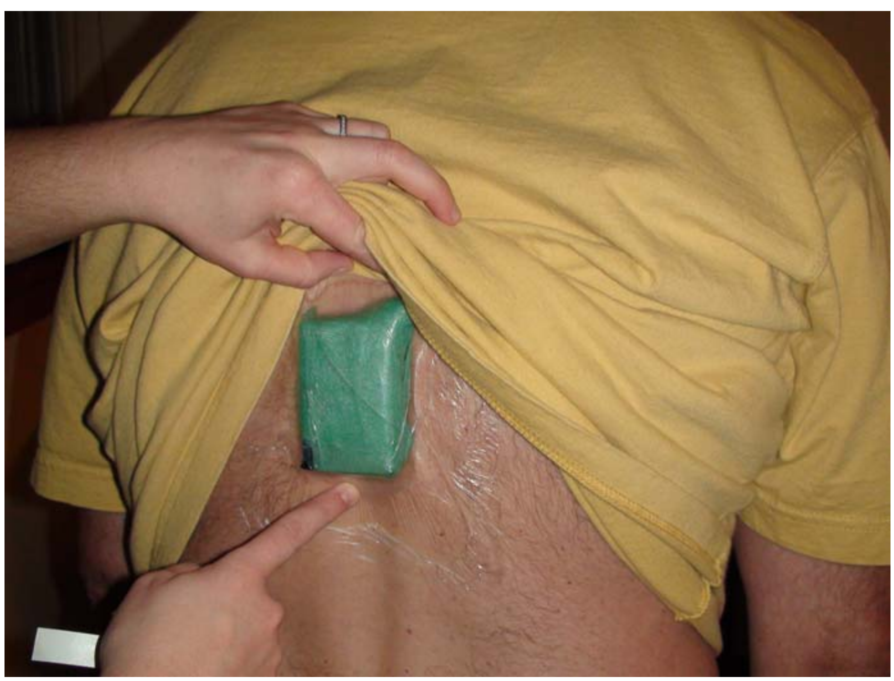
Wraps and Tape Used to Attach the Virtual Corset Devices

FIGURE 5.5

Virtual Corset Inclinometer Mounted to the Upper Arm

FIGURE 5.6

Virtual Corset Inclinometer Mounted to the Trunk



Before the inclinometers were attached, they were launched and calibrated using Virtual Corset 3.2.3 software (Microstrain, Inc; Williston, VT) installed on three laptop computers to correspond with the three inclinometers. Launching the inclinometers was a function of the software and used to erase data collected after battery insertion. To begin use of the inclinometers, two AA batteries were inserted. Once the batteries were inserted, the inclinometers began collecting data. Once the inclinometer preparation was complete, each inclinometer was connected to a laptop computer. Based on the inclinometer design and software, only one inclinometer could be connected to a computer at a time. While it was not ideal to setup three computers in the field, this allowed the inclinometers to be launched synchronously. It was important that all inclinometers were capturing the same exposure period. Once a connection was established for each inclinometer, the Virtual Corset software was used to perform a check. The inclinometers were laid flat on the table to verify all were reading approximately zero degrees. The angle of inclination was viewed directly on the computer screens using the Virtual Corset3.2.3 software.

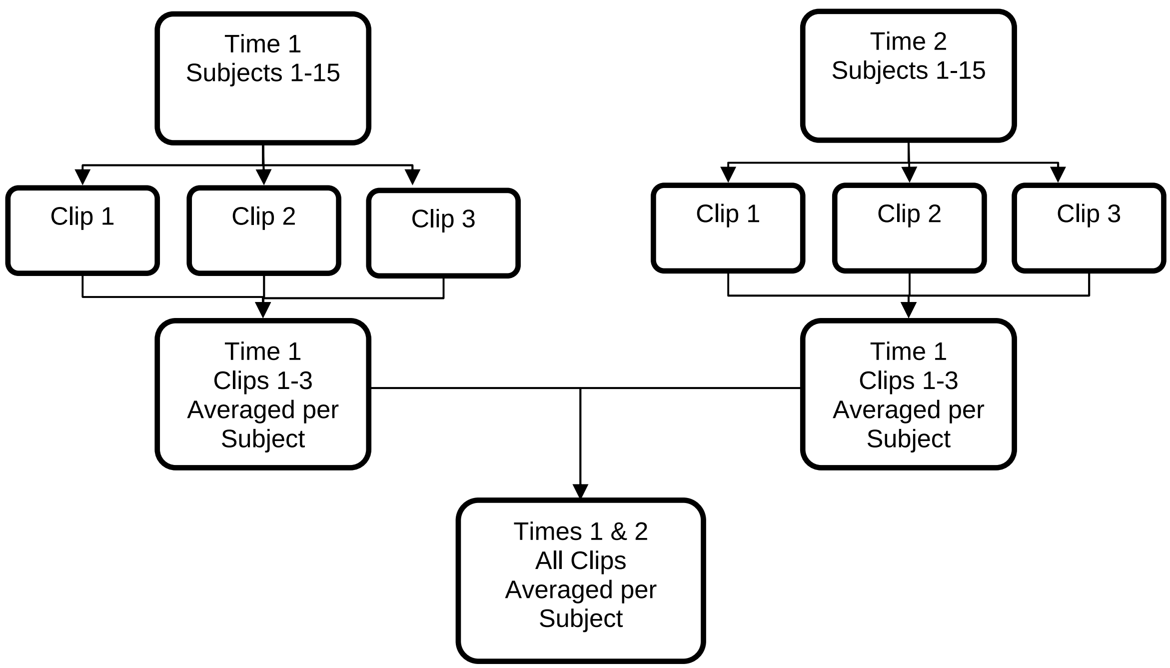
Once the checks were completed, the inclinometers were launched synchronously using the Virtual Corset data. In order to ensure synchronization with the video recordings, a dedicated watch (hh:mm:ss) synchronized with the internal clocks of the three laptop computers was used to verify the launch time. This watch was then used throughout the shifts to time stamp the video recordings.

When mounting the inclinometers on the upper arms, subjects were instructed to stand upright with their arms hanging to the side. While one researcher attached the inclinometer, the other researcher secured the arm at the side of the subjects. Once the attachment location was determined, the inclinometers were calibrated to the individual using the Virtual Corset software. When attaching the inclinometers to the trunk, subjects were instructed to stand upright while staring straight ahead. Again, one researcher determined the location and secured the inclinometer while the other researcher calibrated the inclinometer to the individual using the Virtual Corset software. While attaching the inclinometers on the subjects, a second check was performed. Subjects were instructed to move their arms and trunk to designated postures. Subjects were required to raise their arms from approximately 0° to approximately 90° simultaneously and bend forward at the waist to ensure the inclinometers were functioning properly. The angles of inclination were viewed directly on the computer screens using the Virtual Corset3.2.3 software.

The inclinometers remained attached to the subjects for the duration of their work shifts, with a minimum run time of four hours. The camcorder was used to capture the performed work tasks. Due to the detail of analysis of the video footage and the recording capabilities of the camcorder, only portions of the work shift were recorded. In an attempt to capture task variability over the work shift, video footage was obtained at the beginning, middle, and end of the work shifts. For each separate video footage collection time, approximately five minutes of the task was recorded for later analysis. This resulted in six separate five-minute video clips for each subject with an expected 90 five-minute video clips in total (Figure 5.7). The six separate five-minute segments for both MVTA and the inclinometers, once analyzed, were averaged to obtain one value, mean, for each posture category for each of the 15 subjects (Figure 5.7). As discussed previously, the inclinometers and camcorder video footage were synchronized from the launch time using a dedicated watch. At the beginning of the video clip, the dedicated watch was placed in front of the camcorder so it could be visible during later analysis. The same practice was performed at the end of the clip.

FIGURE 5.7

Schematic of Five-Minute Data Collection Process for Video Observation



Preliminary Data Collection

Preliminary data collection was performed to determine attachment options for the inclinometers, to test calibration and synchronization methods, to test the sampling rate of the inclinometers, and to test the T6 spinous process mounting.

To determine the best process for attaching the inclinometers to the upper arms and trunk, two subjects performing work in a dairy parlor over two work shifts were evaluated. The initial attachment process consisted of armbands and back straps. It was observed that these attachment devices moved on the body in a way that affected the data collection. While the data were usable for some of the subjects, the researchers determined that a better attachment method was necessary. After some troubleshooting, the researchers determined that a combination of athletic wraps and tapes would provide the best option for mounting the inclinometers to the arms and trunk with minimal movement. These methods were then tested on a female undergraduate student who wore the inclinometers for eight hours during two different typical school days. The tape and wrap method proved successful with minimal discomfort to the subject.

In terms of calibration and synchronization, different methods were tested before data collection. The researchers initially planned to take one laptop computer into the field. This single computer would have been used to check and launch the inclinometers. Previous studies that have compared different exposure assessment tools have typically implemented a jig with known angles of inclination or set movements that the subject must perform (Trask et al., 2008; Juul-Kristensen et al., 2001; Spielholz et al., 2001). These angles or movements would result in obvious changes in the inclinometer output. After downloading the data, synchronization of the inclinometers was performed by identifying the known angles or movements and removing data before the known angles or movements. This method was attempted for the current study. A jig was created that would force the inclinometers between known angles of 0° and 90°. While the known angles could be seen in the resultant data, it was very difficult to determine cut-off points. In addition, since each inclinometer was launched at a different time, the time stamp for synchronization with the video recordings produced errors.

The researchers decided to use three laptop computers which allowed simultaneous data collection and a synchronous launch of the inclinometers. In addition, the time stamp for video recording worked more readily since all inclinometers could be launched at the same time. While previous studies (Trask et al., 2008; Juul-Kristensen et al., 2001; Spielholz et al., 2001) utilized a calibration technique of known movements to synchronize different exposure assessment tools, the current study employed two checks to ensure the inclinometers were working properly. This enabled a shorter setup time. The trial period also led to the conclusion that the inclinometers should be tared to each subject after attachment. Taring to the subject adjusted for body morphology differences thus allowing for more direct comparisons between individuals.

In terms of sampling frequency of the inclinometers, the researchers based their initial analysis of data on the reported sampling rate of 7.5 Hz or 7.5 samples per second (Microstrain, Inc; Williston, VT). The researchers questioned the 7.5 Hz sampling rate since a previous study reported 7.6 Hz as the Virtual Corset sampling rate (Trask et al., 2006a; Trask et al., 2006b). In an attempt to verify the sampling rate, the researchers performed a test of the sampling frequency. Two inclinometers were launched using the Virtual Corset 3.2.3 software and allowed to sample for aduration of five minutes on three separate occasions, thus resulting in three five-minute samples for each inclinometer. At a sampling frequency of 7.5 Hz, one would expect the Virtual Corset devices to output 2250 data points based on a five minute sampling duration. The difference between the expected versus actual number of data points averaged over the six samples was 19 data points. The true sampling frequency of the two Virtual Corset devices tested based on the average of the actual number of data points averaged over the six samples divided by the duration sampled (5 minutes) resulted in a sampling frequency of 7.59 Hz or 7. 6 Hz. This result corresponds to the sampling frequency reported by Trask et al. (2006a; 2006b), not by the sampling frequency reported by the manufacturer (Microstrain, Inc; Williston, VT). Therefore, results obtained in the current study were analyzed using a sampling frequency of 7.6 Hz.

The location of attachment of the inclinometers to the trunk should depend on the questions being asked. For the present study, the researchers wanted to understand the implications of mounting error and the possible effects on the resulting data. As discussed previously, the inclinometers were mounted tightly to the trunk at approximately the location of the T6 spinous process (Trask et al., 2006a; Trask et al., 2006b). To approximate the T6 spinous process, the researchers would locate the C7 spinous process for each subject and manually count down the spinous processes until reaching the T6 spinous process. Once located, the inclinomter was centered over the spinous process and attached using a Tegaderm (3M: St. Paul, MN) (Figure 5.5). In an attempt to quantify the possible error associating with mounting the device in the same location for every subject for both occasions, the researchers mounted an inclinometer on the T6 spinous process of a subject in a laboratory setting. Once mounted, the subject was instructed to stand upright while staring straight ahead while the device was launched. The subject was instructed to remain in this position for the first one minute of data collection. After one minute of data collection, the subject was instructed to bend forward at the waist to an approximate 90° trunk flexion for 30 seconds. This position was verified with a manual goniometer. After the 30 second period, the subject was instructed to return to an upright position for 30 seconds, back to 90° degrees for 30 seconds, upright for 30 seconds, back to 90° for 30 seconds, and finally upright again for 30 seconds. This process was repeated with the inclinometer mounted in two other locations: the approximate T5 spinous process and the approximate T7 spinous process.

The data for the three, four-minute data collection periods were then download and analyzed using LabView 8.6 software (National Instruments:Austin, TX). Analysis of this data was performed for the same trunk postures evaluated in the present study (Table 5.1). The mean percent time found in the < 45° trunk posture category was 61.29% with a standard deviation of 1.9 and 95% confidence interval (59.14, 63.40). The mean percent time found in the > 45° trunk posture category was 38.71% with a standard deviation of 1.9 and 95% confidence interval (36.56, 40.86). Based on this data collection, 95% of the means for both posture categories were no more than 2.15% (1.96 standard deviations) from the mean. A separate study could be performed that evaluates the location of mounting inclinometers on the trunk and the resultant implications on posture results. This preliminary data was collected to have some understanding of the implications of mounting the trunk inclinometer in slightly different locations. Based on the results, one would not expect error that significantly affects the final data, if an inclinometer were mounted slightly above or below the T6 spinous process.

Sampling Duration and Sample Size

Three major work areas were evaluated at the manufacturing facility: bottling, canning, and kegging. Two to three different tasks as performed by workers for each of these areas were used in this study. Based on the number of different tasks (seven tasks), and the available crewing (three crews), it was determined that 14 manufacturing workers would be available for full-shift exposure assessment. Recruitment was performed by the researchers with assistance from the facility health and safety personnel. Due to some changes in personnel throughout the study, 17 workers were available for recruitment. Workers were sampled over two separate work shifts. Inclinometry was collected over 8-hour work shifts, with minimum sample times of four hours for all manufacturing workers over both occasions. Video footage was collected as described previously. Sampling duration and analysis for video footage was expected to total 15 minutes per subject per shift.

Sampling duration for the video footage was based on the nature of the tasks, total sample size, and the detail of the analysis. Sampling duration was also based upon previous research using inclinometers and video recordings (De Looze et al., 1994; Hansson et al., 2001; Juul-Kristensen et al., 2001; Spielholz et al., 2001; Lowe, 2004b; Hansson et al., 2006). Sample duration for the above referenced studies ranged from short cyclic tasks (2 minutes) to full work shifts (8 hours). Given the nature of the tasks to be assessed, it was expected that stable and representative exposure estimates could be obtained in a relatively brief period of time. Previous research (Juul-Kristensen et al., 2001; Spielholz et al., 2001) comparing video observation with direct technical measurements has typically involved recording a continuous segment (20-30 min) of work time for 6-19 subjects performing stationary, repetitive tasks while simultaneously obtaining direct measures via electromyography, electrogoniometry, or accelerometry. While sampling times for the video footage were short, the data collected were used to cross-validate two methods and not to associate exposure with a health outcome. However, if one were to use short video clips in attempt to associate exposure with health outcomes, misclassification of exposure could occur if the video clips are not representative of the exposure. However, the misclassification would be non-differential; meaning the amount of measurement error in classifying a task by awkward postures is the same in those with and without the outcome. This can be better described as random error distributed evenly among all observations and therefore will never create an appearance of association with the health outcome. This random error will always underestimate exposure.

Data Analysis

Data obtained from the inclinometers was downloaded using Virtual Corset 3.2.3 software (Microstrain, Inc; Williston, VT). The Virtual Corset software creates a Microsoft Excel comma-separated file with degrees of angle inclination for the X and Y axis. The Virtual Corset data files were then further analyzed using LabView 8.6 software (National Instruments: Austin, TX). The LabView software was programmed specifically for the data obtained by the Virtual Corsets. Upon uploading the data into the LabView software, the entire data file was reviewed and the specific 5-minute segments corresponding to the video footage were extracted and analyzed. An Exposure Variation Analysis for the posture categories of interest was created for each 5-minute clip. This provided the percent of time spent in the posture categories of interest, which was the same information obtained from analysis of the video footage.

Video footage was analyzed by one experienced analyst using Multimedia Video Task Analysis (MVTA) (Ergonomics Analysis and Design Research Consortium, 2003). MVTA is a video-based exposure assessment program that can be used to automate time and motion studies and ergonomic analyses from video. This program uses a computer controlled videocassette recorder that allows interactive study of video footage enabling researchers to time log data such as posture events. Video is reviewed at any speed and in any sequence (real time, slow/fast motion, or frame by frame). In regards to posture, MVTA allows for the evaluation of any pre-determined posture over a continuous period of time. The primary variable of posture for both devices, the inclinometers and MVTA, was the duration of task time spent in specific pre-determined postures (Table 5.1). Therefore, the percentage of time spent in each posture category as obtained from the two measurement tools for every 5-minute segment was compared to evaluate inter-method reliability. Analysis took place in the Ergonomics Research Lab at Colorado State University located in Fort Collins, Colorado. After completion of each analysis, MVTA was used to generate a time study report. These reports provided the percentage of total cycle time spent in each posture category for all body parts. These percentages were then used in the statistical analyses.

The analyst was an experienced user of the MVTA program to estimate postures from video and had been performing similar analyses for approximately five years. Since the inclinometers recorded elevation of the upper arms and inclination of the trunk relative to gravity, anatomical referencing as discussed in a previous study (Dartt et al., 2009) conducted by the researchers was not feasible for this study. For example, if a subject was bending at the waist and flexing their upper arms, the upper arms may be at 90° relative to an anatomical reference, but not 90° relative to gravity. The concern when comparing to different measurement devices that measure postures is that both methods are using the same definition for posture estimation. If the posture definitions are not the same, it is likely that the two different devices could differ based on this factor. In order to ensure the best approximation of posture from video and to ensure that the same posture definitions were used for each device, several techniques were used to estimate the specific postures of the left and right upper arms and trunk from video (Table 5.1). First, the analyst had been trained previously in recognizing specific degrees of excursion from neutral. Second, the analyst utilized known angles drawn on transparencies to assist in analyzing the anatomical positions viewed on the computer monitor. Third, the analyst used known landmarks in the video footage to reference postures relative to gravity. Landmarks included structural upright components of the building or machinery. For example, the analyst was trained to recognize > 90° upper arm elevation, utilize a 90° reference angle, and to record greater than 90° upper arm elevation each time the upper arm (mid-humerus) exceeded 90° relative to gravity. In regards to the trunk, the analyst was trained to identify ≥ 45° flexion, utilize a 45° reference angle, and to record ≥ 45° flexion when an approximate visual line was drawn through the area of the T6 spinous process and this line was equal to or exceeded 45° relative to gravity.

Statistical Analyses

Various statistical measures were used to evaluate the data. All statistical analyses were completed using SPSS 18.0 statistical package. The percent time logged in the specific posture categories of the upper arms and trunk were compared across the two measurement tools. Pearson Product Moment correlation coefficients were used to evaluate correlation between the measurement tools. Correlation values are referred to as a measure of precision or reliability for pair-wise inter-method comparisons (Armstrong et al., 1994; Burdorf, 1995).

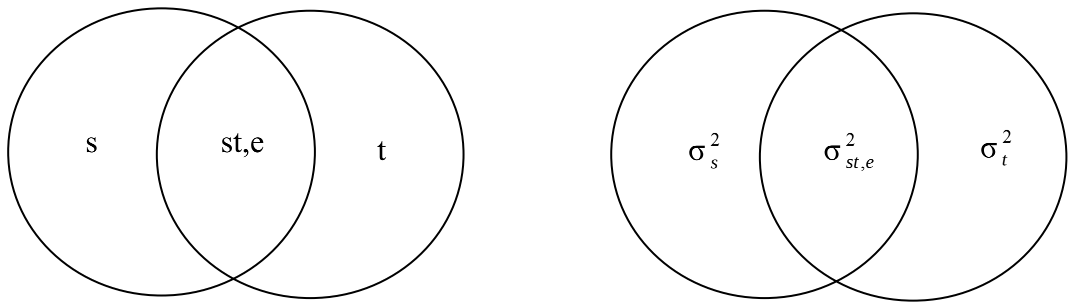
As an alternative method to evaluating the inter-method reliability between the video observation and inclinometry tools, Generalizability Theory was applied.

Generalizability theory (G-theory) is a standard measure of reliability commonly used in the social sciences. The statistical model that drives G-theory is the analysis of variance (ANOVA) (Burns, 1998). Subjects as the object of measurement are a common source of variation in most G-study models. Other variables, or facets, in the model can be occasions, raters, or anything a researcher specifies (Burns, 1998). G-theory facets are analogous to factors or variables in ANOVA. G-theory provides a summary coefficient expressed as rho (ρ) that is analogous to the intraclass correlation coefficient (Shavelson and Webb, 1991).

Variance estimates for the Generalizability Theory analysis were obtained for a subject x tool fully-crossed random effects model. The model consisted of the percent time spent in the specific postures of the upper limbs and trunk (averaged across the five-minute segments and two shifts) and the two tools evaluated (video observation and inclinometry). Figure 5.8 provides a pictorial description of the sources of variability and variance components, where 's' represents subject, 't' represents tool, and 'e' represents residual error. Figure 5.8 is modeled after example figures in Generalizability Theory: A Primer (Shavelson and Web, 1991). To compute the variance estimates, the data for each posture category were split by body part (left upper arm, right upper arm, and trunk). This eliminated the need to treat body parts as a fixed variable since body parts were used to organize the output. Subjects and tools were treated as random variables for variance component estimation. Variance estimates were obtained for each posture category (Table 5.1) of the upper arms and trunk as separate analyses. Posture analyses were completed separately since the posture categories were not independent of each other. For example, if the analyst rated a posture as trunk flexion $\geq$ 45°, the other trunk posture could not be selected; however, the percent time spent in the posture categories must equal 100%. Explanations of the variance components are provided in Table 5.2.

FIGURE 5.8

Venn Diagrams for the Subject(s) x Tool(t) Design



(c) Sources of Variability (b) Variance Components

In the residual (e) variance component, the subject(s) x tool(t) interaction is confounded with unmeasured or unsystematic variability. The subject(s) x tool(r) component reflects whether the postures of the subjects differed across tools. The residual (e) component reflects unsystematic or random error sources. It also includes systematic error from variables not explicitly included or controlled for in the study.

TABLE 5.2

Variance Components

Variable	Symbol	Description
Subject variance	σ_s^2	Variance in postures across subjects or how much subjects differed by postures
Tool variance	σ_t^2	Variance in postures across tools or how much the tools differed by postures
Subject x Tool variance, residual	$\sigma_{st,e}^2$	Variance of postures due to subject x tool interaction and variance that cannot be explained by tools

Variance estimates were then used to compute Generalizability Theory reliability coefficients for each posture category among the body parts (Table 5.4). G-coefficients were computed using equations published in Shavelson and Webb (1991) as well as Deshon (2002). Since the tool variable was treated as a random factor to obtain variance estimates, the variance estimates had to be modified in order to address tool as a fixed factor. The tool variable was treated as a fixed factor since the evaluation was limited to two specific tools. To adjust the variance estimates based on averaging over conditions of the fixed tool factor, new variance component estimates were computed (Table 5.3) for the random subjects factor using an equation published in Shavelson and Webb (1991). Reliability coefficients were calculated for interpretation of absolute decisions. This type of generalizability indexes the absolute level of the posture assessment for each subject with no comparative reference to the posture assessments of the other subjects (VanLeeuwen, 1997).

TABLE 5.3

Equation Used to Average the Random Factor Over the Conditions of the Fixed Factor

Equation
$\sigma^2{}_{s*} = \sigma^2{}_s + \left(\frac{1}{n_s} \right)\left(\sigma^2{}_{st,e}\right)$

TABLE 5.4

Formula Used to Estimate the Generalizability Theory

Inter-Method Reliability Coefficents

Study Design		
Subjects	**Tool**	**Coefficient estimation formulas**
15	2	$\dfrac{\sigma_s^2}{\left\{ \sigma_s^2 + \frac{\sigma_t^2}{2} + \frac{\sigma_{st,e}^2}{2} \right\}}$

To investigate whether differences existed between work areas when comparing the two measurement tools, a repeated measures ANOVA was performed for each body part and posture category with the measurement tool as a within-subjects factor and work area (bottling, kegging, or canning) nested as a between-subjects factor. A significance level of 0.05 was used for all comparisons. The measurement tools are the within-subjects factor because the two measurement tools were used to measure the dependent variable (posture) repeatedly for all subjects. The between-subjects factor, work area, is the categorical variable used to group the subjects. If the F –test for the within-subjects factor, measurement tool, is significant, then one concludes that it is not true that the percent time spent in a specific posture for a specific body part does not change over the measurement tools. If the F-test for an interaction involving the within-subjects factor is significant, than one concludes that the change over the measurement tools in the percent time spent in a specific posture for a specific body part is not the same for all work areas.

Two additional measures, the Levene's test (Levene, 1960) and Hartley's test (O'Brien, 1981) were computed to further investigate the differences in recorded postures between the video observation and inclinometry measurement tools. Homogeneity of variance is one of the assumptions of using the ANOVA and can be tested in various ways. For the present study, the homogeneity of variance was evaluated using the Levene's test (Levene, 1960) as performed as part of the repeated measures ANOVA using SPSS Statistical Package 18.0. Ideally, error variance of each repeated measures dependent variable should be the same across groups formed by the between-subjects factors. If the Levene statistic is significant at the 0.05 significance level, the null hypothesis that the between-subjects factors have equal variances is rejected. The failure to meet the assumption of homogeneity of variances is not fatal to ANOVA models, which are relatively robust, particularly when groups are of equal sample size.

The F_{max} or Hartley's test (O'Brien, 1981) was used to verify whether different groups have a similar variance. This test is used in the ANOVA and involves computing the ratio of the largest group variance to the smallest group variance. The resulting ratio is then compared to a critical F value as obtained from the F sampling distribution for the specified degrees of freedom and level of significance. Hartley's test assumes that the data are normally distributed and that each group has an equal number of members, similar to the assumptions of the ANOVA and Levene's test. However, the Levene's test is robust even when there are departures from normality.

RESULTS

Of the 15 subjects enrolled in the present study, 100% were male, 80% were right-hand dominant, mean age in years was 35.4 (24-59), mean height in inches was 70.6 (66-73), and mean months worked in the particular job areas assessed was 18.7 (3-42). Of the 15 subjects analyzed, six were recorded in the kegging area, five were recorded in the canning area, and four were recorded in the bottling area. In total, 85 approximate five-minute analyses out of an expected 90 were completed using both MVTA and the inclinometer data and used for statistical analysis. Five analyses could not be included in the data analysis due to missing data from either the MVTA or inclinometer data collection methods. This resulted in an approximate 5.7 (4-6) five-minute analyses completed for each of the 15 subjects. Based on the five missing analyses, 15 subjects were used for statistical analyses for the upper arms, while 14 subjects were used for statistical analyses of the trunk. The mean time analyzed for both the inclinometer data and video observation data across the 15 subjects for both shifts was 14.7 (7.8-15.3) minutes. The mean time to perform an MVTA analysis averaged over the 15 subjects, both shifts, and all body parts was 31.4 (10-50) minutes, while the mean time to extract the five-minute segments from the inclinometer data averaged over the 15 subjects, both shifts, and all body parts was 28.8 (18-39). The mean time, averaged across the 15 subjects and both shifts, to attach and perform the checks for the inclinometer devices was 17.2 (14-23) minutes. This time did not include set-up by the researchers prior to attaching the devices.

Means of the raw data for each posture category of the upper arms and trunk for both the MVTA and inclinometer analyses are provided in Tables 5.5-5.7. The data provided in Tables 5.5-5.7 represent the mean percent time logged in each posture category as analyzed by both measurement methods across the 15 subjects and both shifts for the left upper arm, right upper arm, and trunk with ranges provided in parentheses. Figures 5.9-5.11 provide column charts that describe the average percent time spent in posture categories of the left and right upper arms and trunk as measured by the video observation and inclinometry methods. The average percent time spent in the specific posture categories of the left and right upper arms was highest for 0°-44° elevation and lowest for > 90° elevation for both instruments. The left upper arm, on average across both measurement methods, was logged 5.7% more time in 0°-44° elevation than the right upper arm. The largest average percent difference (2.13 %) was found in the 0°-44° elevation for the right upper arm followed by > 90° elevation for the right upper arm (1.43 %). In addition, standard deviations for the right upper arm were higher than the left upper arm, indicating that the right upper arm had a higher variability or dispersion.

TABLE 5.5

Average Percent Time Logged by Video Observation and Inclinometry for Left Upper Arm Elevation Postures

	Average % Time (Range)		
Measurement Method	**0°-44° Elevation**	**45°-90° Elevation**	**> 90° Elevation**
Video Observation	91.69 (77.20-96.61)	8.05 (2.57-22.80)	0.26 (0.00-1.35)
Inclinometry	90.41 (79.98-97.30)	8.71 (2.56-17.01)	0.67 (0.10-2.48)

FIGURE 5.9

Average Percent Time Logged by Video Observation and Inclinometry

for Left Upper Arm Elevation Postures

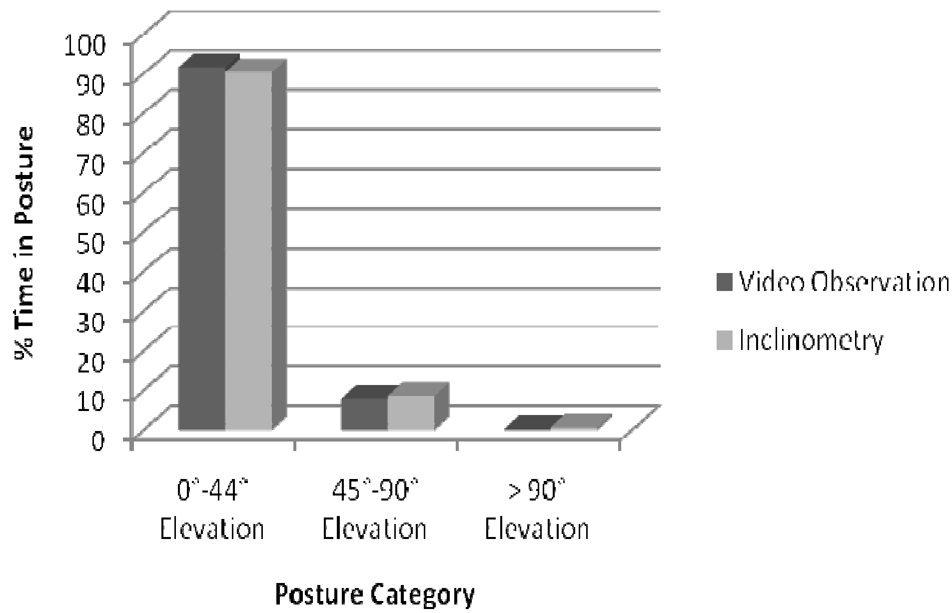


TABLE 5.6

Average Percent Time Logged by Video Observation and Inclinometry

for Right Upper Arm Elevation Postures

	Average % Time (Range)		
Measurement Method	**0°-44° Elevation**	**45°-90° Elevation**	**> 90° Elevation**
Video Observation	86.39 (54.64-94.06)	12.68 (5.44-47.13)	0.93 (0.00-2.30)
Inclinometry	84.26 (48.90-98.36)	13.56 (4.25-41.60)	2.36 (0.15-9.48)

FIGURE 5.10

Average Percent Time Logged by Video Observation and Inclinometry

for Right Upper Arm Elevation Postures

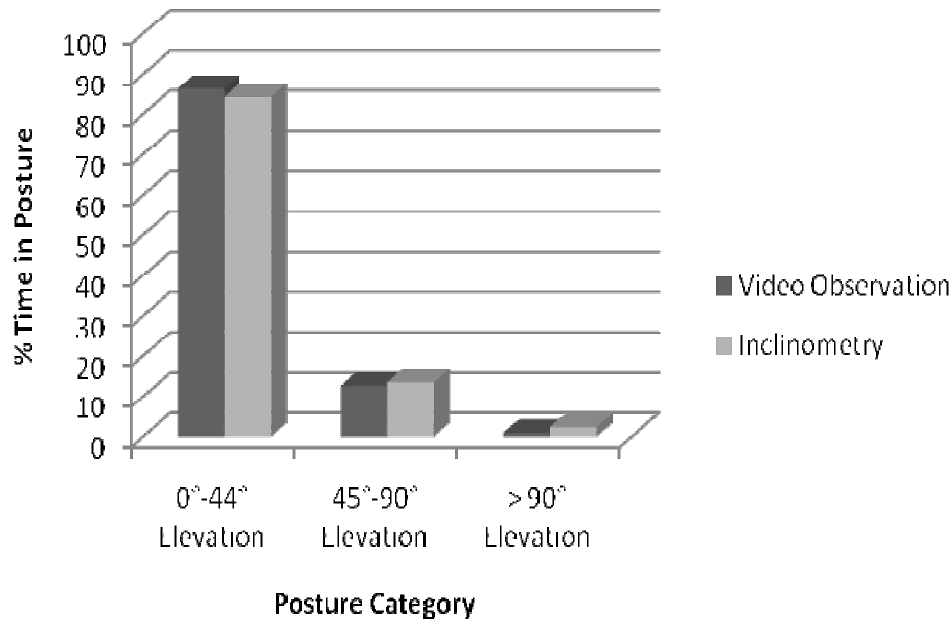


TABLE 5.7

Average Percent Time Logged by Video Observation and Inclinometry

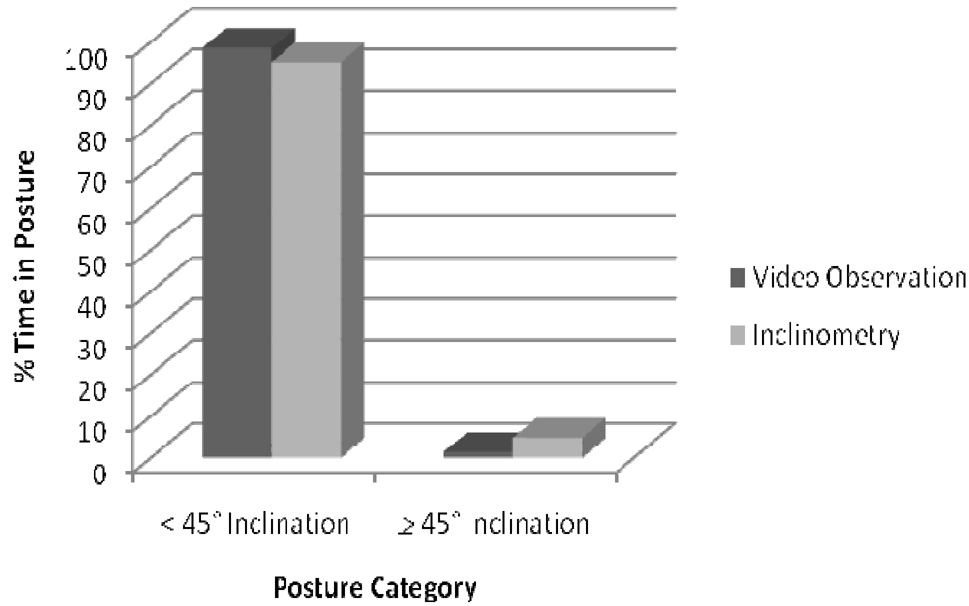
for Trunk Inclination Postures

Measurement Method	Average % Time (Range) < 45° Inclination	≥ 45° Inclination
Video Observation	98.57 (96.73-99.67)	1.42 (0.16-3.27)
Inclinometry	95.13 (90.59-98.33)	4.84 (0.94-9.41)

FIGURE 5.11

Average Percent Time Logged by Video Observation and Inclinometry

for Trunk Inclination Postures



Plots were constructed to evaluate differences between the video observation and inclinometry methods on the subject level (Figures 5.12-15) and also provide a visual representation of variability. The plots provided in the present study were used to visually demonstrate the variability of the 15 analyses as performed by the two measurement tools. Overall, the right upper arm elevation postures were more scattered or variable than the left upper arm elevation postures for both measurement tools (Figures 5.12-5.14). In addition, upper arm postures were more variable than trunk postures (Figures 5.12-5.15). For upper arm postures > 90° (Figure 5.14), the right upper arm demonstrated more variability between than two measurement tools than the left upper arm. For trunk inclination < 45°, the inclinometer measurements were consistently lower than the video observation measurements. This was opposite for trunk inclination ≥ 45° in that the inclinometer measurements were consistently higher than the video observation measurements.

FIGURE 5.12

Plots of Left and Right Upper Arm Elevation Postures for 0°-44°

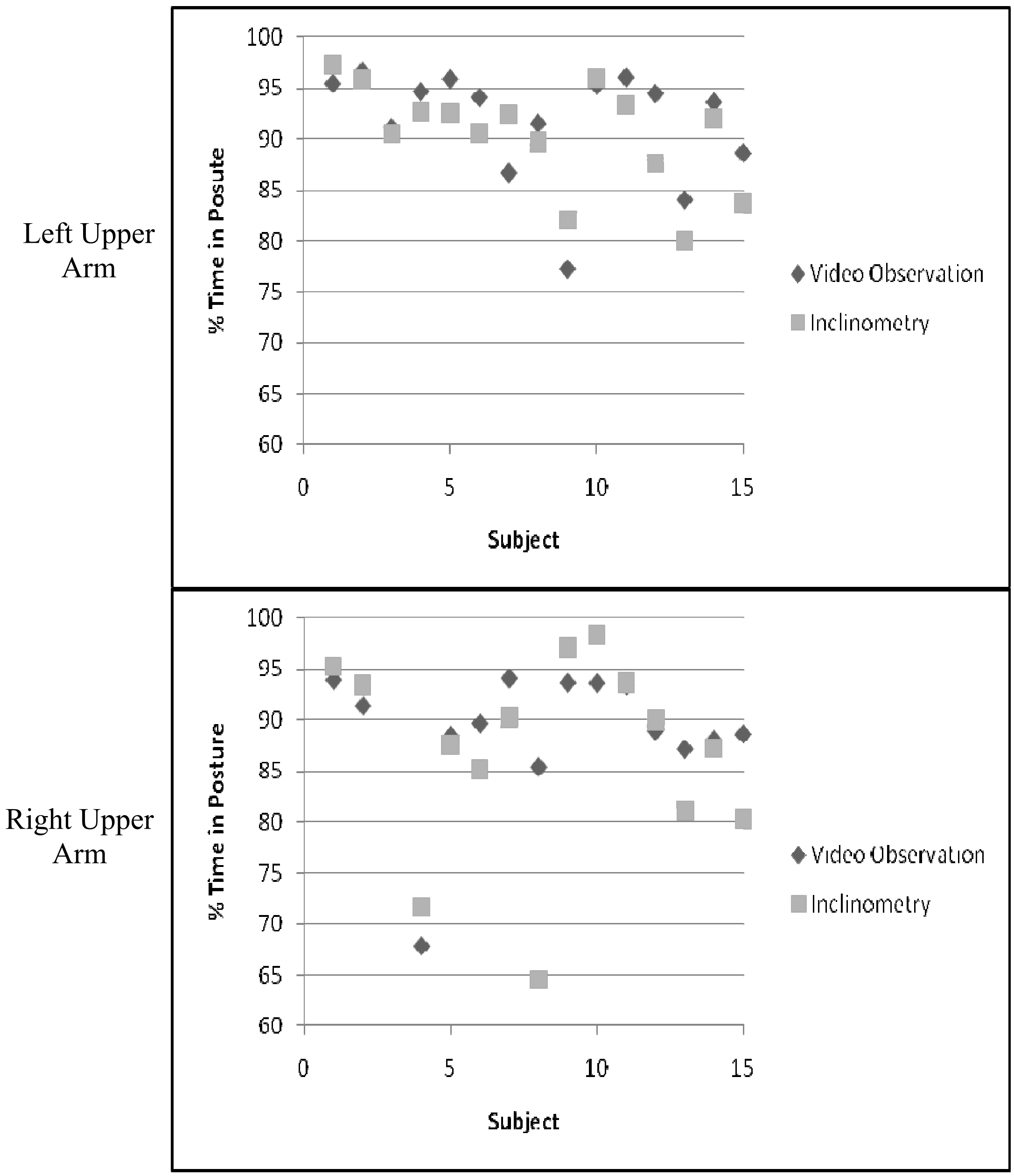


FIGURE 5.13

Plots of Left and Right Upper Arm Elevation Postures for 45°-90°

Left Upper Arm

Right Upper Arm

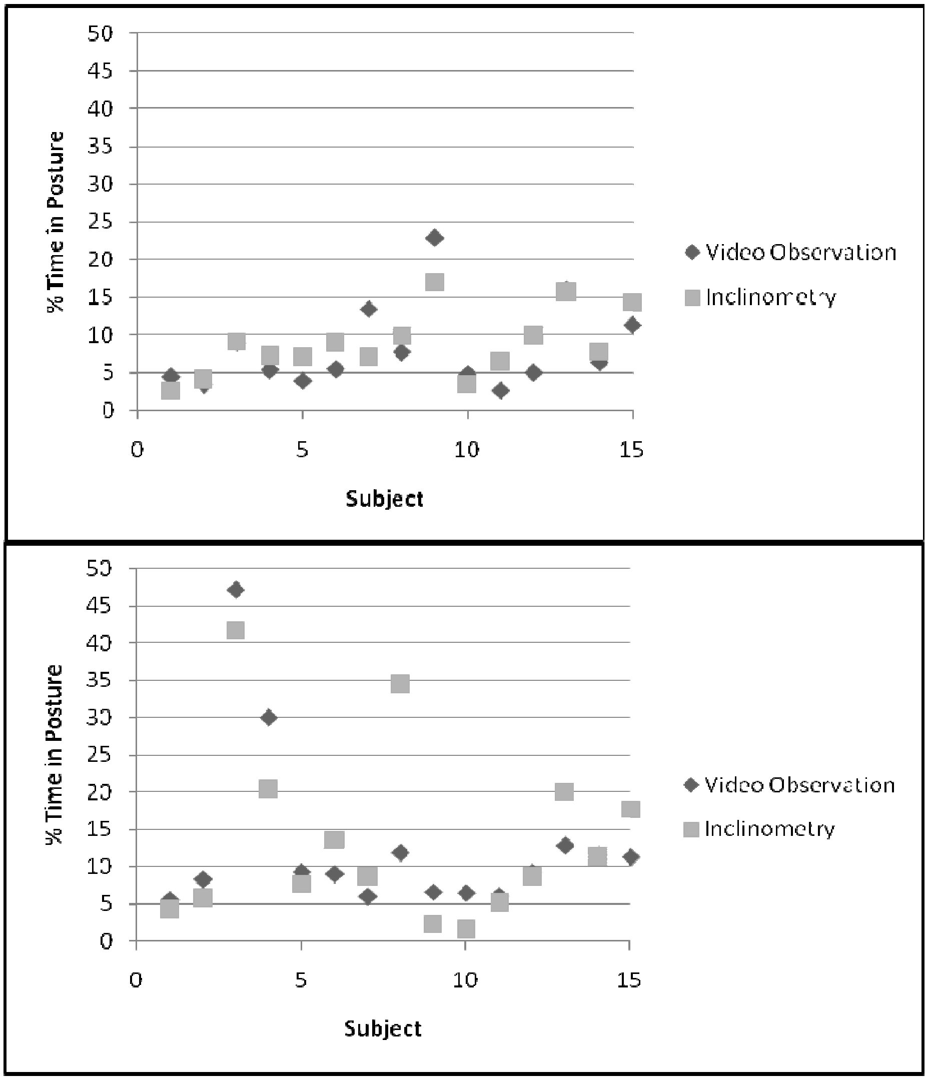
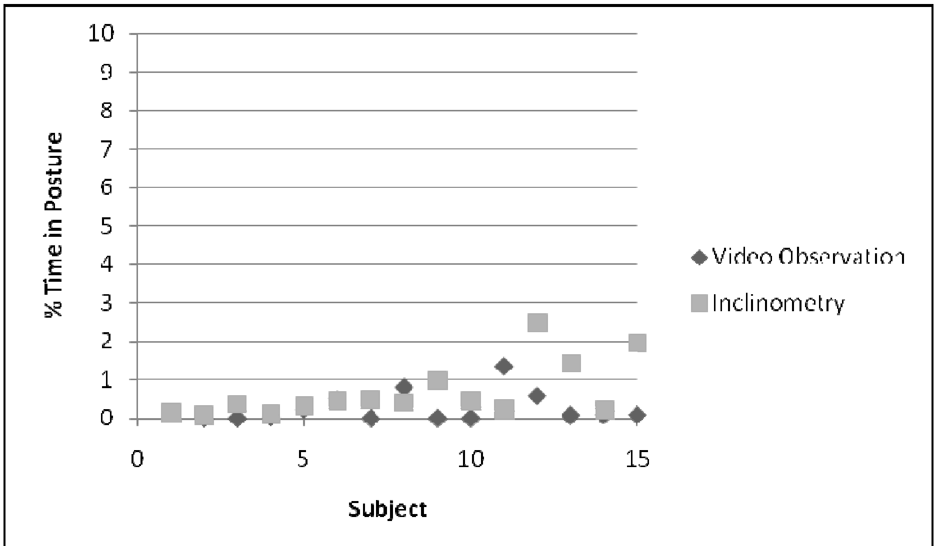


FIGURE 5.14

Plots of Left and Right Upper Arm Elevation Postures for > 90°

Left Upper Arm



Right Upper Arm

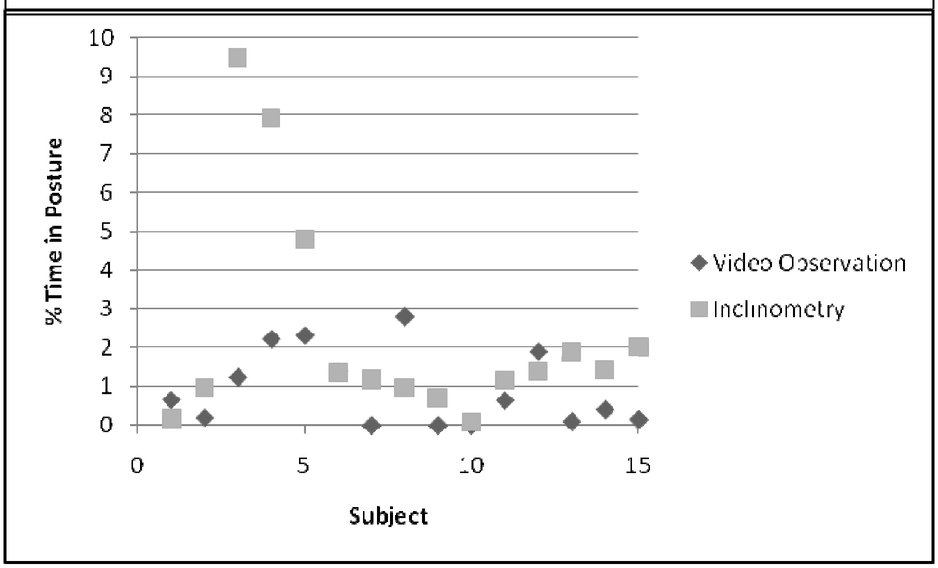
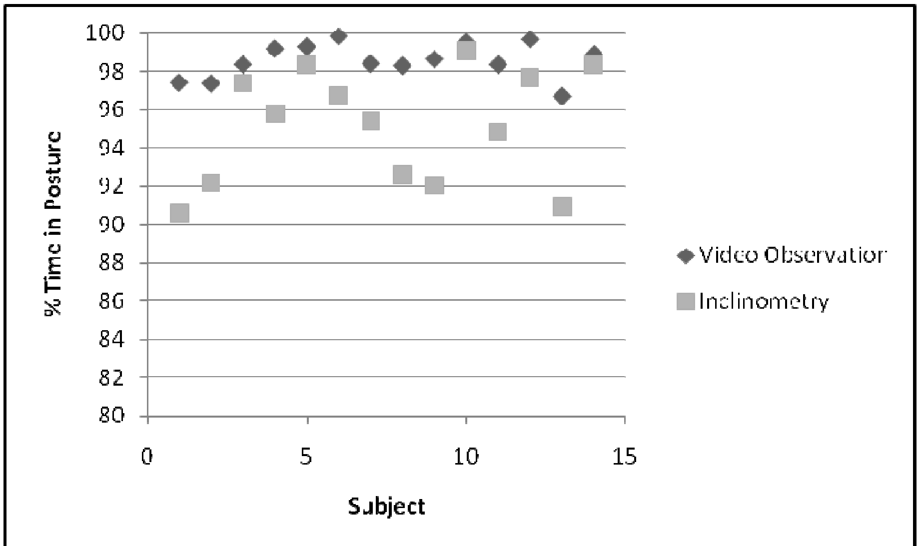


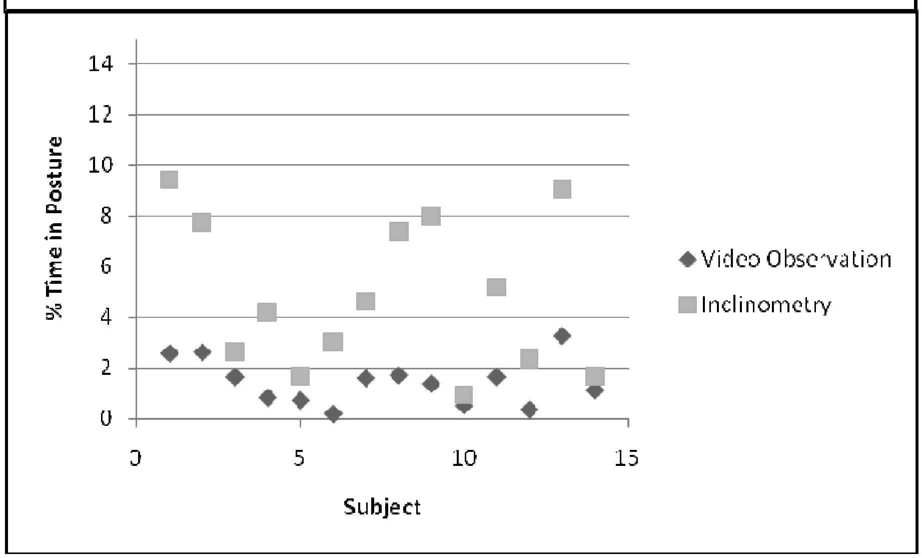
FIGURE 5.15

Plots of Trunk Inclination Postures for < 45° and ≥ 45°

< 45°



≥ 45°



Pearson product moment correlation coefficients for postures of the left and right upper arms and trunk were computed to represent the inter-method reliability of the video observation and inclinometry measurement methods (Table 5.8). Pearson correlations are either 1.00 if there is an increasing linear relationship, -1.00 if there is a decreasing linear relationship, or some value between 1.00 and -1.00. The correlation values indicate the degree of linear dependence between two variables. It should be noted that Pearson's correlation coefficient is not recommended for sample sizes <15 when there are two ratings (Walter et al., 1998). For samples sizes <15, Pearson's overestimates the correlation. The use of an intraclass correlation coefficient is recommended for sample sizes <15 (Walter et al., 1998).

In regards to specific posture categories, two correlation coefficients were not significant at the 0.05 level (Table 5.8). Significance means there is evidence to reject the null hypothesis ($r = 0$) or that there is a statistically significant relationship between the specific pair of raters evaluated. The following posture categories had non-significant correlations between inclinometry and video observation: left and right upper arm elevation > 90°. As a supplementary measure of correlation, scatter plots were created for the posture categories of the left upper arm, right upper arm, and trunk (Figures 5.16-5.18). Figure 5.16 displays scatter plots of video observation versus inclinometry for the three posture categories of the left upper arm: 0°-44° elevation, 45°-90° elevation, and > 90° elevation. The posture categories of 0°-44° elevation and 45°-90° elevation for the left arm demonstrated high positive correlation, while > 90° elevation for the left arm demonstrated virtually no correlation (Figure 5.17 and Table 5.8). Figure 5.17 displays scatter plots of video observation versus inclinometry for the three posture categories of the right upper arm: 0°-44° elevation, 45°-90° elevation, and > 90° elevation. The posture categories of 0°-44° elevation and 45°-90° elevation for the right arm demonstrated high positive correlation, while > 90° elevation for the right arm demonstrated low positive correlation (Figure 5.17 and Table 5.8). This was similar to the findings for the left arm. Figure 5.18 displays scatter plots of video observation versus inclinometry for the two posture categories of the trunk: < 45° inclination and ≥ 45° inclination. Both posture categories of the trunk demonstrated high positive correlation (Figure 5.18 and Table 5.8).

TABLE 5.8

Inter-Method Reliability Results Using Pearson Product Moment Correlation

	r-Coefficients		
Body Part	**0°-44° Arm Elevation or < 45° Trunk Inclination**	**45°-90° Arm Elevation or ≥ 45° Trunk Inclination**	**> 90° Arm Elevation**
Left Upper Arm	0.79*	0.81*	0.02
Right Upper Arm	0.88*	0.78*	0.44
Trunk	0.81*	0.82*	-

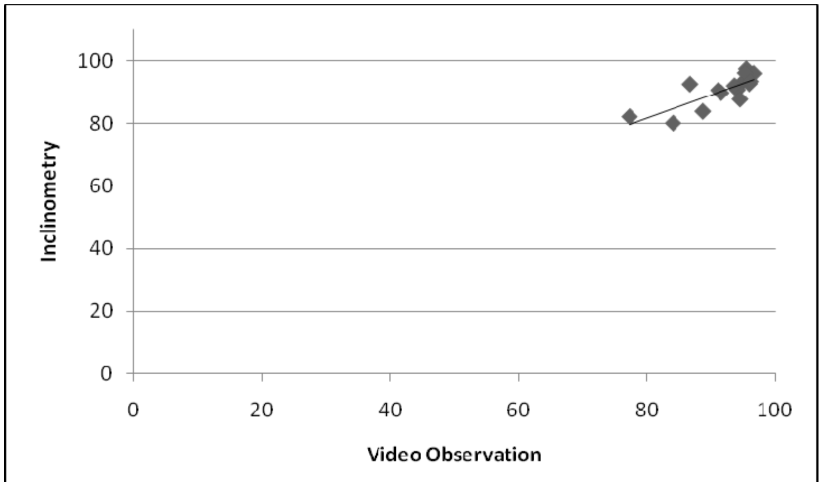
*Indicates significance > 0.05

FIGURE 5.16

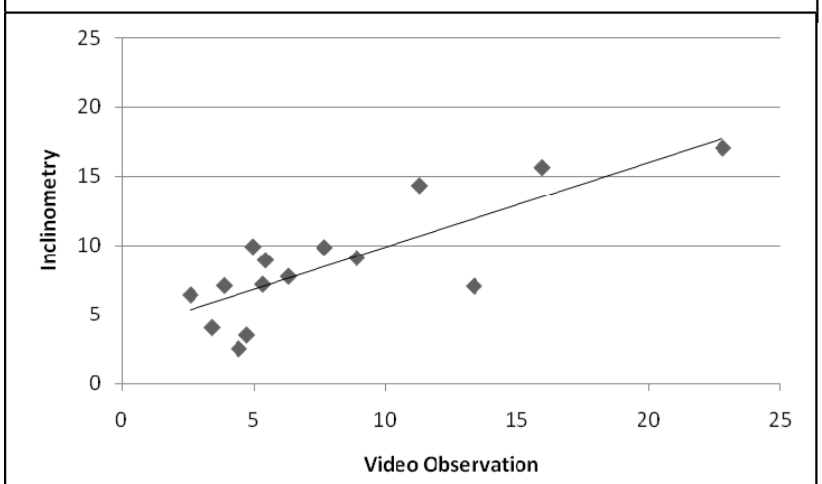
Scatter Plots of Video Observation versus Inclinometry

for Left Arm Elevation Postures

0°-44°



45°-90°



> 90°

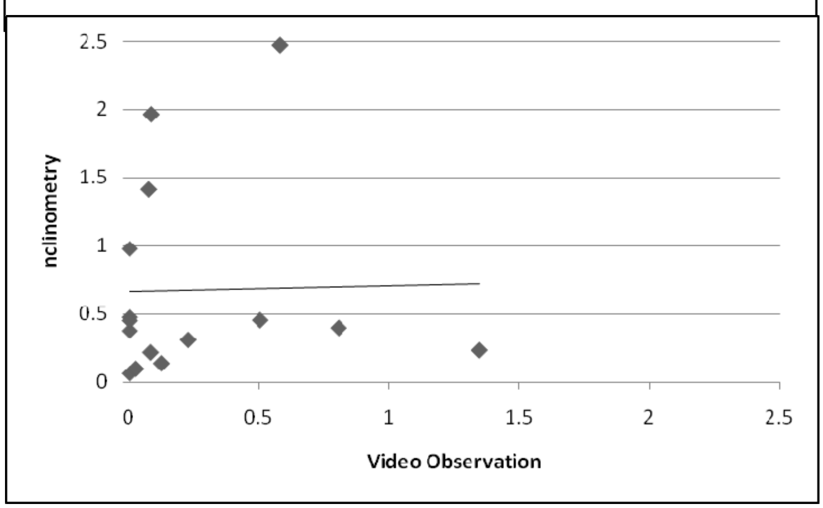


FIGURE 5.17

Scatter Plots of Video Observation versus Inclinometry

for Right Arm Elevation Postures

0°-44°

44°-90°

> 90°

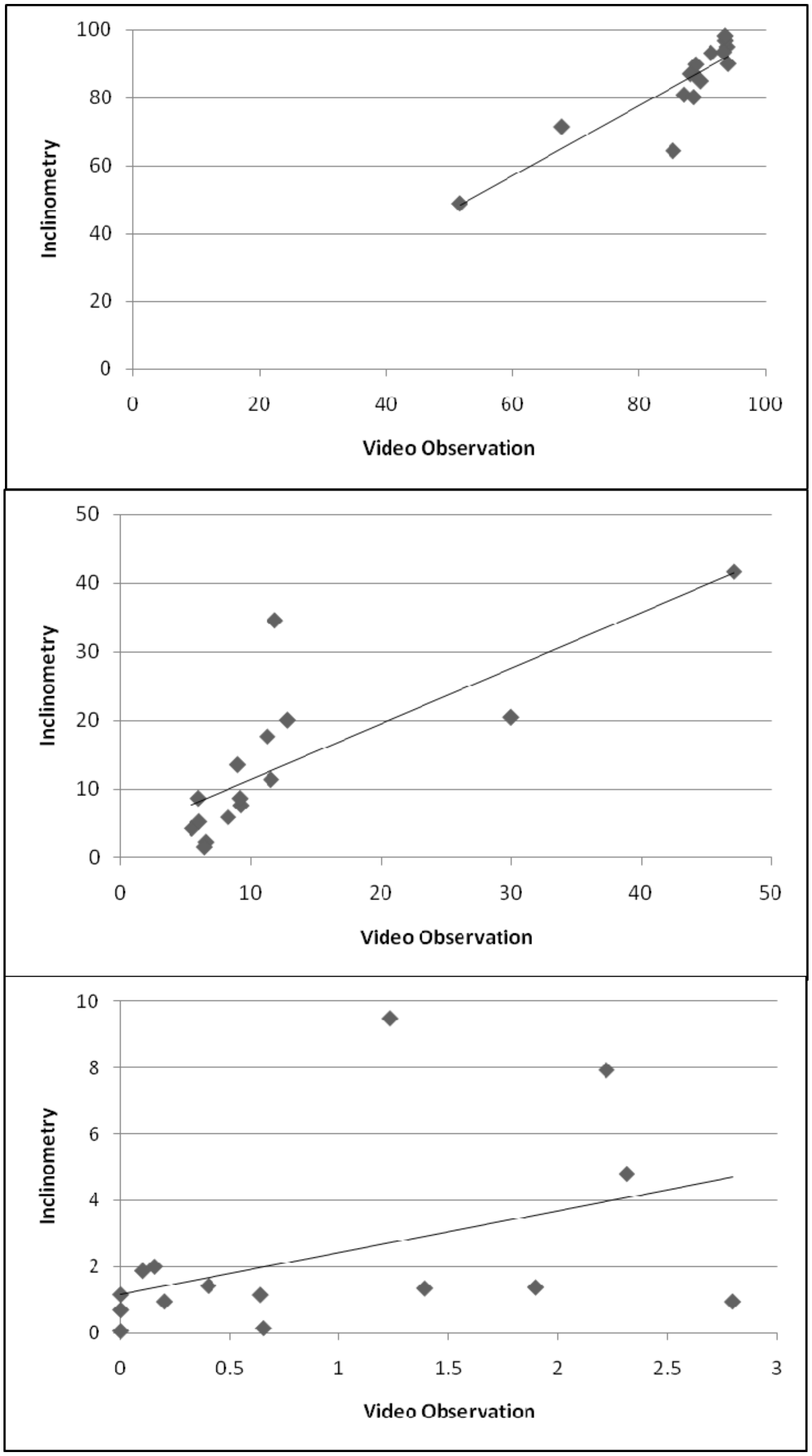
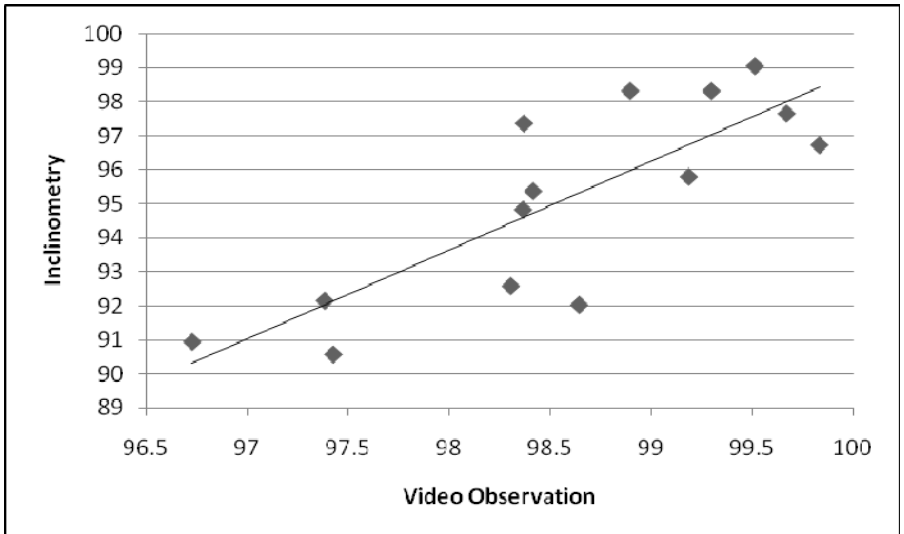


FIGURE 5.18

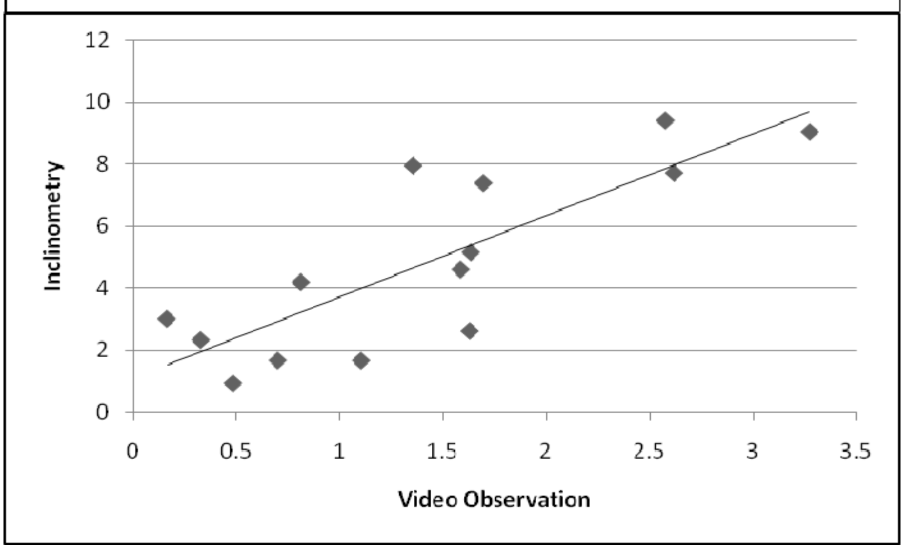
Scatter Plots of Video Observation versus Inclinometry

for Trunk Inclination Postures

< 45°



≥ 45°



Results of the analysis of variance computed for the inter-method reliability assessment using Generalizability Theory are provided in Table 5.9. Variance component estimations as well as the percent of total variance for each variance component are provided. Explanations for each variance component are provided in Table 5.2. The percent of total variance provides insight as to which variables accounted for the largest amounts of variability. For the purposes of variance regarding assessment of postures, it is optimal for most of the variability to occur in the 'subjects' variance component. Variability within the other variance components leads to decreased reliability. Variability in the 'residual' variance component is undesirable because it is difficult to postulate the cause or causes of this variability. The right upper arm had higher overall variance than both the left upper arm and trunk (Table 5.9). This was expected since the majority of the subjects recruited were right hand dominant and expected to use their right upper arm more often than their left upper arm while performing tasks. Trunk variance was low compared to the right and left upper arms. Based on observations made during the data collection, this was expected since most of the work areas did not require trunk inclination $\geq$ 45° for the majority of the tasks performed. Variance estimates for the st,e variance component composed 27% of the total variance when averaged across all body parts and postures. The highest variance estimates for the st,e variance component were found in the > 90° elevation category for the upper arms.

Inter-method reliability coefficients computed using Generalizability Theory, for the posture categories of the left and right upper arms and the trunk across the two measurement tools are provided in Table 5.10. Right upper arm elevation of 0°-44° had the highest generalizability coefficient while left upper arm elevation > 90° had the lowest. Generalizability coefficients obtained for the trunk postures were much lower than the Pearson correlation coefficients and the right upper arm elevation > 90° Generalizability coefficient was half the value as the Pearson correlation coefficient.

TABLE 5.9

Inter-Method Reliability Variance Estimates for Postures of the Left and Right Upper Arms and Trunk

Body Part	Posture	Source of Variation	Estimated Variance Component	% Total Variance
Left Arm	**0°-44° Elevation**	subjects (s)	25.063	79.8
		tools (t)	0.422	1.3
		st,e	5.934	18.9
	45°-90° Elevation	subjects (s)	21.886	80.3
		tools (t)	0.000	0.0
		st,e	5.377	19.7
	> 90° Elevation	subjects (s)	0.174	30.4
		tools (t)	0.064	11.2
		st,e	0.333	58.4
Right Arm	**0°-44° Elevation**	subjects (s)	148.471	87.4
		tools (t)	0.876	0.5
		st,e	20.572	12.1
	45°-90° Elevation	subjects (s)	117.021	81.2
		tools (t)	0.000	0.0
		st,e	27.110	18.8
	> 90° Elevation	subjects (s)	2.799	41.0
		tools (t)	0.799	11.7
		st,e	3.228	47.3
Trunk	**< 45° Inclination**	subjects (s)	3.498	29.6
		tools (t)	5.744	48.6
		st,e	2.570	21.8
	≥ 45° Inclination	subjects (s)	3.510	29.9
		tools (t)	5.656	48.2
		st,e	2.566	21.9

TABLE 5.10

Inter-Method Reliability Results Using Generalizability Theory

Body Part	G-Coefficients		
	0°-44° Arm Elevation or ≤ 44° Trunk Inclination	45°-90° Arm Elevation or > 44° Trunk Inclination	> 90° Arm Elevation
Left Upper Arm	0.78	0.78	0.02
Right Upper Arm	0.87	0.79	0.25
Trunk	0.29	0.29	-

Results of the repeated measures ANOVA are presented in Table 5.11. The effect of measurement tool was significant for right upper arm elevation > 90°, trunk inclination < 45°, and trunk inclination > 45°. Significance indicates that the percent time of posture recorded for right upper arm elevation > 90°, trunk inclination < 45°, and trunk inclination > 45° did change over the measurement tools, video observation and inclinometry. The effect of work area was significant for all three right upper arm posture categories. Significance indicates that the percent time of posture recorded for the right upper arm postures did change over the work areas. No significant effects were found for the tool*work area interaction. This non-significance for the tool*work area interaction indicates that the percent time of posture recorded for all postures and body parts as recorded by the two measurement tools did not change for the different work areas.

TABLE 5.11

Repeated Measures ANOVA Results by Measurement Tool

Body Part and Posture	**DF**	**Mean Square**	***p*-value**
Left Upper Arm Elevation 0°-44°			
Tool	1	13.19	0.18
Work Area	2	53.53	0.37
Tool x Work Area	2	2.39	0.70
Left Upper Arm Elevation 44°-90°			
Tool	1	4.27	0.39
Work Area	2	38.65	0.45
Tool x Work Area	2	6.21	0.35
Left Upper Arm Elevation > 90°			
Tool	1	1.20	0.08
Work Area	2	0.42	0.32
Tool x Work Area	2	0.31	0.43
Right Upper Arm Elevation 0°-44°			
Tool	1	35.72	0.22
Work Area	2	960.63	0.02*
Tool x Work Area	2	17.40	0.46
Right Upper Arm Elevation 44°-90°			
Tool	1	6.26	0.67
Work Area	2	664.90	0.04*
Tool x Work Area	2	3.33	0.90
Right Upper Arm Elevation > 90°			
Tool	1	15.03	0.05*
Work Area	2	24.26	0.00*
Tool x Work Area	2	4.14	0.30
Trunk Inclination < 45°			
Tool	1	79.05	0.00*
Work Area	2	13.97	0.13
Tool x Work Area	2	3.18	0.31
Trunk Inclination ≥ 45°			
Tool	1	77.99	0.00*
Work Area	2	13.91	0.14
Tool x Work Area	2	3.10	0.32

* Indicates significance at $\alpha = 0.05$

Results of the Levene's test for homogeneity of variance varied depending on the measurement tool, posture, and body part. Evaluation of the eight postures evaluated across the three body parts (left/right upper arms and trunk) and across the two measurement tools resulted in16 test statistics. Of the 16 test statistics, six were significant at the 0.05 level, meaning that the three work areas had unequal variances for six of the test statistics. The six significant test statistics were: left upper arm elevation 0°-44° for both measurement tools, left upper arm elevation 45°-90° for video observation, right upper arm elevation 0°-44° and 45°-90° for video observation, and right upper arm elevation > 90° for the inclinometer. Note that some of the statistics were only significant for one of the measurement tools indicating that for the same posture and body part, the error variance of the two measurement tools was different across work areas. This supports that for particular postures and body parts, the percent time of recorded posture by the two measurement tools varied across work areas.

The Hartley's test was applied to the data differently than the Levene's test. The two measurement tools, video observation and inclinometry, were compared within each work area (bottling, canning, and kegging). This resulted in eight test statistics (one for each posture category within the three body parts) for each work area with a total of 24 test statistics. For the present study, 15 degrees of freedom for the numerator and denominator and a significance level of 0.05 were used to obtain the critical F value for left and right upper arm postures while 14 degrees of freedom for the numerator and denominator and a significance level of 0.05 were used to obtain the critical F value for the trunk. This resulted in a critical F value of 2.4 for the upper arms and 2.48 for the trunk. Seven of the eight F_{max} statistics for the bottling area, five of the eight F_{max}

statistics for the kegging area, and 4 of the F_{max} statistics for the canning area had values greater than the critical F value. All six F_{max} statistics for the trunk, six of the nine F_{max} statistics for the right upper arm, and four of the nine F_{max} statistics for the left upper arm had values greater than the critical F value. Both of the trunk postures had F_{max} statistics greater than the critical F value for all three of the work areas. Based on these results, it was concluded that the standard deviation of the percent time recorded by inclinometry for trunk postures was significantly greater than the standard deviation of the percent time recorded by video observation for trunk postures for all three work areas. Left and right upper arm posture > 90° had Fmax statistics greater than the critical F value for the bottling and kegging work areas. Based on these results, several conclusions were made. The standard deviation of the percent time recorded by inclinometry for right upper arm posture > 90° was significantly greater than the standard deviation of the percent time recorded by video observation for right upper arm posture > 90° for the bottling and kegging work areas. This was the same conclusion made for left upper arm posture > 90° for the bottling work area. However, for upper arm posture > 90° for the kegging work area, the standard deviation of the percent time recorded by video observation was significantly greater than the standard deviation of the percent time recorded by inclinometry.

DISCUSSION

This study evaluated the inter-method reliability of assessing upper limb postures of workers performing manufacturing tasks using a video-based observation technique and inclinometry. Inter-method reliability measures the ability of ‘different instruments which measure the same underlying exposure to yield similar results on the same subjects' (Armstrong et al., 1994). Based on the overall results of the present study, the researchers concluded that the percent of time in recorded in a posture category by video observation and inclinometry was similar and most of the inter-method reliability coefficients were moderate to high (> 0.50) (Tables 5.8 and 5.10). There were several exceptions to these overall conclusions. Pearson correlation coefficients (Table 5.8) where greater than or equal to 0.79, except for left and right upper arm elevation > 90°, which were 0.02 and 0.44, respectively. In addition to the low values, these correlation coefficients were found to be significantly different at an alpha of 0.05. The Generalizability intraclass correlation coefficients (Table 5.10) were lower overall than the Pearson correlation coefficients. The Generalizability coefficients (Table 5.10) were greater than or equal to 0.78 for left and right upper arm postures of 0°-44° and 45°-90°, while the other coefficients were below 0.30 and left upper arm elevation > 90° was extremely low at 0.02 indicating almost no correlation.

The repeated measures ANOVA was used to compare the results as obtained by video observation and inclinometry based on the three work areas (bottling, kegging, and canning) and to investigate whether any differences existed. The repeated measures ANOVA (Table 5.11) revealed that the percent time recorded in right upper arm elevation > 90°, trunk inclination < 45°, and trunk inclination ≥ 45° changed significantly between video observation and inclinometry. These findings correspond to the Generalizability coefficients except for left upper arm elevation > 90°, which had very low Generalizability and Pearson coefficients, but was not found as significant at the 0.05 level in the repeated measures analysis. The percent time recorded for all right upper arm postures changed significantly between the work areas. This was expected since the percent time recorded in right upper arm posture 0°-44° for the kegging work area was on average 14.4% less than the bottling and canning work areas, 14.2% more for right upper arm posture 45°-90°, and 2.6% more for right upper arm posture > 90°. While the percent times for the right upper arm postures were similar in the bottling and canning, they were different in kegging therefore leading to the significant finding of the work area factor in the repeated measures ANOVA. This also demonstrated that the kegging work area required more right upper arm posture > 45° than the other work areas. Since none of the interactions between tool and work area were significant, the researchers concluded that the percent time recorded for all postures and body parts by the two measurement tools did not change over the different work areas.

The Levene's test expanded the results of the repeated measures ANOVA and demonstrated that the error variance of the percent time recorded by the two measurement tools varied across the three work areas for particular postures and body parts. The Levene's test also revealed that the error variance for the two measurement tools was different for right upper arm elevation 0°-44°, left and right upper arm elevation 45°-90°, and right upper arm elevation > 90°. This corresponded to the findings of the repeated measures ANOVA. The Hartley's test demonstrated different results than those found in the Levene's test. Based on the calculated F statistics, the researchers concluded that the standard deviation of the percent time recorded by inclinometry for trunk postures was significantly greater than the standard deviation of the percent time recorded by video observation for all three work areas. This corresponded to the findings of the repeated measures ANOVA. The results from the repeated measures ANOVA, Levene's test, and Hartley's test demonstrated that while the two measurement tools may have had acceptable correlation coefficients for a specific posture, the two measurement tools varied significantly in the standard deviation or dispersion for particular postures and recorded percent time in posture more or less often than the other.

Possible Sources of Error Between the Measurement Tools

Differences in results found between the video observation and inclinometry could be contributed to multiple sources of error. These possible sources of error included error associated with the statistical methodology, error associated with the inclinometers, and error associated with the video observation measures.

Statistical Methodology

Some of the differences between the Pearson correlation coefficients and the Generalizability could be attributed to the statistical methods themselves. Statistical methods used to assess inter-method reliability of posture variables are similar or the same to those used to evaluate rater reliability. These methods included percentage of agreement, the Pearson Product moment correlation coefficient, repeated measures ANOVA, the Student's t-test, Cohen's Kappa statistic, and the Spearman Rank correlation coefficient (van Eerd et al., 2009; Hansson et al., 2001; Juul-Kristensen et al., 2001; Spielholz et al., 2001; Burdorf et al., 1992; Burdorf, 1995). The intraclass correlation coefficient (ICC) has been used in rater reliability studies (Bao et al., 2009; Dartt et al., 2009; Lowe, 2004(a); Burt and Punnett, 1999; Genaidy et al., 1993; Stetson et al., 1991), but not typically used to evaluate inter-method reliability. The ICC is used to evaluate continuous data and is conceptualized as the ratio of between-groups variance to total variance.

Generalizability Theory, as applied in the present study, produces a correlation coefficient analogous to the intraclass correlation coefficient. Generalizability theory extends the use of reliability measures, such as percentage of agreement, kappa, and correlation, by considering multiple sources of error simultaneously and allows a more accurate assessment of the measurement situation (VanLeeuwen, 1997). Therefore, one can estimate the magnitude of each source of error separately in a single analysis and use this information to optimize the reliability of the measurement (Shavelson and Webb, 1991). Generalizability theory distinguishes between relative and absolute decisions, in the same manner as computing classical intraclass correlation coefficients (Shavelson and Webb, 1991). Absolute agreement measures whether raters assign the same absolute score and is used when systematic variability due to raters is relevant. Relative agreement, or consistency, considers raters consistent as long as their relative ratings are similar. In addition to providing an intraclass correlation coefficient, or Generalizability coefficient, Generalizability theory utilizes the variance components estimated to design a more efficient and effective measurement procedure to be used in the future. For example, a researcher can estimate the number of raters necessary to achieve a particular level of reliability based on pilot reliability assessment.

One concern with using the ICC and the same for Generalizability Theory is that provided there is the same magnitude of variations among raters' estimates, a posture with smaller variation will result in lower correlation coefficients as compared with postures with larger variations (Bao et al., 2009). In the present study, this same phenomenon tended to occur. Left and right upper arm elevation > 90° and both trunk postures, which all had smaller variance than the other postures evaluated, had lower

Generalizability coefficients (0.02, 0.25, and 0.29, respectively) (Tables 5.9-5.10). However, one should not assume that lower variance will always lead to lower reliability coefficients. For example, left upper arm elevation of 0°-44° and 45°-90° had lower variance compared to the same postures of the right upper arm, but the resulting reliability coefficients were similar (Tables 5.9-5.10). This demonstrates the importance of evaluating the specific variance components rather than evaluating variance in general. In addition, it uncovers the strength of using Generalizability Theory as compared to the results obtained using the Pearson correlation coefficient. Based on the present study, four sources of variance were identified: differences among the subjects, differences among the measurement tools, the subject by tool interaction, and random or unidentified events. Variance estimates indicated that the majority of variance was found within the subjects, with little variance found within the measurement tools, except for trunk postures. In addition, variance estimates indicated some variance in the st,e variable that represents variance for the interaction between the subjects and tools as well as unidentified residual variance. The interaction represents inconsistencies of the tools recorded postures for a particular subject.

Like the ICC and Generalizability Theory, the Pearson correlation coefficient is expected to be lower when there is a variance restriction. If the variance is truncated or restricted, attenuation of the correlation coefficient can occur. In the present study, the low Pearson coefficients for left and right upper arm elevation > 90° could be partly attributed to the lower variance or the amount of time recorded for these two categories. Left and right upper arm elevation > 90° was recorded for 1.1% of the total analyses on average for both measurement tools with a range of 0% to 9.5% (Tables 5.5-5.6 and

Figures 5.9-5.10). This was lower than the other arm elevation postures, potentially leading to the lower Pearson correlation coefficient. However, trunk inclination ≥ 45° had a similar percent time recorded (3.1% recorded on average for both measurement tools), but still had a Pearson correlation coefficient of 0.82 (Tables 5.7 and 5.9). Another statistical cause for the reduced coefficients could have been the binning of the continuous data into set of categories. Another statistical explanation for the discrepancies between the Pearson coefficients and Generalizability coefficients are the actual structures of the statistics. Calculation of the G-coefficient (ICC) centers and scales the data using a pooled mean and standard deviation, whereas the Pearson correlation centers and scales each variable by its own mean and standard deviation.

If the present study would have simply reported the Pearson correlation coefficients, both postures of the trunk would have been concluded to demonstrate high correlation. When interpreting the Generalizability coefficients, the opposite conclusion would have been made, since the Generalizability coefficients were low for the trunk postures. In addition, while the Pearson and Generalizability coefficients were the same (0.02) for left upper arm elevation > 90°, the Generalizability coefficient was nearly half the magnitude of the Pearson coefficient for right upper arm elevation > 90°. This demonstrates that the statistical method used can result in different conclusions. Previous research using similar statistics to assess reliability have reported that results may differ greatly depending on the statistic used (Bao et al., 2009; Burt and Punnett, 1999).

Plots (Figures 5.12-5.15) of inclinometry and video observation for each body part and posture measured demonstrated that while the overall correlation was high for most postures, specific data points (subjects) had much different results. For example, when evaluating right upper arm elevation of 45°-90°, the inclinometer recorded approximately 35% in this posture while the video observation tool recorded approximately 12% for subject number eight. Upon review of the of the video observation and notes maintained by the analyst, it was determined that the right upper arm was occasionally occluded from camera view by stacks of kegs and a fork truck moving in and out of the area. The task performed by subject number eight primarily consisted of loading empty kegs onto the kegging line. It was difficult for one camera view to adequately capture the nature of this task without having blocked views of the right arm. Findings were similar for the other right upper arm postures for subject number eight.

Direct Instrumentation and Observation

A study by Spielholz et al. (2001) reported Pearson correlations of 0.07-0.33 and mean agreement values of 0.49-0.82 for analyses of wrist flexion/extension, wrist deviation, and forearm rotation using video analysis versus electrogoniometry of 18 workers performing three different jobs. Wrist deviation was the only posture category comparison between video analysis and electrogoniometry that did not show a significant difference. While correlation coefficients were higher for most postures in the present study, this was expected since the postures evaluated in the present study involved postures of larger body segments and angular deviations as compared with the wrist. Decreased reliability due to smaller joint movements has been found as a contributing factor to error when estimating postures from observation (Keyserling, 1986; Stetson et al., 1991; Burt and Punnett, 1999).

Juul-Kristensen et al. (2001) reported mean percent time differences > 20% for upper arm elevation between video observation and inclinometry for 21 workers performing two poultry processing jobs. Correlation coefficients were not presented, so comparison between studies was difficult. However, mean percent time spent in the upper arm postures as reported in the present study did not differ as greatly as the numbers reported by Juul-Kristensen et al. (2001). Differences could also be attributed to the arm postures assessed. The Juul-Kristensen et al. (2001) study evaluated upper arm flexion/abduction of 0º-30º, 30º-60º, and > 60º whereas the present study evaluated upper arm flexion/abduction of 0º-44º, 45º-90º, and > 90º. Previous research of reliability of observation of posture found moderate to high reliability for estimation of shoulder postures > 90° (Bao et al., 2009; Dartt et al., 2009; Lowe, 2004a; Ketola et al., 2001; Burt and Punnett, 1999; Stetson et al., 1991; Keyserling 1986). Therefore, the higher inter-method reliability reported in the present study could be attributed to the posture ranges evaluated.

Burdorf et al. (1992) reported moderate Spearman rank correlation coefficients for bent trunk postures > 20º of 30 workers performing either a sedentary or dynamic work task. However, the authors (Burdorf et al., 1992) found considerable differences between direct observation and inclinometry for individual data points. While findings were similar for the present study, different statistical methods were used so direct comparison was difficult. In addition, the postures evaluated were different. Burdorf et al. (1992) evaluated bent trunk postures > 20º, while the present study evaluated trunk inclination of < 45º and ≥ 45º. Therefore, differences could be attributed to differences in the postures evaluated. A previous study conducted by Bao et al. (2009) examining inter-rater reliability of observation of postures from video concluded that wider posture category widths usually resulted in better proportions of agreement compared with smaller posture category widths and that 30° was the appropriate minimal posture category width for postures of the upper arms, neck , trunk, elbows, forearms, and wrists.

In regards to measurement error in the inclinometers used in the present study, the researchers attributed several factors that could have caused differences: errors in attachment of the inclinometers, errors in the set-up and taring of the inclinometers, sampling rate discrepancies between inclinometer devices, and movement in inclinometer attachment throughout the sampling period. The inclinometers were mounted directly to the skin at approximate locations. For the upper arms, the inclinometers were mounted to the posterior aspect of the upper arm, mid-way between the shoulder and elbow. Inclinometers mounted lower or higher than mid-way between the shoulder and elbow in addition to differences in upper arm morphology may have changed the position of the inclinometers relative to gravity potentially contributing to differences between the inclinometers and video observation. However, once mounted, the inclinometers were tared. Therefore, the researchers did not expect that slight mounting differences led to large differences between the inclinometry and video observation.

While not performed for the upper arms, a preliminary study, as described in the methods section, was performed by the researchers of the present study that evaluated the implications of mounting error when mounting the inclinometers at the T6 spinous process. Based on this limited preliminary study, results of mounting the inclinometer at the T5, T6, and T7 spinous processes found that 95% of the means for both trunk posture categories were no more than 2.15% (SD 1.96) from the mean. Based on the limited preliminary study, the researchers concluded that mounting differences on the trunk could change the outcome of the inclinometers, but that mounting errors within one spinous process are not likely to cause significant changes in posture outcome. However, the mounting error could be exacerbated by the length of sampling time. The researchers of the present study recommend further research into the implications of small attachment discrepancies on posture outcomes and how these discrepancies could change based on the sampling duration.

Errors in taring and set-up of the devices could have been another source of error that created differences between the inclinometry and video observation. During set-up, the three inclinometers were synced and launched using three laptop computers. Since launching was performed manually by the researchers, there could have been slight errors in the launch process. However, launch errors were expected to be within one second and since the inclinometers sampled at approximately 7.6 Hz, one second errors should not have affected the data. Taring of each inclinometer was performed after the inclinometers were attached to the subjects. Previous research attributed differences between video observation and inclinometry on using different reference positions (Juul-Kristensen et al., 2001). Juul-Kristensen et al. (2001) attributed differences based on the observation method using the arm hanging relaxed along the side of the body as the reference position whereas for the inclinometer, the arm was positioned along the line of gravity with a 2 kg dumbbell. The researchers (Juul-Kristensen et al. (2001) adjusted the reference positions statistically *a posterior* to investigate the change in reliability. Based on these findings (Juul-Kristensen et al., 2001) the researchers of the present study ensured that reference positions between video observation and inclinometry were the same to avoid error in the set-up and taring process. Based on these findings, the researchers of the present study recommended that reference positions between data collection tools be as similar as possible and that holding the arms relaxed at the side is a better reference position for inclinometry than using a weight to align to gravity.

Another possible source of error related to the inclinometers could have been sampling frequency discrepancies. The Microstrain Virtual Corset manual (MicroStrain, Inc; Williston, VT) indicated two programmable sampling frequencies: 7.5 Hz and 15 Hz. However, Trask et al. (2006) reported a 7.6 Hz sampling frequency. Based on this discrepancy, the researchers performed a small preliminary study to examine the 'true' sampling frequencies of the inclinometers since the sampling frequency used could affect the results of the data analyzed using the LabView 8.6 software (National Instruments). Two inclinometers were tested over a known sampling period, as discussed previously in the methods section. The mean sampling frequency over the timed trials was 7.59 Hz and corresponded to the Trask et al. (2006) study. Based on this finding, the researchers of the present study utilized a sampling frequency of 7.6 Hz to analyze all results to minimize error in sampling frequency. It should be noted that not all of the inclinometers used in the study were tested; therefore, it cannot be verified that all inclinometers sampled at 7.6 Hz. It could also not be verified that the sampling frequency remained constant over the sampling periods. Based on these findings, the researchers of the present study concluded that sampling frequency discrepancies could have affected the inter-method reliability results. Therefore, it is recommended that all research verify sampling frequencies of the devices used to ensure proper data analysis and findings. While a 0.01 Hz discrepancy only caused a 2.5 second difference between 7.5 Hz and 7.6 Hz over the five-minute test sample periods, this would result in a 5.1 minute discrepancy for an entire shift. While 5.1 minutes does not seem to be a lot of time, it could have affected the present study since only five-minute periods were extracted from the data. This could have resulted in different five-minute periods from the inclinometry as compared to the video observation. This discrepancy or error was not found in previous research.

Movement of the inclinometers during the sampling period could have also been a source of error in the inclinometry results. While the researchers don't believe that this was an issue in the present study, it cannot be disregarded since it was not assessed or quantified. Information regarding the attachment of inclinometers throughout long sampling periods could not be identified in the research literature. Initial trials of attaching the inclinometers, as discussed previously in the methods, utilized armbands and harnesses. Trask et al. 2006 and 2007) used a combination of harnesses to mount the inclinometers. Pilot data collected by the present researchers indicated that the armbands and harnesses shifted positions throughout the sampling periods, affecting the data collected. It was then determined that a combination of athletic wraps and Tegaderms (3M: St. Paul, MN) could better secure the inclinometers with little to no movement during sampling. While the researchers of the present study concluded that the athletic wraps and Tegaderms were a better attachment device than armbands and harnesses, there was still potential for some movement during the sampling periods. Visual inspection of the inclinometers at the end of the sampling periods did not indicate shifting of the devices, but since this possible error was not quantified, it still must be considered a reason for differences between the video observation and inclinometry.

In regards to measurement error in the video observation tool and techniques used in the present study, the researchers attributed several factors that could have caused differences which included: occluded views, parallax error, the dynamic character of the work, and application of the definitions used to estimate postures as compared with the inclinometry. Spielholz et al. (2001) listed occluded views and parallax error noted during the video observation analysis as possible sources of error between the video observation and direct measurements obtained using electrogoniometers. While the present study did not use electrogoniometers or measure wrist postures, occluded views and parallax error were noted by the video analyst as reasons for difficulty in estimating postures of the upper arms and trunk. Occluded views and parallax error are two common issues that have been recorded in previous research investigating observation of postures (Bao et al., 2009; Dartt et al., 2009; Spieholz et al., 2001; Douwes and Dul, 1991).

For optimal estimation of postural angles, the plane of the joint movement estimated should be orthogonal to the camera (Douwes and Dul, 1991). This allows the observer to analyze posture from a perpendicular view to the plane of interest for a particular posture or anatomical area. Deviations from this perpendicular view may lead to parallax errors. During video analysis of certain subjects, the analyst had difficulty estimating postural angles to specific degrees when the viewing angles were inadequate. At times, equipment, products, or fork trucks blocked views upper arms and trunk. Douwes and Dul (1991) found that differences between direct and indirect measurements of posture were within 3° if the viewing angle of the posture was no more than 5° deviated from the perpendicular view of the body part. They also found that under conditions with more than 60° deviation from a perpendicular viewing angle, agreement between direct and indirect measurements differed by more than 9°. Liu et al. (1991) reported an estimated error of up to 10 due to parallax. Based on findings of the present study and previous studies, the researchers concluded that occluded views and parallax error may have contributed to the variation in estimates between the video observation and inclinometry.

The dynamic character of the tasks performed in the three different work areas has been used as an explanation for differences between upper arm postures for observation versus inclinometry in previous studies (Juul-Kristensen et al., 2001; Ericsson et al., 1993; Keyserling, 1986). A previous study has reported a tendency to overestimate angles closer to 0º and underestimate angles closer to the extreme (Genaidy et al., 1993). Since correlation coefficients were lower for upper arm postures > 90º, the tendency to underestimate angles could have attributed to the differences between video observation and inclinometry; however, the present study demonstrated that the inclinometers registered a larger amount of percent time for upper arm elevation > 90º for both arms. This could mean that there was an underestimation of posture closer to the extreme, contrary to previous research (Genaidy et al., 1993). Results of the Levene's test and Hartley's test support the conclusion that the dynamic character of the tasks performed in the three work areas could have affected the inter-method reliability. For example, right upper arm elevation > 90º was found to have unequal variance across the three work areas and was also found to have a low correlation coefficient. Therefore, the low correlation coefficient could be explained by the more dynamic nature of one or more of the work areas since significant variance was found between the work areas.

The application of the definitions used to estimate postures for the video observation as compared with the inclinometry could have also affected the inter-method reliability. Previous studies comparing observation with inclinometry have reported different definitions of estimating postures as a possible source of error (Juul-Kristensen et al., 2001; Burdorf et al., 1992). As described previously, Juul-Kristensen et al., 2001) evaluated work postures using video observation and direct technical measurements mounted inclinometers on the left and right upper arms of 21 women performing two primary jobs in the poultry processing industry. The superior edge of each inclinometer was mounted at the origin of the deltoid muscle and the the posterior edge of the each inclinometer was mounted at a line from the acromion and laterial epicondyle. The researchers (Juul-Kristensen et al., 2001) concluded that shoulder postures were measured differently by video observation as compared with the inclinometry.

While the present study cannot be directly compared to the Juul-Kristensen et al. (2001) study, the definitions used to estimate posture versus the location of the inclinometers for the upper arms could have affected inter-method reliability even though present study attempted to minimize this error through precise definitions and video observation estimation techniques. In the present study, the inclinometers were mounted to the posterior aspect of the upper arm mid-way between the shoulder and elbow. An imaginary line drawn from the acromion through the mid-humerus to the elbow was used to estimate upper arm elevation from video. Zeroing the inclinometers to the subjects was expected to reduce this error; however, this was not quantified and cannot be concluded definitively. Due to the low correlation coefficients in upper arm elevation > 90º, it was concluded that definitions of posture estimation differences between video observation and inclinometry affected the extreme upper arm postures more than the postures closer to 0º.

Burdorf et al. (1992) attributed the definitions of the angles of trunk bending being applied differently between direct observations and inclinometry as a source of error to the location of the inclinometer on the back. In the Burdorf et al. (1992) study, the trunk inclinometer was mounted at the L2/L3 spinous processes, whereas the present study mounted the inclinometer at the T6 spinous process. Burdorf et al. (1992) explained that trunk flexion as recorded using OWAS was defined as an angle of 20º between the straight line through the pelvis and shoulders and a vertical line. In using this method, flexion of both the thoracic and lumbar spine contribute to the observed angle. A study by Burton (1986) reported that measurements of trunk inclination at the T12 and L4 showed more trunk flexion at T12 than T4. The inclinometry obtained by Burdorf et al. (1992) at the L2/L3 focused on the position of the lumbar section of the spine in order to be representative of the position of the trunk as a whole. One would then expect that the inclinometry would systematically underestimate the angle of trunk bending and therefore presented lower estimates as compared with the direct observation (Burdorf et al., 1992). However, Burdorf et al. (1992) concluded the opposite since their study results did not indicate underestimation. Inter-method reliability coefficients for the present study were higher than those found by Burdorf et al. (1992). The percent time recorded by the trunk inclinometer for the present study was lower than video observation for trunk inclination $< 45^{\circ}$ and thereby higher than video observation for trunk inclination $\geq 45^{\circ}$. While possible, the researchers of the present study did not contribute location of the trunk inclinometer as compared to the video observation estimation of trunk inclination as a major source of error since posture estimation from video was aligned with the location of the inclinometer on the trunk.

CONCLUSIONS

Video observation is a qualitative method used to assess ergonomic risk factors such as awkward postures. Inclinometry is a quantitative direct measure also used to assess awkward postures. Ultimately, the goal of ergonomic practice and research when evaluating exposure to ergonomic risk factors such as awkward postures is to apply exposure assessment measurement tools that are not only reliable and valid, but also efficient, low in cost, and easy to use; and apply to large populations and in all work environments. Since various exposure assessment methods ranging from questionnaires to direct observation to video observation to direct instrumentation have been used and are currently used to assess exposures to awkward postures, it is important to verify that these exposure assessment tools are measuring the same variable (inter-method reliability). This provides further improvement of ergonomic exposure assessment methods, increases the opportunity to compare research studies that have used different exposure assessment methods, and provides researchers with additional information regarding which assessment method is most appropriate for their research studies.

This study demonstrated generally moderate to high inter-method reliability between video observation and inclinometry depending on the postures and body parts assessed and statistical method used. Assuming the inclinometry measured the 'true' postures, this study also demonstrated that video observation can have adequate external validity depending on the postures and body parts assessed. Based on the results of this study, the authors recommend using direct instrumentation for posture assessment, but that video observation can be useful if errors associated with video observation found in previous studies (Bao et al., 2009; Dartt et al., 2009; Burt and Punnet, 1999) can be minimized.

SECTION SIX

STUDY IV: SAMPLING STRATEGY

ABSTRACT

One facet of ergonomics involves exposure assessment of physical risk factors in the workplace, such as awkward postures. While exposure assessment methods vary based on research objectives, direct-measure exposure assessment instrumentation is recommended due to greater detail of the data collected and improved accuracy as compared to observation or self-reports (Juul-Kristensen et al., 2001; Spielholz et al., 2001; Winkel and Mathiassen, 1994; Westgard and Winkel, 1997). While direct instrumentation is recommended, challenges of optimal utilization, specifically how long to sample to obtain representative data, remains problematic (Burdorf and van der Beek, 1999). The purpose of this field study was to determine the length of time needed to adequately assess worker postures using inclinometry during work tasks at a brewery . Posture exposures to the upper arms and trunk were directly measured by Virtual Corset inclinometers for the full shift among workers performing manufacturing tasks. Sample durations of 4-hrs, 2-hrs, 1-hr, and 30-min were compared to full-shift exposures in an attempt identify optimal (i.e., accurate and efficient) durations of measurement. Among four durations randomly sampled from the full shift duration, results suggest that durations of 2-4 hrs measurement provide representative exposure estimates for the anatomical areas, postures, and tasks assessed. Based on these results, the investigators recommend sampling a minimum of 2 hours when evaluating posture exposures of the upper extremity and trunk among workers performing manufacturing tasks similar to those evaluated. However, shorter sampling durations may be adequate for more highly repetitive tasks.

INTRODUCTION

An essential tool of ergonomics is exposure assessment of physical risk factors in the workplace such as force, awkward posture, and repetition. The purpose of ergonomic exposure assessment are varied and the exposure assessment tools and methods can range from questionnaires to job surveys to observation to direct technical measurement. Exposure assessment of physical risk factors allows ergonomists to prioritize job areas and tasks in an effort to reduce the risk of injury/illness, increase productivity, and increase efficiency in the workplace. In addition to the workplace, ergonomics research utilizes and applies exposure assessment tools and methods to investigate associations and relationships between ergonomic risk factors and musculoskeletal outcomes.

Previous research has reported associations between exposures of the shoulder and back to awkward postures and musculoskeletal outcomes. These include increased risk of shoulder disorders with increased proportion of time with the arm above the shoulder and increased risk of low back disorders with mild trunk flexion more than 10% of the time (Punnett et al., 2000; Svendsen et al., 2004; Punnett et al., 1991). A study that evaluated the intramuscular pressure of the infra- and supraspinatus muscles in relation to arm posture > 60° found that intramuscular pressure increases in association with upper arm elevation for both shoulder muscles (Palmerud et al., 2000). The elevation plane (flexion versus abduction) also influenced the intramuscular pressure; however, the elevation angle affected it more. In addition, hand load greatly influenced the intramuscular pressure. A recent study has shown an association and increased odds between upper arm flexion greater than or equal to 45° and rotator cuff syndrome (Silverstein et al., 2008). In addition, frequency or cycle time has been associated with shoulder and back disorders. Research has demonstrated an increased risk of shoulder disorders if the upper arm is in postures above the shoulder one time or more per minute and that there is an increasing risk of being at high risk for back disorders if the number of lifts exceeds 120 per hour (Punnett et al., 2000; Marras et al., 1995).

Assessing exposure to physical risk factors can be problematic since subjective and observation-based methods are commonly used. The use of subjective and observation-based methods can lead to an over- or underestimation of exposure, thereby leading to erroneous conclusions regarding associations and relationships between ergonomic risk factors and musculoskeletal disorders (Burt and Punnett, 1999). Concerns of subjective and observation methods include reliability and validity. Direct technical measurements are used in ergonomic exposure assessment; however, much of their use historically has been in the laboratory or for short periods of time. Investigation of the existence of a causal relationship between ergonomic risk factors and musculoskeletal outcomes requires accurate and precise methods for exposure assessment (Mathiassen et al., 2003). Exposure assessment methods using direct technical instrumentation have been recommended since they offer detailed data collection with better accuracy than observations or self-reports (Juul-Kristensen et al., 2001; Spielholz et al., 2001; Winkel and Mathiassen, 1994; Westgard and Winkel, 1997). The use of inclinometry as a direct technical measure of posture has been applied in previous ergonomics research (Hansson et al., 2006; Trask et al., 2006; Moller et al., 2004; Mathiassen et al., 2003; Bernmark and Wiktorin, 2002; Allread et al., 2000) including evaluation of the reliability and validity of inclinometry (Bernmark and Wiktorin, 2002; Hansson et al., 2001) as well as exposure variability of various work environments.

Advantages of inclinometers include cost effectiveness, ease of use, (Li and Buckle, 1999) and most recently, enough memory and battery capability to allow for multi-shift measurement. One limitation of inclinometers is the inability to measure rotation. Therefore, in body segments, particularly the upper extremities which rotate around the long axis simultaneously with other movements, data must be interpreted with caution (Bernmark and Wiktorin, 2002). In addition, research involving inclinometry has been historically limited to laboratory settings (Hansson et al., 2006; Mathiassen et al., 2003; Bernmark and Wiktorin, 2002) not dynamic work situations (Li and Buckle, 1999). Recent research has utilized inclinometry in work environments outside of the laboratory (Trask et al., 2007; Moller et al., 2004; Jansen et al., 2001) and for full shift measurements (Trask et al., 2007; Trask et al., 2006; Moller et al., 2004; Jansen et al., 2001) to evaluate full-shift exposure measures, determine cost and feasibility, evaluate job enlargement, and exposure variability.

While direct instrumentation is recommended to obtain exposures to awkward postures, the challenges of optimal utilization of resources based on the goals of the study, determination of the postures to be assessed, and the sources of variation in exposure to the postures of interest remain to be a problem (Burdorf and van der Beek, 1999). Variations in exposure have been described using frequency analysis, occurrence of certain events, and by exposure variation analysis (EVA) (Mathiassen and Winkel, 1991). Variation of exposure across time has been studied less than exposure amplitudes, and typically using questionnaires (Wells et al., 2007). Fewer studies in variation of exposure across time has been attributed to the relative difficulty of measurement (Wells et al., 2007). Exposure variability information is critical in determining statistical precision of exposure estimates (Burdorf and van der Beek, 1999; Burdorf et al., 1997; Burdorf, 1995; van der Beek et al., 1995). In addition, direct instrumentation such as inclinometry, does not provide instructions on when and how to sample exposures (Gold et al., 2006). Many times, sampling of selected tasks or short durations are extrapolated to predict a full-shift exposure, while continuous full-shift measurements remain uncommon (Trask et al., 2007; Trask et al., 2006; Moller et al., 2004; Jansen et al., 2001). Exposure assessments based on short duration samples include assumptions about the measured risk factors of the shorter duration being representative of the entire exposure (Gold et al., 2006).

The purpose of the present study was to identify optimal sampling duration when obtaining inclinometry for postures of the upper arms and trunk of workers performing manufacturing tasks. Full-shift inclinometry was compared to shorter sample durations to evaluate the shorter durations as representative measures of the full-shift exposure (Trask et al., 2009). In addition, the full-shift inclinometry was used to assess exposure variability of postures of the upper arms and trunk of three major work areas in a brewing facility using full-shift using exposure variation analysis (EVA). Exposure variation analysis has been used as a data reduction method for electromyography and was applied in a similar manner in the present study. When used with the inclinometry data, EVA describes the percent time spent in specific pre-determined posture categories as well as the length of time (duration) at each posture category. This allows for measurement of multiple exposure dimensions simultaneously.

The development and improvement of ergonomic exposure assessment methods was a 1st Decade priority area outlined in the National Occupational Research Agenda

(NORA). The present study also addresses the NORA 2nd Decade 'manufacturing sector' and the 'cross-sector programs' of exposure assessment and musculoskeletal disorders.

The primary objectives of the present study were to:

- Compare full-shift inclinometry with shorter sampling durations to determine optimal sampling duration for postures of the upper arms and trunk.
- Assess exposure variability of the percent time spent in specific postures of the upper arms and trunk as well as the length of time (duration) at each posture category of workers performing manufacturing tasks using full-shift inclinometry.

The hypothesis of the present study was:

3. There would be no statistically significant difference between the full-shift inclinometry measures for the upper arms and trunk as compared to the shorter sampling durations of four hours, two hours, one hour, and thirty minutes.

METHODS

Research Design

Fifteen workers performing manufacturing tasks in a brewing facility were recruited for this study. Direct measures of upper arm and trunk postures (Table 6.1) were obtained using Virtual Corset (MicroStrain, Inc; Williston, VT) inclinometers. Inclinometry was collected on each worker performing the same task over two different shifts. Upon completion of the data collection, subsequent inclinometer analysis was completed using LabView 8.6 software (National Instruments: Austin, TX). The full-shift inclinometry was used to evaluate exposure variability and sampling duration strategies for postures of the upper arms and trunk. Figure 6.1 provides a visual schematic of the research design.

TABLE 6.1

Postures of the Shoulders and Trunk Evaluated Using Inclinometry

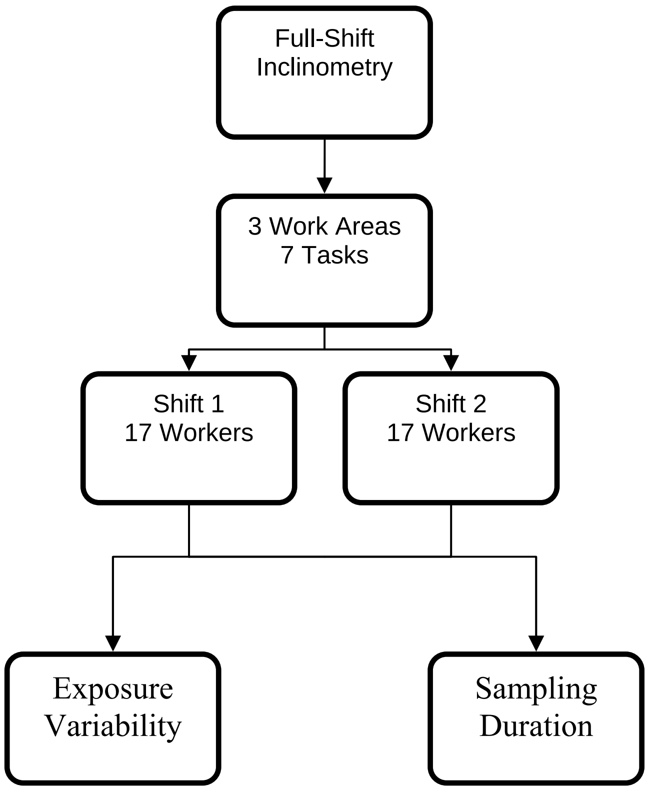
Posture Category	Upper Arms (Elevation)	Trunk (Inclination)
1	0°-45˚	< 45˚
2	45˚-90˚	≥ 45˚
3	> 90˚	-

Evaluation of posture for the present study was limited to the upper arms and trunk based on the function and design of the inclinometer. Rationale for posture selection included biomechanical and pathophysiological factors as well as previous research studies (Armstrong et al., 1982; Stetson et al., 1991; McAtamney and Corlett, 1993; Juul-Kristensen et al., 2001; Spielholz et al., 2001; Lowe, 2004a; Lowe, 2004b). There is some disagreement in the literature regarding postural thresholds for shoulder postures that define risk for musculoskeletal outcomes. The NIOSH review (1997) of 13 studies that have examined awkward postures and their relationship to shoulder musculoskeletal disorders concluded that there is evidence for a relationship between repeated or sustained shoulder postures with greater than 60° of flexion or abduction and shoulder musculoskeletal disorders. The NIOSH review (1997) considers shoulder flexion/adduction greater than 60° as awkward due to the greatest mechanical pressure on the supraspinatus tendon at arm elevations between 60°-120°. Much of the laboratory and clinical focus has been on upper arm elevation greater than 60°. A study that evaluated the intramuscular pressure of the infra- and supraspinatus muscles in relation to arm posture found that intramuscular pressure increases in association with upper arm elevation for both shoulder muscles (Palmerud et al., 2000). The elevation plane (flexion versus abduction) also influenced the intramuscular pressure; however, the elevation angle affected it more. In addition, hand load greatly influenced the intramuscular pressure. A new research study has shown an association and increased odds between upper arm flexion greater than or equal to 45° and rotator cuff syndrome (Silverstein et al., 2008). Shoulder postures were assessed by analysis of video footage using MVTA. A total of 733 subjects performing manufacturing tasks in 12 industries were evaluated. Based on the most recent findings (Silverstein et al., 2008), a cut point of 45° was used in the present study when evaluating shoulder postures using both MVTA and inclinometry (Table 5.1).

In regards to the trunk (low-back), the NIOSH review (1997) provided evidence that awkward postures are associated with low-back disorders. The review of 12 studies that examined low-back postures evaluated a variety of different back postures simultaneously and did not differentiate between specific back postures as associated with low-back disorders. Determining posture cut points for the trunk is not as straightforward for those of the shoulder since so many other risk factors affect the outcome of low back disorders. Bloswick and Villnave (2000) proposed the use of three cut points based on the range of motion of the back as one bends forward. This model uses 90° as the extreme range of motion for forward bending (flexion). This 0°-90° range is quartered and results in 4 posture categories: 0°-23°, 23°-45°, 45°-67°, and 67°-90°. The current study proposes that these categories be simplified further into two major categories: 0°-45° and 45°-90° (Table 6.1).

FIGURE 6.1

Research Design Schematic



Site and Subject Selection

The participant group was comprised of workers recruited at a single northern Colorado brewery. Only male participants were recruited due to low numbers of female employees for the tasks evaluated. Subjects with significant upper extremity pathology were no included in the study. This was determined through subject inquiry during the recruitment process.

Task Description

Subjects from three work areas (bottling, kegging, and canning) were recruited to participate in the study. Each work area consisted of several primary tasks, with each task typically performed by one worker. The bottling line was split into the wet and dry sections. The dry section tasks were the focus of data collection for this study. The two primary tasks on the bottling line involved loading cardboard trays, carriers, and cases into the machine. Cardboard carriers are filled with bottles. The filled carriers were enclosed by cases and the filled cases were loaded onto trays. The extent of the work was automated by the line; however, the workers had to retrieve the cardboard materials on pallets and load them into the machine throughout their shifts. It was noted in observations made prior to the study that the bottling tasks were much more dynamic than tasks performed in the kegging and canning areas. Loading the materials was not repetitive (1-2 stacks/minute); however, much of the awkward postures were noted during troubleshooting and non-standard tasks. Workers did not rotate within a shift.

The kegging line was comprised of two major tasks: loading kegs and offloading kegs. Workers interacted with two keg sizes, full kegs and 1/6 kegs. In regards to full-size kegs, one worker loaded the empty kegs (approximately 30 lbs) onto the line from pallets. The other worker used a lift-assist device to offload the filled kegs (approximately160 lbs) from the line. In regards to 1/6 kegs, both workers loaded the empty kegs (approximately 14 lbs) and offloaded the filled kegs (approximately 56 lbs) without mechanical assistance. Observations made prior to the study indicated repetitive upper arm movements based on the keg rate (3-4 full kegs/minute; 4-5 1/6 kegs/minute) and the position of the lift assist device. In addition, loading kegs involved lifting a keg from floor level to near waist/chest height at times. Kegging line workers used a fork lift periodically to bring empty kegs to the line or take away filled kegs.

The canning line was comprised of three major tasks: loading cans, gluing cases, and loading cases into trays. Canning line tasks were performed by contract labor, who tended to have less tenure than the kegging and bottling line workers. As the full cans exited the machine, the workers loading cans procured 12 cans and loaded them into a cardboard case. This task had two stations and was typically performed by two workers. The primary can loader was located closest to the filling machine and the secondary can loader was located immediately after the primary can loader. The workers erected the initially flat cases and then used a jig to secure the case before loading the cans. Observations made prior to the data collection indicated possible awkward upper arm and trunk posture for loading cans. After the cans were loaded into a case, the case was sent to the next position, where a worker operated a gluing mechanism to seal the case. Once the glue set, the worker moved the case to the next position, where a different worker put the cases into cardboard trays. This worker erected the cardboard tray and lifted the cases into the trays at approximate waist to chest height. Each tray held two cases. Typically, four workers operated the canning line tasks; however, if the crew was short, the third person typically operated the gluing mechanism and loaded the cases into trays.

Data Collection Instrumentation

The present study employed full-shift inclinometry to assess exposures to postures of the shoulders and trunk of 17 manufacturing workers. Inclinometers are instruments used for measuring angles of slope (tilt), inclination, or elevation of an object relative to gravity. The inclinometer (Figure 6.2) used in the present study was a pager-sized datalogging device called the Virtual Corset (MicroStrain, Inc.; Williston, VT). The Virtual Corset combines an inclinometer and datalogger (1 MB) into a pager-sized enclosure weighing less than two ounces that can be programmed to record the actual angle of inclination in a continuous stream of data in two dimensions. With programmable sampling rates, the Virtual Corset can collect continuous data over an entire work shift or over multiple days up to 80 hours based on memory capabilities. Full and multiple shift measuring capability has not been practical previously due to memory and battery limitations.

The inclinometer logs inclination in one degree increments over +/-180° in the flexion/extension axis and +/-70° in the lateral axis. When set in the 'Linear' mode, data are recorded in the actual angle of inclination in a continuous stream from beginning to the end of a session. The inclinometer can be programmed to record the X and Y axis separately or together (Figure 6.3). The X axis represented the angle developed by movement in the sagittal or lateral plane of the body during anterior or posterior movement. The Y axis represented the angle developed by movement in the coronal or frontal plane of the body during lateral movement. The orientation of the inclinometer, once mounted, will affect the orientation and recorded output for the X and Y axis. The inclinometer allows researchers to account for variation in a subject's anatomy by 'zeroing' the inclinometer relative to a standardized position. Differences in subject's anatomy can be normalized for the data collection process by employing this 'zeroing' capability. Accuracy of the inclinometer as reported by Mircostrain Inc. (Williston, VT) is estimated at +/- 0.5 degrees. Previous research utilizing accelerometer-based inclinometers have reported reliability/precision of approximately 1° for static positions (Hansson et al., 2001).

FIGURE 6.2

MicroStrain Virtual Corset Inclinometer

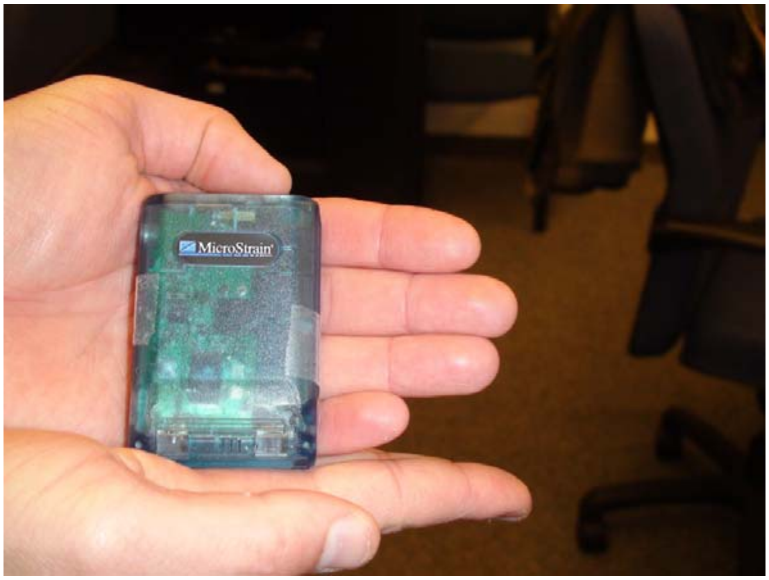
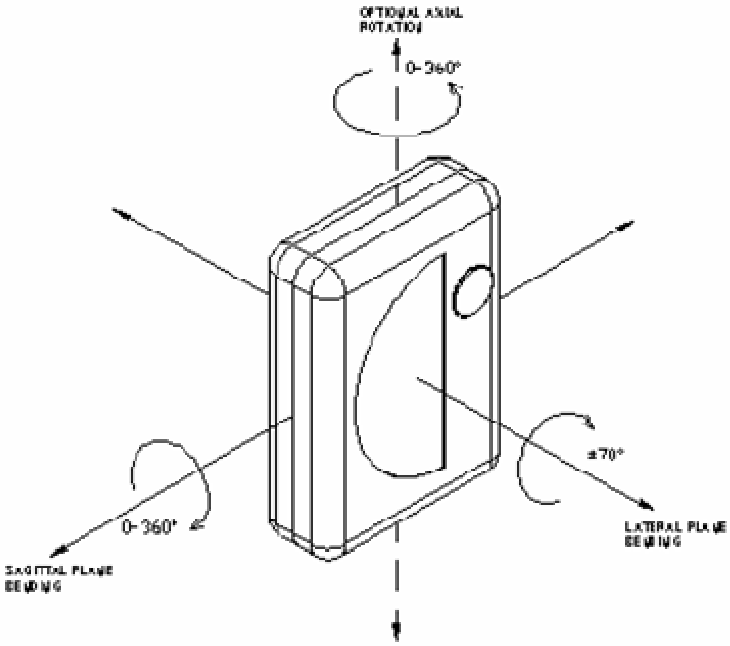


FIGURE 6.3

Diagram of the Virtual Corset Inclinometer



Data Collection

To assess exposures to postures of the shoulder and trunk, inclinometers were directly attached to the upper arms and trunk of each subject and were programmed to sample at 7.6 Hz. One inclinometer was attached to each of the upper arms by mounting the device to the posterior aspect of the upper arm, mid-way between the shoulder and elbow. The reference position for the upper arm (0° elevation) was defined with the subject standing, the arm hanging relaxed to the side. A combination of athletic wraps and tapes (Figure 6.4) were used to secure the inclinometers to the upper arms. First, the upper arm was wrapped with pre-wrap to create a barrier between the skin and plastic housing of the inclinometer. Once the attachment position was determined, the inclinometer was secured with a strip of athletic tape, then with several layers of Coban (Figure 6.5).

Another inclinometer was secured tightly to the trunk at approximately the location of the T6 spinous process (Trask et al., 2006a; Trask et al., 2006b). To approximate the T6 spinous process, the researchers located the C7 spinous process for each subject and manually counted down the spine until reaching the T6 spinous process. Once located, the Virtual Corset was centered over the spinous process and attached using a Tegaderm (3M: St. Paul, MN)(Figure 6.5). The Tegaderm created a seal around the Virtual Corset while maintaining flexibility during movements.

FIGURE 6.4

Wraps and Tape Used to Attach the Virtual Corset Devices

FIGURE 6.5

Virtual Corset Inclinometer Mounted to the Upper Arm

FIGURE 6.6

Virtual Corset Inclinometer Mounted to the Trunk

Before the inclinometers were attached, they were launched and calibrated using Virtual Corset 3.2.3 software (Microstrain, Inc; Williston, VT) installed on three laptop computers to correspond with the three devices. To begin use of the inclinometers, two AA batteries were inserted. Once the batteries were inserted, the inclinometers began collecting data. Due to some battery connection issues, the lids of the inclinometers were secured with electrical tape to ensure the battery connection was maintained throughout the sampling shifts. The inclinometers that were attached to the trunk were wrapped in a layer of plastic wrap followed by a layer of pre-wrap. This protected the inclinometers from sweat build-up due to the Tegaderm. Once the inclinometer preparation was complete, each device was connected to a laptop computer. Based on the inclinometer design and software, only one device could be connected to a single computer at a time.

While it was not ideal to setup three computers in the field, this allowed the inclinometers to be launched synchronously. It was important that all inclinometers were capturing the same exposure period. Once a connection was established for each inclinometer, the Virtual Corset software was used to perform a position check. The inclinometers were laid flat on a table to verify all inclinometers were reading approximately zero degrees. The angle of inclination was viewed directly on the computer screens using the Virtual Corset 3.2.3 software. Once the position checks were completed, the inclinometers were launched synchronously using the Virtual Corset 3.2.3 software. Any previous data was erased when the inclinometers were launched. A dedicated watch (hh:mm:ss) synchronized with the internal clocks of the three laptop computers was used to verify the launch time and used when recording observations in the field.

When mounting the inclinometers on the upper arms, subjects were instructed to stand upright with their arms hanging to the side. Inclinometers were mounted to one arm at a time, beginning with the left arm. While one researcher mounted the inclinometer, the other researcher secured the arm at the side of the subjects to ensure a standard position across subjects for zeroing purposes. Once the attachment location was determined, the inclinometers were zeroed to the subject. When attaching the inclinometers to the trunk, subjects were instructed to stand upright while staring straight ahead. Again, one researcher determined the location and secured the inclinometer while the other researcher zeroed the device to the individual using the Virtual Corset 3.2.3 software. While mounting the inclinometers on the subjects, a second check was performed. Subjects were instructed to move their arms and trunk to designated postures. For the arms, the subjects were required to raise their arms from approximately 0° to approximately 90° simultaneously to ensure the inclinometers were functioning properly For the trunk, subjects were instructed to bend forward at the waist to ensure the inclinometers were functioning properly. The angles of inclination were viewed directly on the computer screens using the Virtual Corset 3.2.3 software. The inclinometers remained attached to the subjects for the duration of their work shifts (approximately eight continuous hours), with a minimum run time of four hours.

Sampling Duration and Sample Size

Three major work areas were evaluated at the manufacturing facility: bottling, canning, and kegging. Several work tasks as performed in these areas were used in the present study, resulting in a total of seven tasks evaluated. Based on the seven tasks and the available crewing (three crews), it was determined that 14 manufacturing workers would be available for full-shift exposure assessment over two separate shifts, resulting in 28 full-shift samples. Recruitment was performed by the researchers with assistance from the facility health and safety personnel. Due to some changes in personnel throughout the study, 17 workers were available for recruitment; however, only 15 of the workers were available for inclinometry over two shifts. Data collection occurred over entire 8-hour work shifts, with minimum sample times of four hours for all manufacturing workers over both occasions.

Sample size and duration was based on previous research using inclinometry (Trask et al., 2007; Hansson et al., 2006; Trask et al., 2006; Moller et al., 2004; Mathiassen et al., 2003; Bernmark and Wiktorin, 2002; Juul-Kristensen et al., 2001; Hansson et al., 2001). Previous research using inclinometry has included investigations of inter-method comparisons, evaluation of exposure variability, and evaluation of cost and feasibility (Trask et al., 2007; Hansson et al., 2006; Trask et al., 2006; Moller et al., 2004; Mathiassen et al., 2003; Bernmark and Wiktorin, 2002; Juul-Kristensen et al., 2001; Hansson et al., 2001). These studies have varied in the number of subjects (5-125 subjects), time analyzed (3 minutes to approximate 8-hr full-shifts), the number of tasks/jobs/industries evaluated (2-3 laboratory-based tasks to several tasks to 50 different worksites), and have included both laboratory and field data collection.

With regard to statistical power, Mathiassen et al. (2003) demonstrated in a relatively constrained industrial work, the number of subjects needed to detect differences depended on the sizes of the within and between subject components of variability. For upper arm inclinometry, Mathiassen et al. (2003) reported moderate to high levels of within and between subject variability for postures of the upper arms. The authors (Mathiassen et al., 2003) concluded that the number of subjects required to obtain an acceptable precision in group mean exposure varied depending on the exposure parameter. The authors (Mathiassen et al., 2003) used the median elevation of the right and left upper arms to perform their statistical analyses. The median right and left upper arm elevation was reported approximately 45°. Estimates of the number of subjects necessary to achieve a 95% confidence interval, detection of a 10% difference in mean exposure levels, and a power of 0.80 were 15 for the right arm and 8 for the left arm (Mathiassen et al., 2003). These estimates were based on the mean exposure and exposure variability found for the constrained industrial tasks evaluated in the study and were based on 120 distinct securings of thread fasteners over six combinations of tool and work locations (Mathiassen et al., 2003). Based on the findings by Mathiassen et al.

(2003), the researchers of the present study determined that 15 subjects should allow for the detection of at least a 10% difference in mean exposure levels with 80% power.

Data Analysis

Data obtained from the inclinometers was downloaded using Virtual Corset 3.2.3 software (Microstrain, Inc; Williston, VT). The Virtual Corset software creates a Microsoft Excel comma-separated file with degrees of angle inclination for the X and Y axis. The Virtual Corset data files were processed using LabView 8.6 software (National Instruments: Austin, TX). The LabView software was programmed to create an EVA for each of the 15 workers across both full-shift samples for the specific postures categories (Table 6.1) of the upper arms and trunk. Exposure variation analysis is a data reduction method that has been used for data obtained by electromyography (Mathiassen and Winkel, 1991). Previous research utilizing EVA to reduce inclinometry data was not found in the literature. When using EVA to reduce inclinometry data, the analysis describes the percent time spent in a pre-defined posture caregory as well as the length of time (duration) at each of those pre-defined categories (Table 6.2). The duration categories were defined as the percentage of time spent within a specific posture category (0-1 seconds, 1-3 seconds, 3-5 seconds, and > 5 seconds).

Upon completion of the full-shift EVA's, shorter intervals (4-hr, 2-hr, 1-hr, and 30-min) of EVA were obtained from the full-shift data. Comparison of the shorter sampling durations was performed to investigate the optimal combination of measurement accuracy as compared to the full-shift data and to investigate optimal efficiency to reduce sampling time and resources (Trask et al., 2009). The full-shift data were re-sampled *a posterior* for shorter intervals: 4-hr, 2-hr, 1-hr, and 30-min for each of the 15 subjects and both shifts for every posture category of the upper arms and trunk. For each of the shorter intervals (4-hr, 2-hr, 1-hr, and 30-min), a randomly selected start time obtained from a random time generator was used to re-sample the full-shift data. The re-sampling resulted in five sets of data for each of the 15 workers and two shifts assessed.

Statistical Analyses

Various statistical measures were used to evaluate the data. All statistical analyses were completed using SPSS 18.0 statistical package. Summary statistics (mean and standard deviation) were computed for the full-shift inclinometry data for the percent time spent in each posture category and the duration at each posture category across the fifteen subjects. Exposure variation analyses were used to evaluate the percent time logged in the specific posture categories and the length of time (duration) in each posture category of the upper arms and trunk.

Comparison of the re-sampled shorter intervals (4-hr, 2-hr, 1-hr, and 30-min) to the full-shift measures was completed using several statistical measures. Pearson correlation coefficients were computed to determine the strength of the relationship between the full-shift and shorter sampling intervals. Repeated measures ANOVA was used to test for significant differences between the full-shift measures and the shorter sampling intervals. Repeated measures ANOVA were completed for each body part and posture category with the full-shift measures and the re-sampled shorter durations as a within-subjects factor with five level. Work area was nested as a between subjects factor. A significance level of 0.05 was used for all comparisons. The absolute difference and percent difference (absolute difference divided by the full-shift value) were used to examine the level of deviation from the full-shift exposure measures and each shorter sampling interval (Trask et al., 2008). Agreement between the full-shift exposure measures and each shorter sampling interval were also evaluated using bias and limits of agreement (Bland and Altman, 1986; Trask et al., 2008). Bias was calculated as the mean of the differences between data from the full-shift exposure measures divided by the standard deviation of those differences. The upper and lower limits of agreement were calculated by adding or subtracting the mean of the differences between data from the full-shift exposure measures to the standard deviation of those differences multiplied by two (Bland and Altman, 1986).

The Hartley's test (O'Brien, 1981), was computed to further investigate the differences between the full-shift measures and shorter sampling intervals. The F_{max} or Hartley's test (O'Brien, 1981) was used to verify whether the different sampling durations had a similar variance. This test involves computing the ratio of the largest group variance to the smallest group variance. The resulting ratio is then compared to a critical F value as obtained from the F sampling distribution for the specified degrees of freedom and level of significance. Hartley's test assumes that the data are normally distributed and that each group has an equal number of members, similar to the assumptions of the ANOVA.

RESULTS

Of the 15 subjects enrolled in the present study, 100% were male, 80% were right-hand dominant, mean age in years was 35.4 (24-59), mean height in inches was 70.6 (66-73), and mean months worked in the particular job areas assessed was 18.7 (3-42). Of the 15 subjects analyzed, six were recorded in the kegging area, five were recorded in the canning area, and four were recorded in the bottling area. Second shift measurements were not obtained for two of the 17 subjects due to attrition. These two subjects were contractor laborers who were moved to a different company during the study period. Therefore, the data obtained for the 15 subjects with two full-shift measures was used in the subsequent analyses. The mean time analyzed for the inclinometer data across the 15 subjects with two shifts was 6.54 (4.47-7.90) hours. The mean time, averaged across the 15 subjects and both shifts, to attach and perform the checks for the inclinometer devices was 17.2 (14-23) minutes.

Means of the raw data for each posture category of the upper arms and trunk as obtained by the inclinometry are provided in Tables 6.2-6.4. The data provided in Tables 6.2-6.4 represent the mean percent time logged in each posture category as obtained by inclinometer and subsequently analyzed using exposure variation analysis across the 15 subjects for the left upper arm, right upper arm, and trunk. Means and standard deviations are presented for: overall data which included both shifts and all three work areas, data specific to the first shift, data specific to the second shift, and data specific to the work areas. Figures 6.7-6.9 provide column charts that illustrate the mean percent time spent in posture categories of the left and right upper arms and trunk as measured by the inclinometry. Figures 6.7-6.9 provide the mean percent time for each of the three work areas along with corresponding standard deviation error bars.

The average percent time spent in the specific posture categories of the left and right upper arms was greatest for 0°-44° elevation and lowest for > 90° elevation (Tables 6.2-6.3). The right upper arm had, on average, 4.6% more time in 45º-90º elevation and 2.31% more time in > 90º elevation than the left upper arm. The percent time spent in the left upper arm elevation categories demonstrated small changes when evaluating the mean by shift and work area. The largest standard deviation for the left upper arm posture categories was observed for 0º-44º elevation for the second shift. Standard deviations for the right upper arm were higher than the left upper arm, indicating that the right upper arm had a higher variability or dispersion. Right upper arm 45º-90º and > 90º had the highest percent time in the kegging work area (Figures 6.7-6.9) Trunk inclination < 44º had much higher percent time than trunk inclination $\geq$ 44º, with the bottling work area having the highest percent time in trunk inclination $\geq$ 44º (Table 6.4, Figure 6.9). Standard deviations were highest for right upper arm elevation postures and were highest for kegging work area tasks (Tables 6.2-6.4).

TABLE 6.2

Mean Percent Time Logged by Inclinometry for Left Upper Arm Elevation Postures

	% Time x (SD)		
Summary Measure Group	**0°-44° Elevation**	**45°-90° Elevation**	**> 90° Elevation**
Overall	89.51 (4.95)	9.54 (4.39)	0.95 (0.92)
Shift 1	90.22 (3.80)	8.78 (3.22)	1.00 (0.90)
Shift 2	88.86 (5.88)	10.24 (5.28)	0.90 (0.96)
Bottling	87.87 (5.59)	11.09 (5.49)	1.04 (0.43)
Kegging	88.18 (3.62)	10.45 (2.57)	1.37 (1.21)
Canning	92.26 (5.14)	7.35 (4.86)	0.38 (0.33)

TABLE 6.3

Mean Percent Time Logged by Inclinometry for Right Upper Arm Elevation Postures

	% Time x (SD)		
Summary Measure Group	**0°-44° Elevation**	**45°-90° Elevation**	**> 90° Elevation**
Overall	82.84 (11.58)	13.90 (9.75)	3.26 (3.72)
Shift 1	83.66 (10.29)	12.20 (7.48)	4.14 (4.78)
Shift 2	82.03 (13.05)	15.60 (11.62)	2.37 (2.02)
Bottling	85.51 (6.55)	12.59 (6.00)	1.90 (0.98)
Kegging	76.85 (11.98)	17.25 (10.66)	5.90 (4.65)
Canning	87.90 (11.86)	10.93 (10.65)	1.17 (1.25)

TABLE 6.4

Mean Percent Time Logged by Inclinometry for Trunk Inclination Postures

	% Time x (SD)	
Summary Measure Group	**< 45° Inclination**	**≥ 45° Inclination**
Overall	94.77 (3.31)	5.23 (3.31)
Shift 1	94.97 (3.40)	5.03 (3.40)
Shift 2	94.57 (3.35)	5.43 (3.35)
Bottling	92.28 (3.73)	7.72 (3.73)
Kegging	95.20 (3.24)	4.80 (3.24)
Canning	96.13 (2.20)	3.87 (2.20)

FIGURE 6.7

Mean Percent Time and Standard Deviation Logged by Inclinometry

for Left Upper Arm Elevation Postures by Work Area

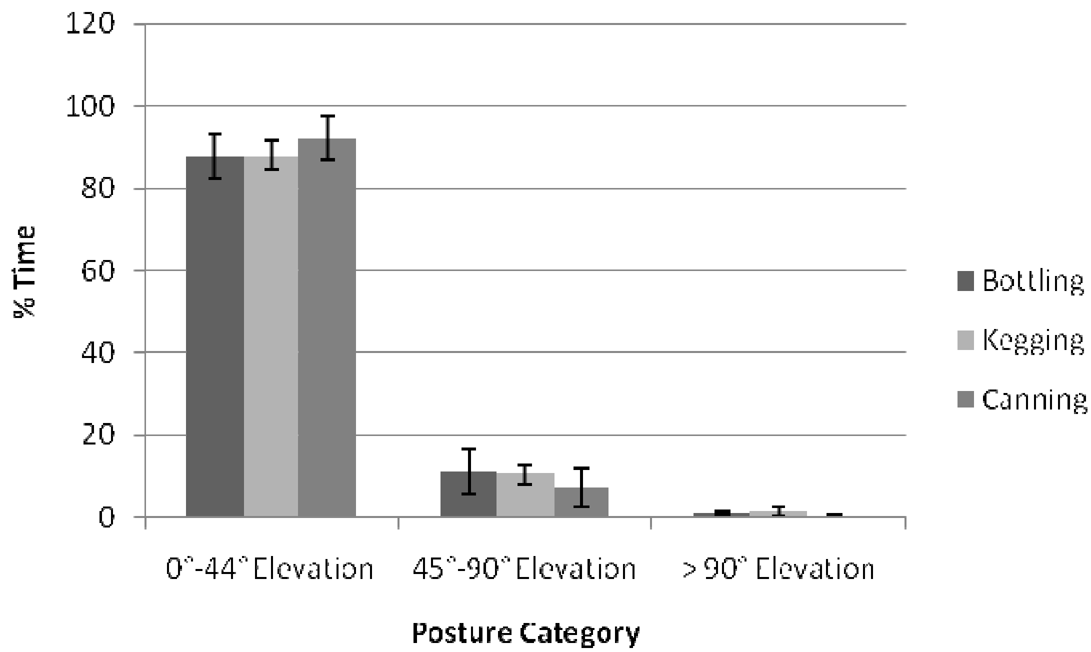


FIGURE 6.8

Mean Percent Time and Standard Deviation Logged by Inclinometry

for Right Upper Arm Elevation Postures by Work Area

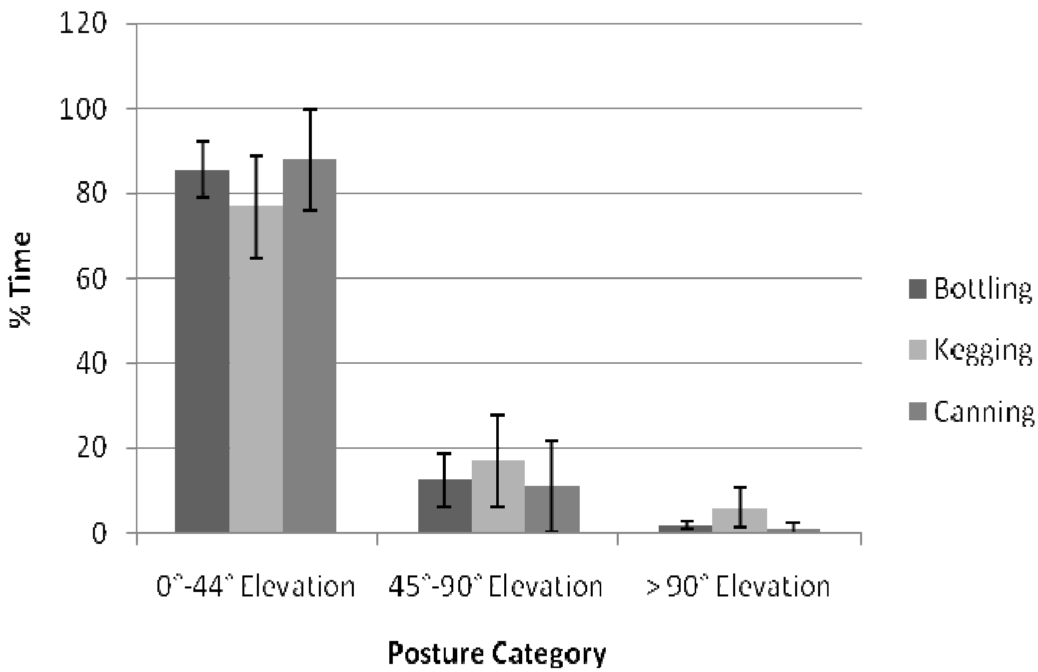
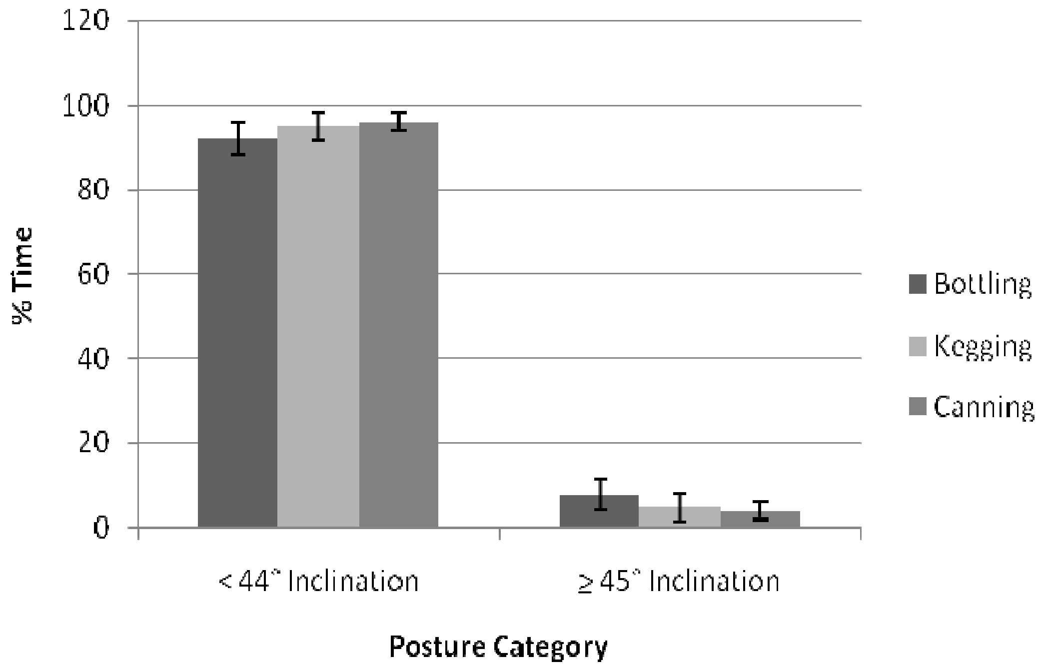


FIGURE 6.9

Mean Percent Time and Standard Deviation Logged by Inclinometry

for Trunk Inclination Postures by Work Area



Summary statistics are presented in Table 6.5 for postures of the left and right upper arms and trunk for the full-shift inclinometry and the sub-samples (4-hr, 2-hr, 1-hr, and 30-min). Summary statistics include all 15 subjects and both shifts. The 30-min sub-sample overestimated the full-shift data for five out of the eight the posture categories of the left and right upper arms and trunk. The 1-hr sub-sample overestimated the full-shift data for four out of the eight posture categories. Percentage difference for the means was highest in both trunk postures when comparing the full-shift data with the 30-min sub-samples. The smallest percentage difference for the means was in right upper arm elevation > 90. The 2-hr and 4-hr sub-samples were similar to the full-shift data.

TABLE 6.5

Summary Statistics for Full-Shift and Shorter Sampling Intervals

Body Part	**Posture**	**Sampling Interval**				
		Full	**4-Hr**	**2-Hr**	**1-Hr**	**30-Min**
Left Arm	**0°-44°**					
	Mean	89.43	89.98	89.27	90.15	88.66
	Std Dev	*4.17*	*4.12*	*5.39*	*5.26*	*6.64*
	45°-90°					
	Mean	9.45	9.10	9.92	9.12	10.32
	Std Dev	*3.75*	*3.54*	*4.93*	*5.10*	*5.95*
	> 90°					
	Mean	0.97	0.92	0.82	0.73	1.03
	Std Dev	*0.76*	*0.66*	*0.72*	*0.75*	*1.49*
Right Arm	**0°-44°**					
	Mean	82.84	82.99	83.88	83.94	83.58
	Std Dev	*10.82*	*11.48*	*11.29*	*12.07*	*12.52*
	45°-90°					
	Mean	13.90	13.48	12.86	12.68	11.97
	Std Dev	*8.99*	*9.42*	*9.61*	*10.18*	*9.52*
	> 90°					
	Mean	3.26	3.53	3.26	3.40	4.45
	Std Dev	*2.94*	*3.31*	*2.80*	*5.30*	*4.62*
Trunk	**< 45°**					
	Mean	94.43	94.30	94.42	94.38	92.19
	Std Dev	*3.53*	*3.33*	*5.48*	*2.85*	*7.94*
	≥ 45°					
	Mean	5.57	5.70	5.59	5.61	7.81
	Std Dev	*3.52*	*3.33*	*5.48*	*2.85*	*7.94*

Pearson product moment correlation coefficients were computed for the full-shift data correlated with each sub-sample (Table 6.6). These coefficients represented correlations across all 15 subjects and both shifts. In regards to specific sampling duration pairs, five correlation coefficients were not significant at the 0.05 level (Table 6.6). The non-significant correlations were: both trunk posture categories for correlation of the full-shift data with the 2-hr and 30-min intervals and left upper arm elevation > 90° for correlation of the full-shift data with the 1-hr interval. Two additional correlations were not significant at the 0.01 level and included: left upper arm elevation 0°-45° and 45°-90° for correlation of the full-shift data with the 1-hr interval. Significance meant there was evidence to reject the null hypothesis ($r = 0$) or that there was a statistically significant relationship between the full-shift sample and the sub-samples.

Scatter plots were constructed to reveal associations between the full-shift data and shorter sampling intervals for the postures of the left and right upper arms and trunk (Figures 6.10-6.17). Scatter plots were used to illustrate the correlation and direction of correlation of two different variables. Figures 6.10-6.12 contain scatter plots for all three posture categories of the left upper arm for each shorter sampling interval versus the full-shift. Figures 6.13-6.15 contain scatter plots for all three posture categories of the right upper arm for each shorter sampling interval versus the full-shift. Figures 6.16-6.17 contain scatter plots for the two posture categories of the trunk for each shorter sampling interval versus the full-shift.

TABLE 6.6

Pearson Product Moment Correlation Coefficient

for Full-Shift and Shorter Sampling Intervals

Body Part	Posture	Sampling Interval			
		Full vs 4-Hr	**Full vs 2-Hr**	**Full vs 1-Hr**	**Full vs 30-Min**
Left Arm	**0°-44°**	0.95*	0.94*	0.58	0.78*
	45°-90°	0.96*	0.97*	0.59	0.73*
	> 90°	0.85*	0.80*	0.50	0.78*
Right Arm	**0°-44°**	0.99*	0.96*	0.84*	0.83*
	45°-90°	0.98*	0.96*	0.90*	0.75*
	> 90°	0.99*	0.94*	0.75*	0.95*
Trunk	**< 45°**	0.88*	0.89*	0.43	0.39
	≥ 45°	0.88*	0.89*	0.43	0.39
Average	**All**	0.94	0.92	0.63	0.70

*Indicates significance > 0.05

FIGURE 6.10

Scatter Plots of Left Upper Arm Elevation 0°-44° for

Each Shorter Sampling Interval versus the Full-Shift Data

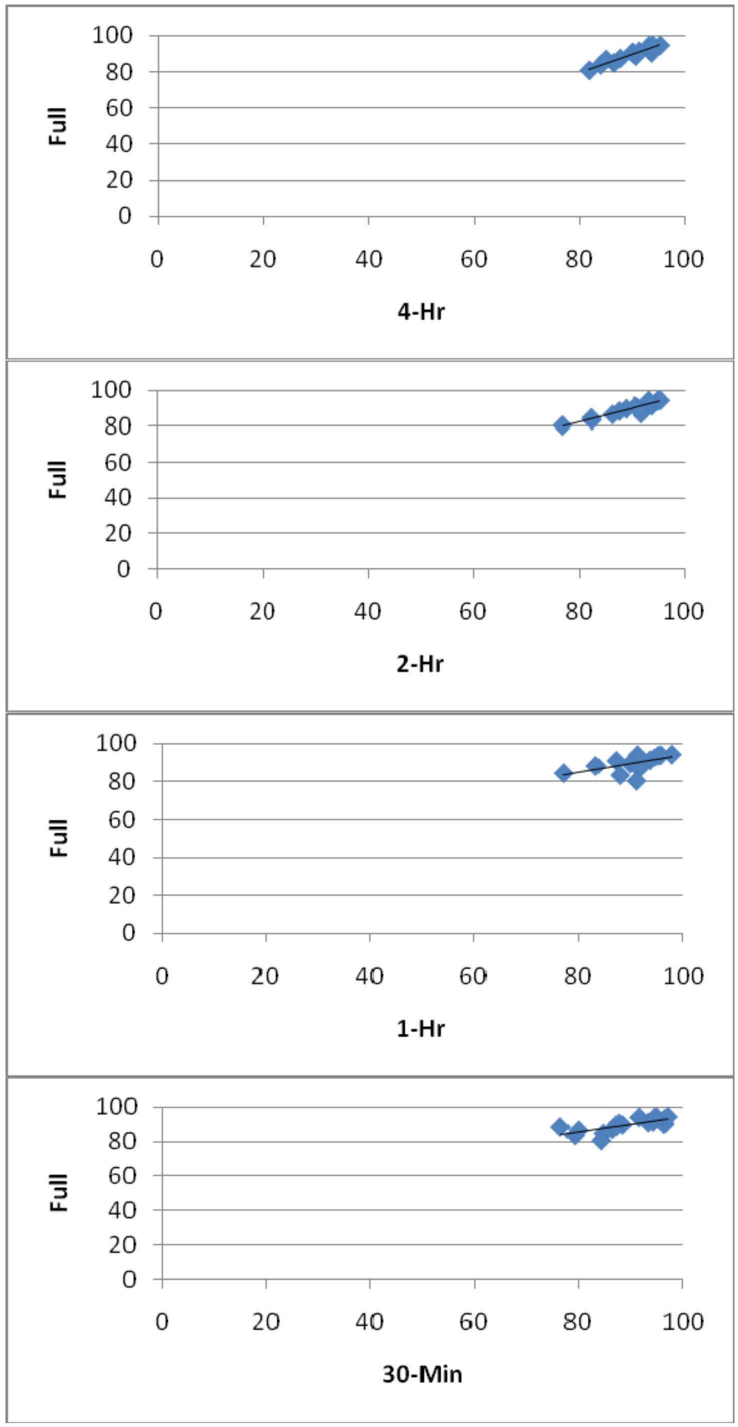


FIGURE 6.11

Scatter Plots of Left Upper Arm Elevation 45°-90° for

Each Shorter Sampling Interval versus the Full-Shift Data

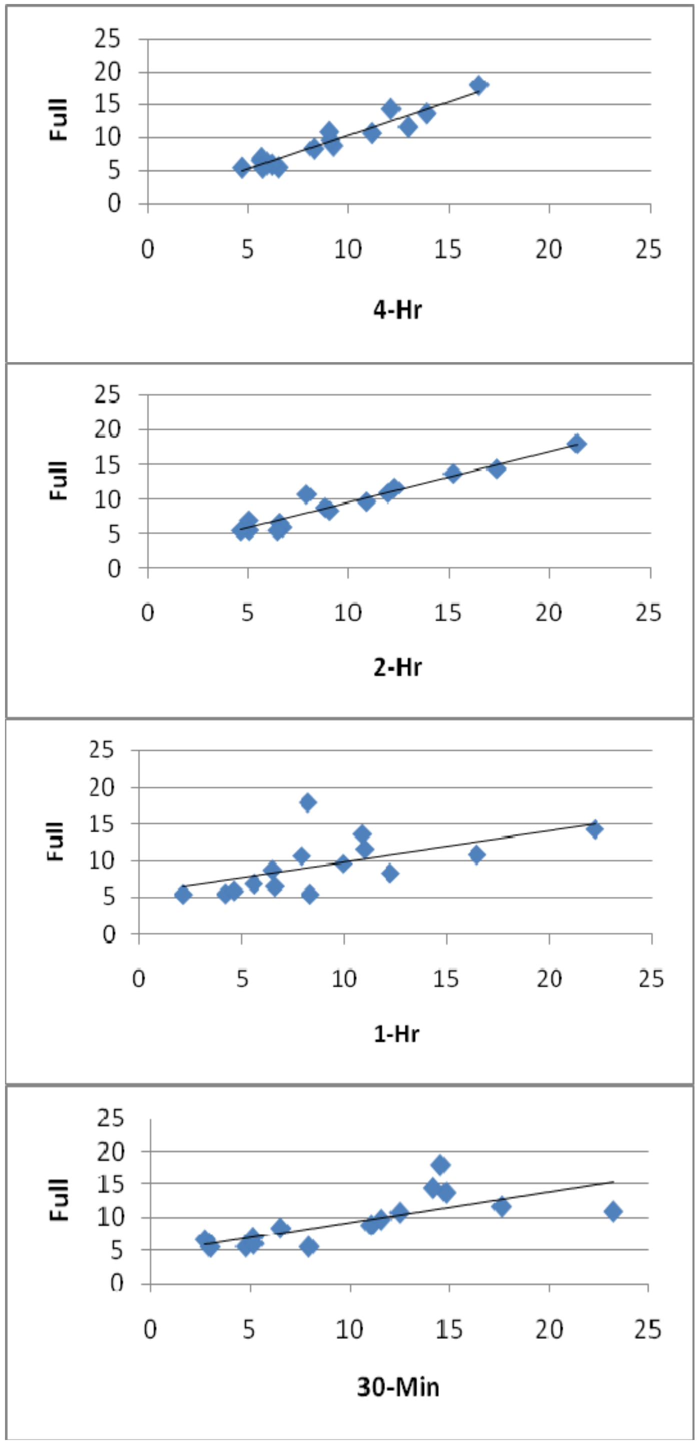


FIGURE 6.12

Scatter Plots of Left Upper Arm Elevation > 90° for

Each Shorter Sampling Interval versus the Full-Shift Data

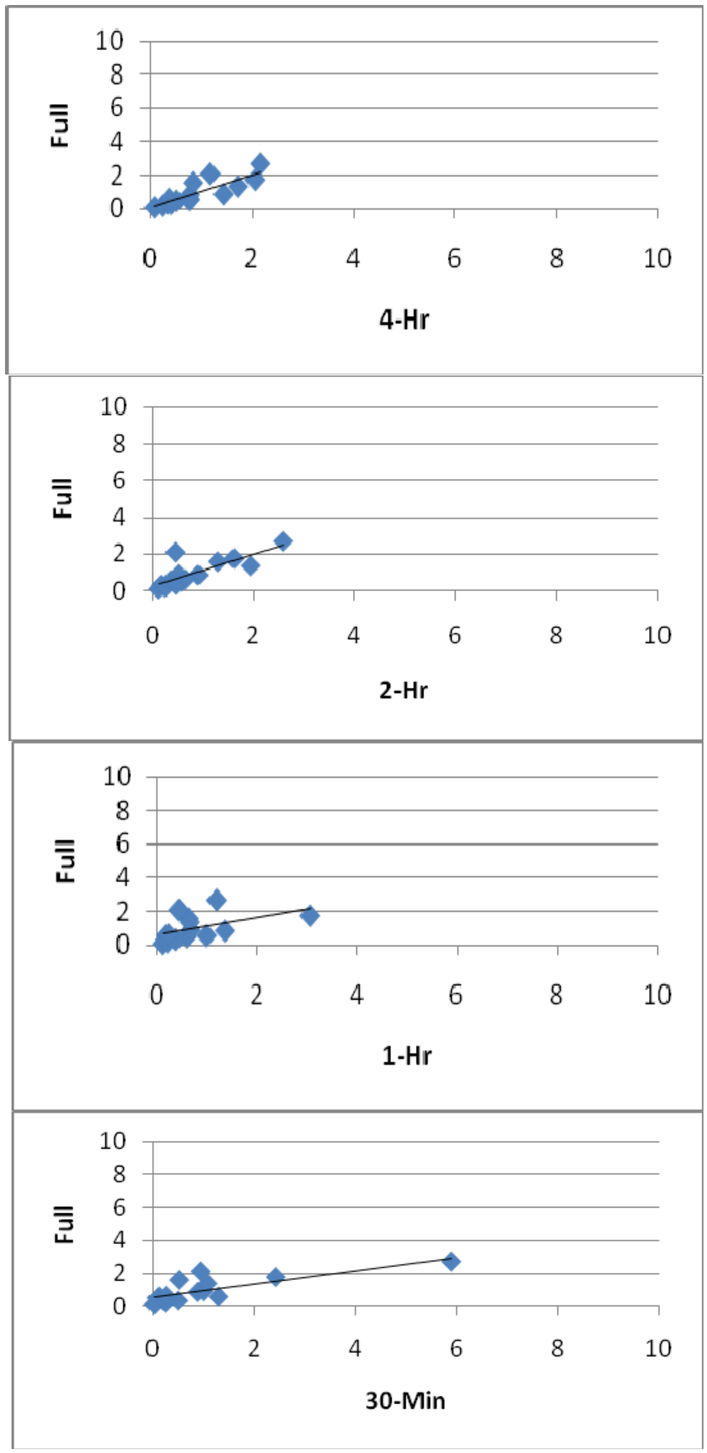


FIGURE 6.13

Scatter Plots of Right Upper Arm Elevation 0°-44° for

Each Shorter Sampling Interval versus the Full-Shift Data

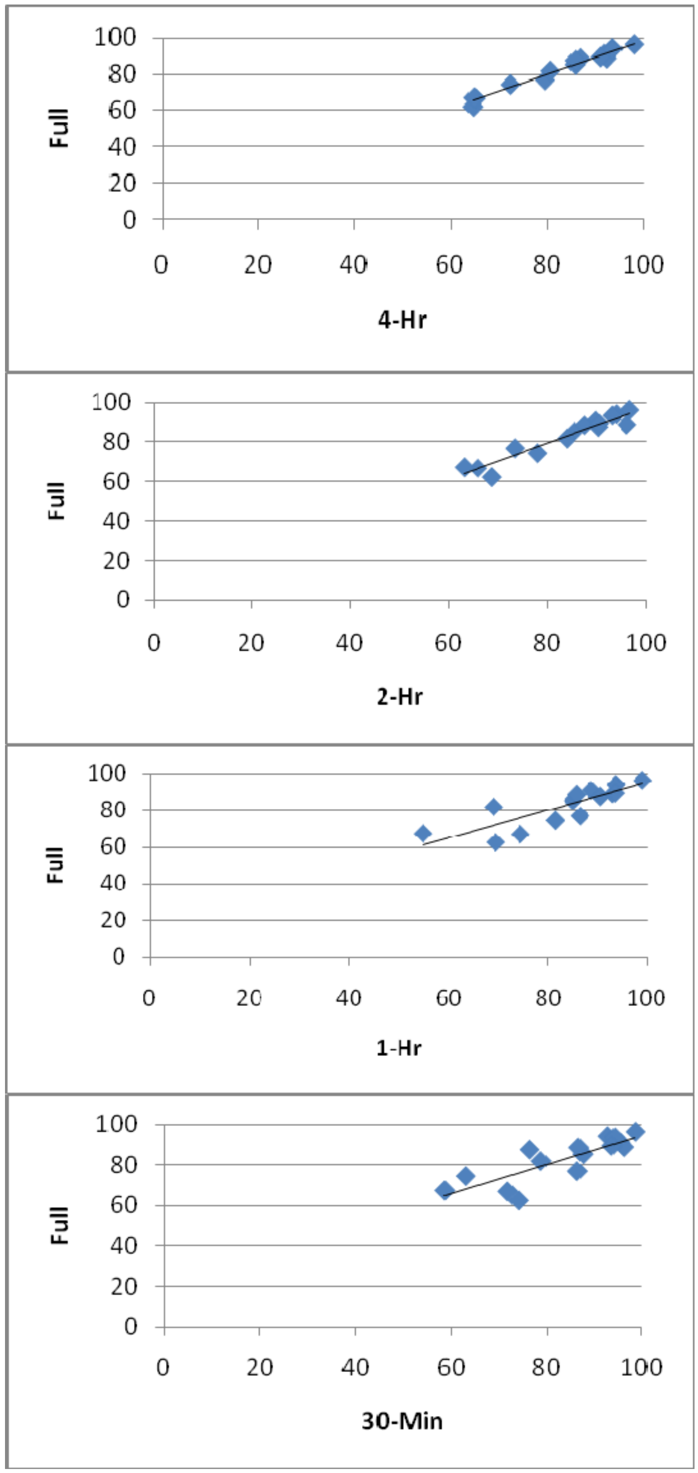


FIGURE 6.14

Scatter Plots of Right Upper Arm Elevation 45°-90° for

Each Shorter Sampling Interval versus the Full-Shift Data

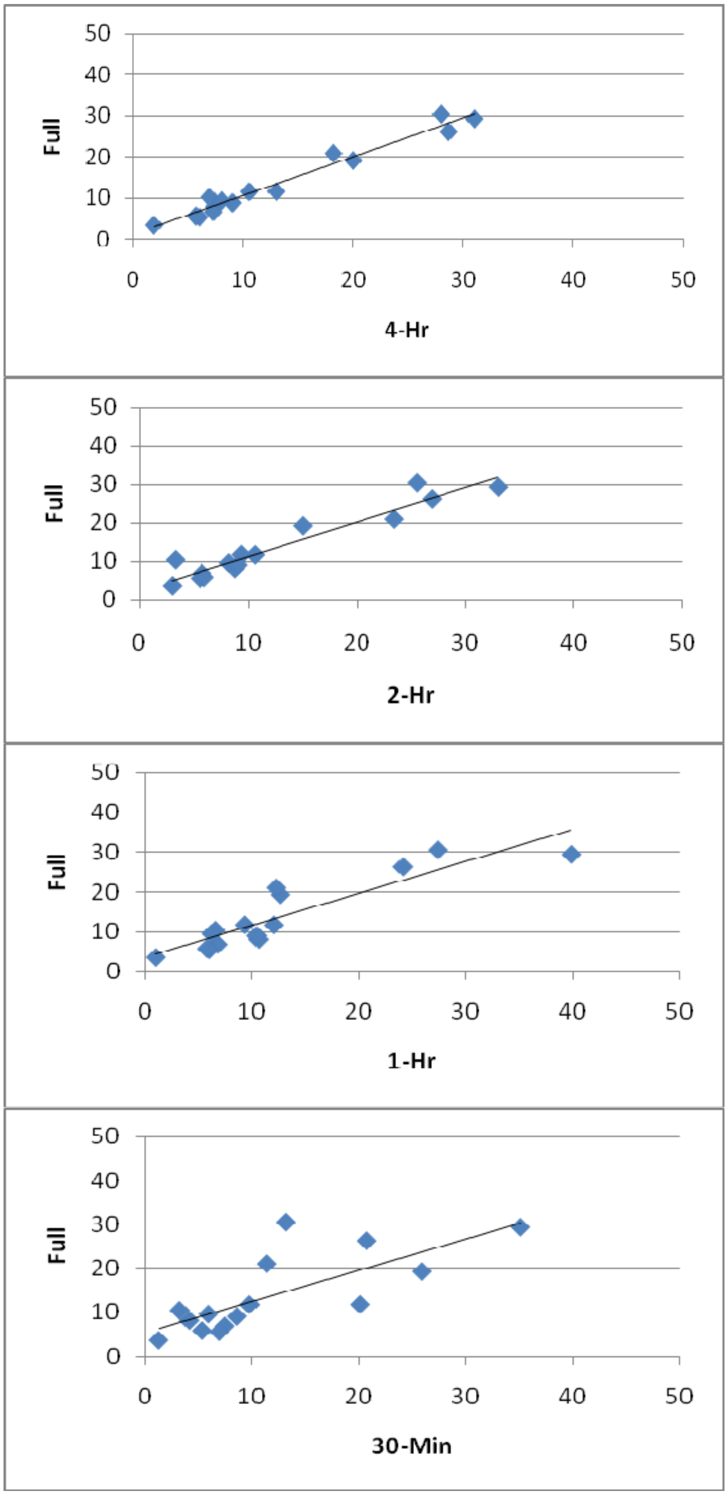


FIGURE 6.15

Scatter Plots of Right Upper Arm Elevation > 90° for

Each Shorter Sampling Interval versus the Full-Shift Data

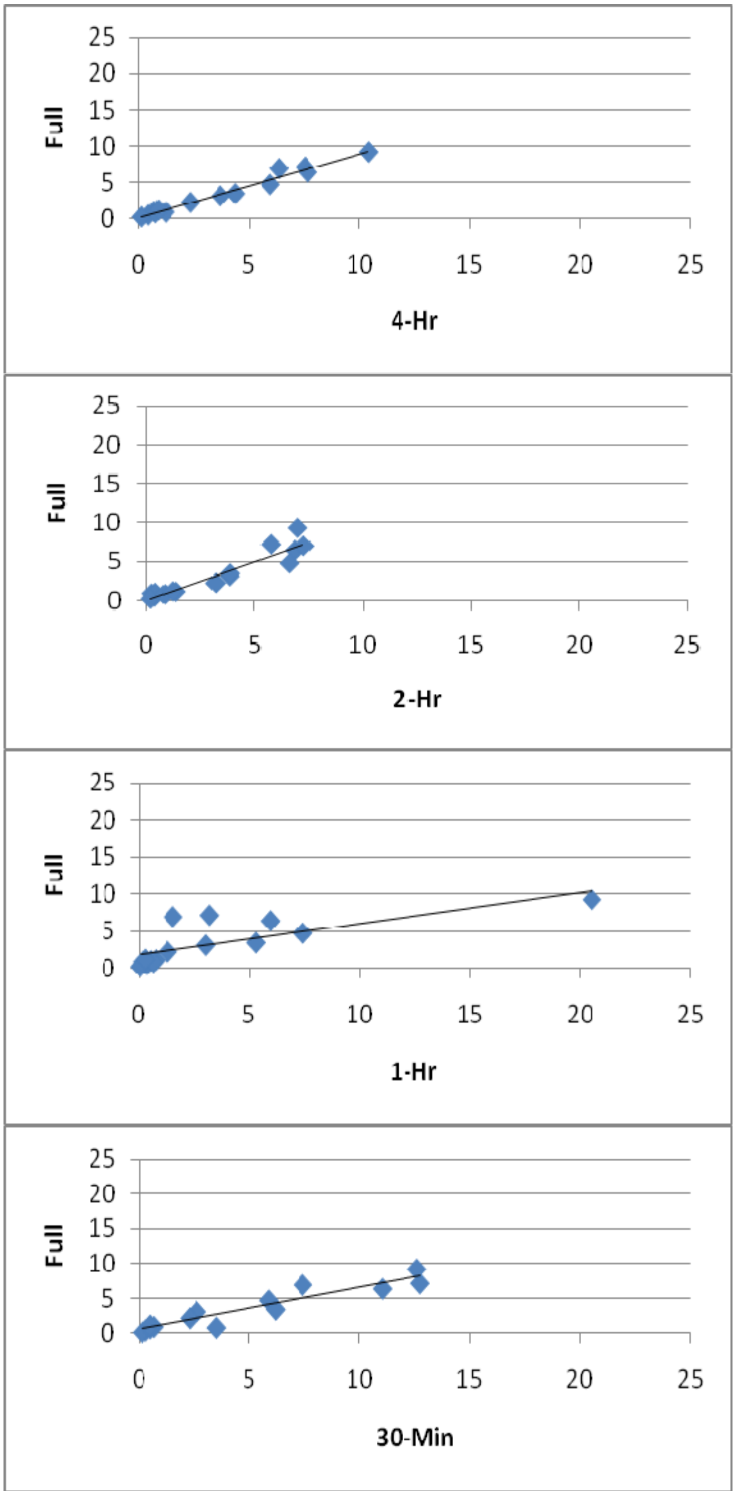


FIGURE 6.16

Scatter Plots of Trunk Inclination < 45° for

Each Shorter Sampling Interval versus the Full-Shift Data

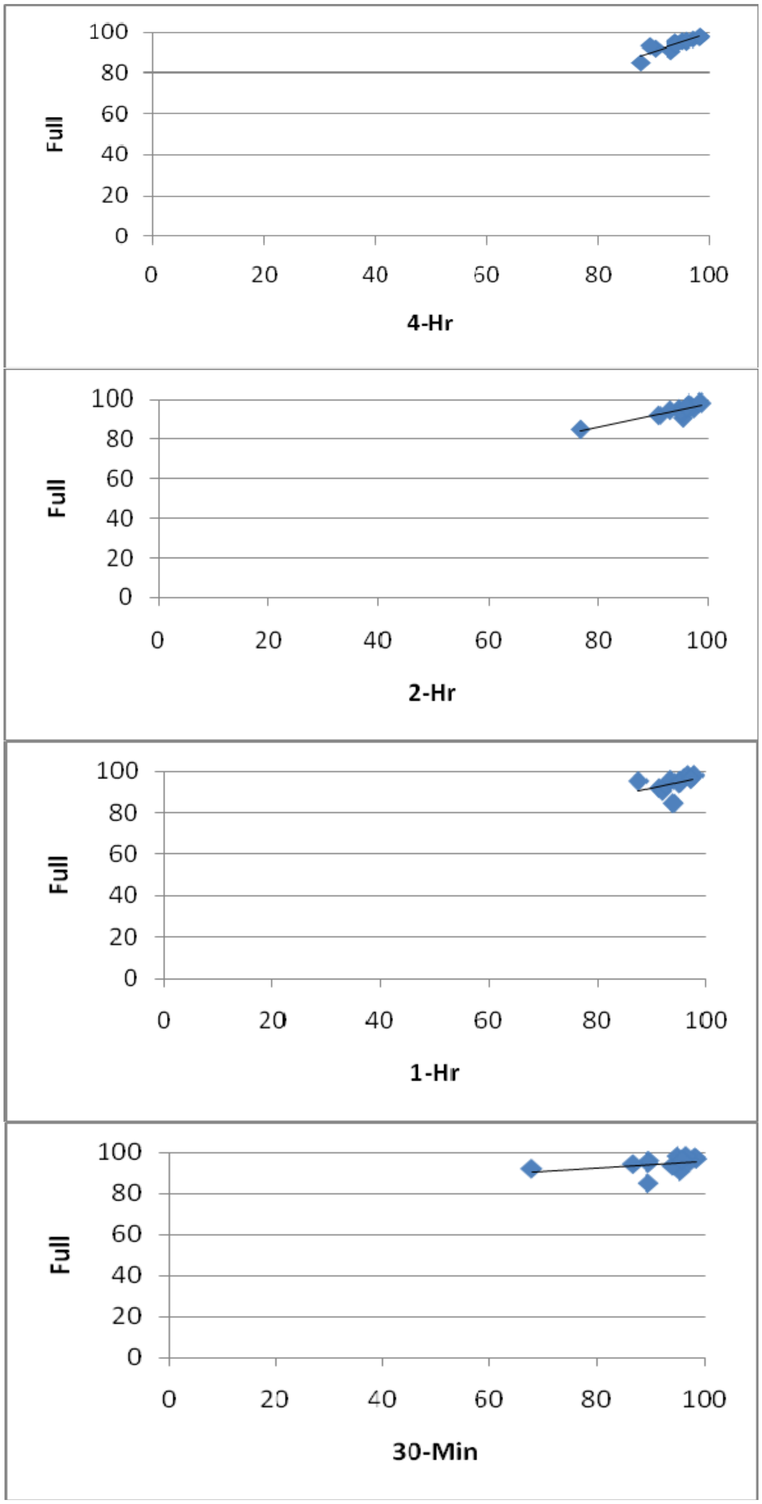
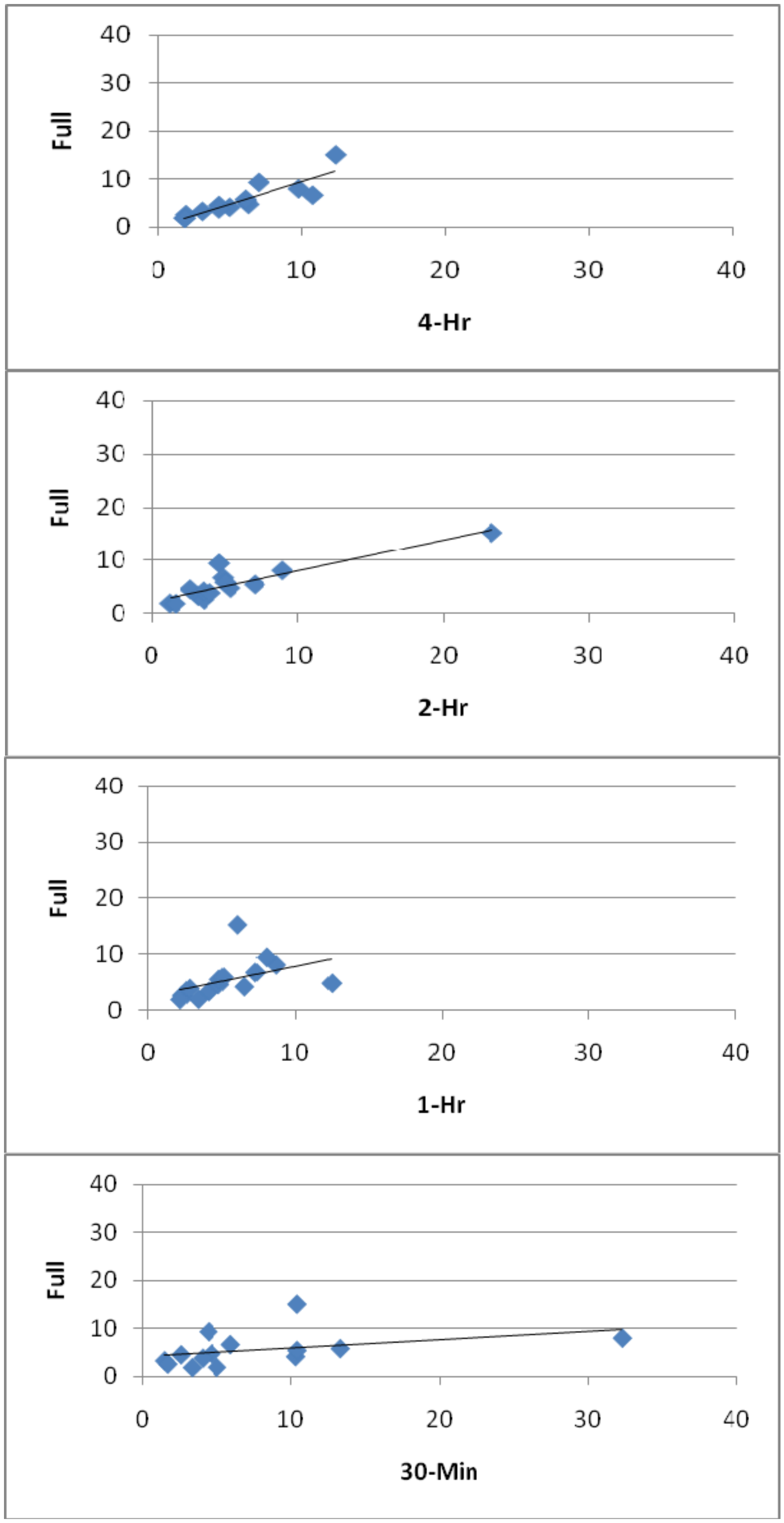


FIGURE 6.17

Scatter Plots of Trunk Inclination ≥ 45° for

Each Shorter Sampling Interval versus the Full-Shift Data



Scatter plots of the left upper arm postures revealed positive correlations for all comparisons of the shorter sampling durations versus the full-shift measures; however, the correlations became weaker as the sampling durations became smaller (Figures 6.10-6.17). Scatter plots of the right upper arm postures revealed positive correlations for all comparisons of the shorter sampling durations versus the full-shift measures; however, the correlations became weaker as the sampling durations became smaller (Figures 6.13-6.15). Scatter plots of the trunk inclination postures had positive correlations for all comparisons of the shorter sampling durations versus the full-shift measures; however, the correlations became weaker as the sampling durations became smaller (Figures 6.16-6.17).

Results of the repeated measures ANOVA are presented in Table 6.7. Duration was the within-subjects factor encompassing the full shift data as well as the shorter intervals (4-hr, 2-hr, 1-hr, 30-min) and work area was a between-subjects factor. The effect of work area was significant for right upper arm elevation > 90°, trunk inclination < 45° and trunk inclination ≥ 45°. The interaction effect of duration and work area was significant for trunk inclination < 45° and trunk inclination ≥ 45°. Significance in the work area factor indicates that the percent time of posture recorded for right upper arm elevation > 90°, trunk inclination < 45°, and trunk inclination ≥ 45°changed over the three work areas of bottling, kegging, and canning. Significance in the interaction effect of duration and work area factor indicates that the pattern of percent time of posture recorded within the different sampling intervals for trunk inclination < 45° and trunk inclination ≥ 45° changed across the three work areas. No significant effects were found for the duration factor, meaning that the percent recorded for all body parts and postures did not change across the different sampling intervals.

TABLE 6.7

Repeated Measures ANOVA Results for Different Sampling Duration Intervals and Work Areas

Body Part and Posture	**DF**	**Mean Square**	***p*-value**
Left Upper Arm Elevation 0°-44°			
Duration	4	5.259	0.643
Work Area	2	128.431	0.310
Duration x Work Area	8	5.494	0.725
Left Upper Arm Elevation 44°-90°			
Duration	4	4.354	0.678
Work Area	2	79.348	0.421
Duration x Work Area	8	3.765	0.849
Left Upper Arm Elevation > 90°			
Duration	4	0.152	0.773
Work Area	2	2.312	0.141
Duration x Work Area	8	0.303	0.530
Right Upper Arm Elevation 0°-44°			
Duration	4	4.173	0.903
Work Area	2	1015.456	0.198
Duration x Work Area	8	19.298	0.321
Right Upper Arm Elevation 44°-90°			
Duration	4	9.578	0.516
Work Area	2	269.025	0.555
Duration x Work Area	8	0.792	0.612
Right Upper Arm Elevation > 90°			
Duration	4	2.381	0.632
Work Area	2	240.076	0.008*
Duration x Work Area	8	3.490	0.487
Trunk Inclination < 45°			
Duration	4	18.568	0.195
Work Area	2	239.150	0.012*
Duration x Work Area	8	29.809	0.023*
Trunk Inclination ≥ 45°			
Duration	4	18.593	0.194
Work Area	2	239.353	0.012*
Duration x Work Area	8	29.748	0.023*

*Significant at α = 0.05

The Hartley's test or F_{max} test was used to compare the variance of each of the shorter sampling durations versus the full-shift measurements. Based on the three body parts and posture categories, 32 test statistics were computed across the different sampling durations. A critical F value of 2.4 for the upper arms and 2.48 for the trunk were used based on 15 degrees of freedom for the upper arms and 14 degrees of freedom for the trunk at a significance level of 0.05. When comparing the full-shift to the 30-min sampling duration across the left and right upper arms and trunk, six of the eight test statistics had F_{max} values greater than the critical F value. These six F_{max} values included all three of the left upper arm postures, right upper arm elevation > 90° and both of the trunk postures. One F_{max} value, right upper arm elevation > 90°, for the full-shift versus the 1-hr sampling duration was greater than the critical F value. Based on these results, the standard deviation of the percent time logged for the 30-min sampling duration was significantly greater than the standard deviation of the percent time logged for the full-shift measure for all left upper arm postures, right upper arm elevation > 90°, and both trunk postures. The standard deviation of the percent time logged for the 1-hr sampling duration was significantly greater than the standard deviation of the percent time logged for the full-shift measure for right upper arm elevation > 90°. In addition, both of the trunk posture F_{max} values were borderline significant when comparing the full-shift to the 2-hr sampling duration.

Results of the assessment of bias and limits of agreement for the right and left upper arms and trunk are provided in Tables 6.8-6.10. Mean differences for the left upper arm tended to increase as the sampling durations became shorter, with the exception of the full shift vs the 30-min sampling duration (Table 6.8). The limits of agreement increased consistently as the sampling duration times decreased for left upper arm postures. The largest widths of agreement were found in the 1-hr and 30-min sampling durations when compared with the full-shift measures for left upper arm postures. Almost half of the estimated biases for the left upper arm postures were negative, indicating that the shorter sampling durations over-estimated exposure as compared to the full-shift measures. Findings were similar for the right upper arm (Table 6.9); however, the widths of agreement for the right upper arm were all consistently higher than the left upper arm. Estimates computed for the trunk had similar patterns as those reported for the left and right upper arms (Table 6.10).

TABLE 6.8

Mean Differences, Bias, and Limits of Agreement for Full-Shift versus Shorter Sampling Intervals for the Left Upper Arm

Posture	**Sampling Interval**			
	Full vs 4-Hr	**Full vs 2-Hr**	**Full vs 1-Hr**	**Full vs 30-Min**
0°-44°				
Mean Difference	-0.557	0.155	-0.726	0.763
Bias	-0.424	0.075	-0.164	0.168
Upper Limit of Agreement	2.070	4.295	8.141	9.842
Lower Limit of Agreement	-3.185	-3.985	-9.592	-8.315
Width of Limit of Agreement	5.255	8.280	17.733	18.157
45°-90°				
Mean Difference	0.349	-0.450	0.328	-0.870
Bias	0.333	-0.289	0.078	-0.211
Upper Limit of Agreement	2.445	2.780	8.724	7.362
Lower Limit of Agreement	-1.747	-3.719	-8.067	-9.101
Width of Limit of Agreement	4.192	6.499	16.791	16.463
> 90°				
Mean Difference	0.046	0.164	0.291	-0.057
Bias	0.111	0.337	0.386	-0.054
Upper Limit of Agreement	0.886	1.133	1.802	2.045
Lower Limit of Agreement	-0.793	-0.806	-1.219	-2.159
Width of Limit of Agreement	1.679	1.939	3.021	4.204

TABLE 6.9

Mean Differences, Bias, and Limits of Agreement for Full-Shift versus

Shorter Sampling Intervals for the Right Upper Arm

Posture	**Sampling Interval**			
	Full vs 4-Hr	**Full vs 2-Hr**	**Full vs 1-Hr**	**Full vs 30-Min**
0°-44°				
Mean Difference	-0.148	-1.041	-1.094	-0.741
Bias	-0.075	-0.324	-0.165	-0.106
Upper Limit of Agreement	3.793	5.395	12.164	13.275
Lower Limit of Agreement	-4.089	-7.477	-14.352	-14.757
Width of Limit of Agreement	7.882	12.872	26.516	28.032
45°-90°				
Mean Difference	0.418	1.040	1.225	1.930
Bias	0.243	0.374	0.278	0.294
Upper Limit of Agreement	3.858	6.597	10.047	15.063
Lower Limit of Agreement	-3.021	-4.517	-7.597	-11.203
Width of Limit of Agreement	6.879	11.114	17.644	26.266
> 90°				
Mean Difference	-0.319	0.018	-0.173	-1.323
Bias	-0.522	0.018	-0.046	-0.635
Upper Limit of Agreement	0.904	2.049	7.390	2.841
Lower Limit of Agreement	-1.541	-2.013	-7.736	-5.486
Width of Limit of Agreement	2.445	4.062	15.126	8.327

TABLE 6.10

Mean Differences, Bias, and Limits of Agreement for Full-Shift versus

Shorter Sampling Intervals for the Trunk

Posture	**Sampling Interval**			
	Full vs 4-Hr	**Full vs 2-Hr**	**Full vs 1-Hr**	**Full vs 30-Min**
< 45°				
Mean Difference	0.135	0.016	0.051	2.243
Bias	0.080	0.006	0.015	0.306
Upper Limit of Agreement	3.498	5.639	6.979	16.886
Lower Limit of Agreement	-3.228	-5.601	-6.876	-12.400
Width of Limit of Agreement	6.726	11.240	13.855	29.286
≥ 45°				
Mean Difference	-0.132	-0.019	-0.039	-2.241
Bias	-0.079	-0.007	-0.011	-0.306
Upper Limit of Agreement	3.221	5.604	6.880	12.406
Lower Limit of Agreement	-3.486	-5.642	-6.957	-16.888
Width of Limit of Agreement	6.707	11.246	13.837	29.294

DISCUSSION

The evaluation of the dimensions of and determinants of exposure variability can be used to plan exposure measurements, assign estimates of exposure, or predict and control future exposures. While a considerable number of studies have evaluated physical workplace exposures, most studies have only evaluated short periods of exposure as a representation of total exposure. While much research has been devoted to improvements of direct technical measurements in regards to reliability and validity, few studies have investigated exposure variability for ergonomics exposure assessment tools (Ortiz et al., 1997; Allread et al., 2000; Anton et al., 2003; Mathiassen et al., 2003; Dahlberg et al., 2004; Moller et al., 2004; Svendsen et al., 2005; Hansson et al., 2006). In addition, even fewer studies have evaluated exposure variability when measured over different sampling durations by direct reading instrumentation (Trask et al., 2008; Mathiassen et al., 2003).

Full-shift measures of inclinometry are few in the ergonomics literature (Trask et al., 2007; Trask et al., 2006; Moller et al., 2004; Jansen et al., 2001). While full-shift measures using direct instrumentation have been recommended to obtain exposure measures to ergonomic risk factors such as awkward postures, application has previously been difficult due to costs, sampling duration, battery limitations, inadequate memory, and large amounts of data to reduce. However, current inclinometry applications, like the Virtual Corset, allow for full-shift sampling over multiple days. To assist with evaluation of the inclinometry data, LabVIEW software (National Instruments: Austin, TX) to create a user interface for interactive software system control. LabVIEW was used to create a user data reduction interface that provided the ability to analyze the data using EVA, amplitude probability distribution function (APDF), and spectral analysis. Only the EVA feature was used in the present study based on the pre-defined posture categories. The amplitude probability distribution function has been used to quantify aspects of exposure variation. However, the APDF omits important aspects of the exposure variation (Mathiassen and Winkel, 1991). The APDF does not consider the length of the analysis period or reflect changes in the distribution of the variable of interest along a real-time scale (Mathiassen and Winkel, 1991). The present study chose to apply exposure variation analysis to the inclinometry data in an effort to simplify the data reduction process and provide the exposure metrics of the percent time spent in a pre-defined posture as well as the length of time (duration) within that posture. The EVA was used to obtain posture data for the full-shift measures as well as the sub-samples. Mathiassen and

Winkel (1991) developed EVA as a data reduction method to assist with effectively quantifying electromyography and electrogoniometry data. For the present study, the EVA was used to describe the percent time spent in pre-defined posture categories.

The present study demonstrated the application of full-shift inclinometry for postures of the upper arms and trunk of workers performing manufacturing tasks. The largest amount of time across all workers, both shift, anatomical areas, and posture categories was spent in the lowest inclination categories. Silverstein et al. (2008) reported an association and increased odds between upper arm flexion ≥ 45 for ≥ 15% of the time with pinch gripand rotator cuff syndrome. The mean percent time logged in left upper arm elevation ≥ 45° for the present study was 10.49%, while the mean percent time logged in right upper arm elevation ≥ 45° was 17.16%. The kegging work area had the highest percent time logged in right upper arm elevation ≥ 45° (23.15%). Based on these findings and the study by Silverstein et al. (2008), the right shoulder for all three work areas and particularly the kegging work area may have been at increased odds for the development of rotator cuff syndrome; however, other factors found by Silverstein et al. (2008) such as age and body mass index would have to be considered. No subjects in the present study reported any musculoskeletal symptoms during the course of the study.

Increased risk of back disorders has been associated with mild trunk flexion more than 10% of the time (Punnett et al., 2000). Other research has demonstrated an increased risk for back disorders if the number of lifts exceeded 120 per hour (Marras et al., 1995). The mean percent time recorded for trunk inclination ≥ 45° measured at the T6 spinous process for the present study was 5.23% with the highest mean percent time recorded for the bottling work area (7.72%). Previous research studies using inclinometry to measure trunk inclination are few (Hansson et al., 2006; Trask et al.,2006; Hansson et al., 2001; Juul-Kristensen et al.,2001; Burdorf et al., 1992) and much of this research was performed to compare inclinometry with other exposure assessment tools. The majority of the above referenced studies did not designate specific posture categories, but utilized percentiles to report the inclinometry data. Burdorf et al., (1992) used a trunk flexion cut-point of 20° and Trask et al. (2006) evaluated trunk flexion > 45° and > 60° when evaluating full-shift inclinometry and EMG. The present study modified the trunk flexion cut-points proposed by Bloswick and Villnave (2000) of 0°-23°, 23°-45°, 45°-67°, and 67°-90° into < 45° and ≥ 45° categories used to measure trunk inclination. While the mean percent time spent in trunk inclination among the manufacturing tasks evaluated in the present study did not reach the 10% time associated with back disorder as reported by Punnett et al. (2000), changing the trunk posture categories to smaller intervals may show an increase in percent time of trunk inclination. Future research should evaluate power spectrum analysis for inclinometry data as well as the angular velocity of movements. This would provide information for four exposure dimensions of posture: time spent in a specific posture, length of time (duration) within that specific posture, repetition, and angular velocity.

Sampling Duration

The length of time researchers should sample postures to obtain an exposure profile representative of a full-shift or full work day is a challenging question. Much ergonomics research has focused on capturing short segments of exposure information, due to time, cost, and measurement restrictions, and extrapolating this information to predict a full shift's exposure. Continuous full-shift exposure measurements are uncommon. The present study performed an initial assessment in an attempt to begin to answer the question of the length of time needed to adequately assess posture using inclinometry. While full-shift sampling provides the most comprehensive exposure profile of a person's exposure to awkward postures throughout their work day, shorter sampling durations have several benefits. A major benefit is the ability to obtain more assessments within the same day, thereby optimizing efficiency and reducing costs. Shorter sampling durations would allow for more individual measurements for the same time and cost (Trask et al., 2008).

To the authors' knowledge, this is the first study to compare measures of posture exposure using different sampling durations with inclinometry. Sampling duration will ultimately depend on the particular exposure metric researchers want to measure. For the present study, the means of the shorter duration sampling intervals did not vary substantially. This finding was similar to that reported by Trask et al. (2008). Trask et al. (2008) evaluated full-shift EMG of 103 workers performing tasks in five heavy industries, with a second full-shift measurement repeated for 35 of the workers. The full-shift EMG was then re-sampled at shorter intervals (4-hr, 2-hr, 1-hr, 10-min, and 2-min) and analyzed using Pearson product moment correlation coefficients, repeated measures ANOVA, percentage difference, bias, and limits of agreement (Trask et al., 2008). The authors (Trask et al., 2008) found that the shorter sampling durations tended to overestimate the full-shift exposure for percentiles below the median (50^{th} percentile) and underestimate the full-shift exposure for percentile above the median.

In the present study, the 30-min sampling duration tended to overestimate the full-shift exposure; however, this was specific to some, but not all postures and anatomical areas. Trask et al. (2008) utilized percentiles to compare the different sampling durations, whereas the present study utilized the percent time spent in specific posture categories to compare the different sampling periods. An overestimation of full-shift exposure in the 30-min sampling can be explained by the fact that for most cases, the 30-min duration did not include a break time. However, some of the 30-min durations in the present study included machine downtime that could be considered equivalent to break time for some situations. Specific posture categories demonstrated the same trend for all of the shorter sampling durations. Right upper arm elevation 0°-44°, right upper arm elevation > 90°, and trunk inclination ≥ 45° all had higher mean estimates for all the shorter sampling durations than the full-shift measures. Other posture categories tended to underestimate the full-shift exposure across all shorter sampling durations. This included right upper arm elevation 45°-90° and trunk inclination < 45°.

While the overall means of the different sampling durations did not vary substantially in the present study, the standard deviations for the means tended to increase as the sampling duration intervals became smaller (Table 6.8). The evaluation of the standard deviations or variance using the Hartley's F_{max} statistics revealed that the standard deviations for the majority of the 30-min sample duration were significantly greater than the full-shift measures. Standard deviation provides a measure of the variability or dispersion. A low standard deviation indicates that the data points tend to be close to the mean, whereas a high standard deviation indicates that the data are spread out over a large range of values. The analysis of the standard deviations indicated that while the comparisons of the shorter durations with the full-shift measures had similar means, the ranges of values were different.

Correlations as reported by Trask et al. (2008) indicated that full-shift measures were significantly correlated (0.344-0.969) to the shorter sampling durations, except for the 2-min measure. Correlations were highest for the 4-hr and 2-hr measures and decreased with shorter sampling durations (Trask et al., 2008). In the present study, full-shift measures were significantly correlated to the shorter sampling durations, with the exception of seven correlations (Table 6.9). In the present study, correlations were highest for the 4-hr and 2-hr sampling durations (0.80-0.99) and tended to decrease with the 1-hr and 30-min sampling durations (0.39-0.95). In some cases the 1-hr correlation was lower than the 30-min correlation. This could be explained by error associated with obtaining one sub-sample for each shorter duration. The two non-significant correlations at a significance level of 0.05 included left upper arm elevation 0°-44° and left upper arm elevation 45°-90° for the 1-hr sampling duration. The five non-significant correlations at a significance level of 0.01 included left upper arm elevation > 90° and both trunk postures for the 1-hr sampling duration and both trunk postures for the 30-min sampling duration. These findings indicate that the correlation of the shorter sampling durations to the full-shift measures can vary depending on the body part and posture evaluated.

It was not entirely clear why the left arm postures had non-significant correlations for the 1-hr sampling duration while the right arm postures had significant correlations. This indicates that some factor affected the left arm, typically the non-dominant arm, differently than the right arm for the 1-hr sampling durations. The researchers expect that since the right arm was dominant for most subjects, the right arm postures would be more consistent throughout the shift, since the subjects used this arm to perform the majority of the arm movements required. This is noted in the higher percentage of time spent in the

45°-90° and > 90° posture categories as compared to the left arm (Tables 6.2 and 6.3). In regards to the trunk, correlations were much lower for the 1-hr and 30-min sampling durations than the 4-hr and 2-hr sampling durations. The lower sampling durations tended to overestimate trunk inclination ≥ 45° as compared to the overall shift. This finding indicates that the shorter sampling durations across the 15 subjects and two shifts likely logged specific work events that required trunk inclination ≥ 45° as compared to the longer sampling durations that captured downtime and breaks.

The repeated measures ANOVA results, as reported by Trask et al. (2008), demonstrated significant differences between the full-shift and 2-min measures for the percentile categories evaluated. The present study also utilized repeated measures ANOVA to evaluate the different sampling durations, but also included work area (bottling, kegging, and canning) as a between-subjects factor to evaluate whether the sampling duration was affected by the work area. The repeated measures ANOVA results of the present study found significant differences between work area for right upper arm elevation > 90° and both trunk postures (Table 6.10). This indicated that the percent time logged in these posture categories significantly differed between the three work areas. This makes intuitive sense since analysis of the data by work area (Tables 6.2-6.4) indicated higher percentages for right upper arm elevation > 90° for the kegging area and a higher percentage of time logged in trunk inclination > 45° for the bottling work area. The repeated measures ANOVA also found significant differences for the duration x work area interaction for both trunk inclination categories. This finding indicates that the percent time in trunk postures as obtained by the sampling durations (full-shift, 4-hr, 2-hr, 1-hr, and 30-min) significantly differed across the three work areas. This indicates that sampling duration requirements may not only be different for certain body parts and postures, but may depend on the nature of the task.

Trask et al. (2008) reported that bias calculated did not exceed 1% of the reference contraction for the mean, 10th percentile, and 90th percentile at any of the sampling durations for evaluation of EMG data. Half of the biases reported by Trask et al. (2008) were negative, indicating an over-estimation of exposure for the shorter sampling durations. The limits of agreement had the smallest range for the 10th percentile and increased consistently with decreasing sampling duration (Trask et al., 2008). The authors (Trask et al., 2008) reported increasing widths of agreement with decreasing sampling duration. The researchers of the present study found similar results to the Trask et al. (2008) study. Bias and the widths of the limits of agreement increased consistently with decreasing sampling duration, with the exception of the full-shift vs 30-min sampling duration for right upper arm elevation > 90°. The 4-hr and 2-hr sampling durations more closely resembled the full-shift data as compared to the 1-hr and 30-min sampling durations. The limits of agreement for the 4-hr sampling durations for upper arm posture of 45°-90° did not exceed 4% in either direction at the calculated two standard deviations from the mean. Since the bias estimates were positive, the 4-hr sampling duration underestimated the exposure found by the full-shift measures. This indicates that the 4-hr sampling duration exposure estimate could underestimate upper arm elevation exposure of 45°-90° by 4%. Depending on the objectives of the research studies, this 4% difference could be important of association with musculoskeletal outcome is dependent not only on the angular deviation of the upper arm, but the percent time spent in that angular deviation as reported by Silverstein et al. (2008). The limits of agreement for the 2-hr sampling durations for upper arm posture of 45°-90° were slightly higher, but did not exceed 6.6%.

Limitations

The present study results and conclusions were based on a limited sample of 15 subjects. However, each subject was sampled over two shifts. The small sample size could have affected the ability to detect differences between the full-shift and shorter sampling durations.

Based on the function of the inclinometers and work environment, the full-shift measures included breaks. When considering representative exposure estimates, breaks are part of a worker's exposure profile and should be considered. Full-shift sampling allows for this consideration, while shorter intervals may not capture breaks. In addition, shorter sampling intervals may not capture postural variation that may be experienced during a full work shift. Exposure estimates obtained for shorter intervals used to estimate a worker's overall exposure should account for breaks; however, this would require *post hoc* statistical adjustment during the extrapolation. When comparing shorter sampling durations to a full-shift, the inclusion of breaks can complicate the comparison, particularly if 15-30 minutes of the 30-min sampling duration was break time. The analysis software was not programmed to extract segments of the data from the full-shift measures. When performing the random sampling of the shorter duration intervals, efforts were made to avoid obvious subject breaks/lunch. The 4-hr sampling durations always included at least one break based on the duration of the sample. Detailed task logs maintained by the researchers were used to verify work breaks.

Another limitation of the present study was the limited re-sampling of the shorter duration intervals. One re-sample for each shorter duration interval was obtained per subject. While limited, the sampling was performed randomly with replacement, allowing the different samples to be independent. Mathematically, this means that the covariance between the different samples is zero. Further analysis utilizing bootstrapping, constructing a large number of re-samples obtained by random sampling with replacement, is being considered as an application to this dataset.

CONCLUSIONS

Full-shift inclinometry of the upper arms and trunk is feasible given adequate attachment procedures, short set-up and attachment time, and programming that allows for quick data reduction into usable exposure metrics. While full-shift samples are considered ideal compared to shorter intervals, the results of the present indicated that sampling durations of 4-hrs or 2-hrs could provide representative exposure estimates depending on the body parts, postures, and nature of the tasks assessed. The 1-hr and 30-min sampling durations consistently over- or under-estimated the exposure for the body parts and postures assessed. The researchers of the present study recommend to sample at least 2-4 hours based on the results of the 15 workers, body parts and postures evaluated, and seven tasks performed in three work areas of a manufacturing facility. Tasks were typically cyclic and repetitive in nature, but involved non-repetitive activities such as troubleshooting, paperwork, testing, and fork truck use. Depending on the purpose of the research, the body parts and postures evaluated, and the type of work being evaluated, shorter sampling intervals (1-hr and 30-min) may still be appropriate. Extremely repetitive, cyclic tasks may only require a 30-min sample to obtain representative exposure estimates.

REFERENCES

Aaras, A., Westgard, R.H., and Stranden, E. Postural angles as an indicator of postural load and muscular injury in occupational work situations. *Ergonomics*, 31: 915-933.

Akesson, I., Hansson, G.A., Balogh, I., Moritz, U., and Skerfving, S. 1991. Quantifying work load in the neck, shoulders, and wrists in female dentists. *International Archives of Occupational and Environmental Health*, 69: 461-474.

Allread, W.G., Marras, W.S., and Burr, D.L. 2000. Measuring trunk motions in industry: variability due to task factors, individual differences, and the amount of data collected. *Ergonomics*, 43(6): 691-701.

Anton, D., Cook, T.M., Rosecrance, J.C., and Merlino, L.A. 2003. Method for quantitatively assessing physical risk factors during variable noncyclic work. *Scandinavian Journal of Work, Environment & Health,* 29(5): 354-362.

Ariens, G.A., van Mechelen, W., Bongers, P, Bouter, L.M., and van der Waal, G. 2000. Physical risk factors for neck pain. *Scandinavian Journal of Work, Environment & Health,* 26: 7-19.

Armstrong, T.J., and Chaffin, D.B. 1979. Carpal tunnel syndrome and selected personal attributes. *Journal of Occupational Medicine*, 21: 481-486.

Armstrong, T.J., Foulke, J.A., Joseph, B.S., and Goldstein, S.A. 1982. Investigation of cumulative trauma disorders in a poultry processing plant. *American Industrial Hygiene Association Journal*, 43(2): 103-116.

Armstrong, T.J., Castelli, W., Gaynor Evans, F., and Diaz-Perez, R. 1984. Some histological changes in carpal tunnel contents and their biomechanical implications. *Journal of Occupational Medicine,* 26: 197-201.

Balogh, I., Orbaek, P., Ohlsson, K. Nordander, C., Unge, J., Winkel, J., and Hansson, G.A. Self-assessed and directly measured occupational physical activities-influence of musculoskeletal complaints, age and gender. Applied Ergonomics, 35: 49-56.

Bao, S., Howard, N., Spielholz, P., Silverstein, B., and Polissar, N. Interrater reliabi,ity of posture observations. 2009. *Human Factors*, 51(3): 292-309.

Bartko, J. J. 1994. Measures of Agreement: A Single Procedure. *Statistics in Medicine,* 13: 737-745.

Baty, D., Buckle, P.N., and Stubbs, D.A. 1986. Posture recording by direct observation questionnaire assessment and instrumentation: a comparison based on a recent field study in N. Corlott and J Wilson (eds), *The Ergonomics of Working Postures*. London. Taylor and Francis, 293-292.

Bernmark, D. and Wiktorin, C. 2002. A triaxial accelerometer for measuring arm movements. *Applied Ergonomics*, 33: 541-547.

Bland, J.M. and Altman, D.G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet,* 8(1): 307-310.

Bloswick, D.S. and Villnave, T. 2000. Ergonomics (Chapter 54). In Harris, R.E. (ed), *Patty's Industrial Hygiene and Toxicology*, 5th ed. Vol. 4, New York: John Wiley and Sons.

Bureau of Labor Statistics. 2005. *Lost-Worktime Injuries and Illnesses: Characteristics and Resulting Days Away from Work, 2003.* United States Department of Labor News. Washington, D.C. USDL 05-521. www.bls.gov/iff/home.htm

Bureau of Labor Statistics. 2007. Workplace injuries and illnesses in 2006. Washington, D.C.: US Department of Labor, Bureau of Labor Statistics.

Bureau of Labor Statistics. 2008. Workplace injuries and illnesses in 2007. Washington, D.C.: US Department of Labor, Bureau of Labor Statistics.

Burdorf, A., Derksen, J., Naaktgeboren, B., and van Riel, M. 1992. Measurement of trunk bending during work by direct observation and continuous measurement. *Applied Ergonomics*, 23(4): 263-267.

Burdorf, A. 1995. Reducing random measurement error in assessing postural load on the back in epidemiological surveys. *Scandinavian Journal of Work and Environmental Health*, 21: 15-23.

Burdorf, A., and van der Beek, A. 1999. Exposure Assessment Strategies for Work-Related Risk Factors for Musculoskeletal Disorders. *Scandinavian Journal of Work and Environmental Health,* 25(4): 25-30.

Burns, K. J. 1998. Beyond Classical Reliability: Using Generalizability Theory to Assess Dependability. *Research in Nursing and Health,* 21: 83-90.

Burt, S. and Punnett, L. 1999. Evaluation of Interrater Reliability for Posture Observations in a Field Study. *Applied Ergonomics,* 30*:* 121-135.

Centers for Disease Control and Prevention. 2001. *Guide to Evaluating the Effectiveness of Strategies for Preventing Work Injuries: How to Show Whether a Safety Intervention Really Works.* Deparment of Health and Human Services, Centers for Disease Control and Prevention, and National Institute for Occupational Safety and Health. DHHS (NIOSH) Publication No. 2001-119.

Chen, P. Y., & Krauss, A. 2004. Inter-rater reliability. In M.S. Lewis-Beck, A. Bryman, and T. F. Liao (Eds.). Encyclopedia of Research Methods for the Social Sciences (Vol. 2), 513-514. Newbury Park, CA: Sage.
Corlett, E. N., Madeley, S. J., and Manenica, I. 1979. Posture Targetting: A Technique for Recording Working Postures. *Ergonomics,* 22(3): 357-366.

Dahlberg, R., Karlqvist, L., Bildt, C., and Nykvist, K. 2004. Do work technique and musculoskeletal symptoms differ between men and women performing the same type of work tasks? *Applied Ergonomics*, 35: 521-529.

Dartt, A., Rosecrance, J., Gerr, F., Chen, P., Anton, D., and Merlino, L. 2009. Reliability of assessing upper limb postures among workers performing manufacturing tasks. *Applied Ergonomics*, 40: 371-378.

de Bruijn, I., Engels, A., and van der Gulden, J. W. J. 1998. A Simple Method to Evaluate the Reliability of OWAS Observations. *Applied Ergonomics,* 29(4): 281-283.

de Looze, M.P., Toussaint, H.M., Ensink, J. and Mangnus, C. The validity of visual observation to assess posture in a laboratory-simulated, manual material handling task. Ergonomics, 37(8): 1335-1343.

Denis, D., Lortie, M., and Bruxelles, M. 2002. Impact of Observer's Experience and Training on Reliability of Observations for a Manual Handling Task. *Ergonomics,* 45(6): 441-454.

DeShon, R. P. 2002. Generalizability theory. In F. Drasgow and N. Schmitt (Eds). Measuring and analyzing behavior in organizations (pp. 189-220). San Francisco, CA: Jossey-Bass.

Douwes, M., and Dul, J. 1991. Validity and Reliability of Estimating Body Angles by Direct and Indirect Observation. In Quéinnec, Y. and Daniellou, F. *Designing for everyone, Proceeding of the 11th Congress of the International Ergonomics Association. Taylor and Francis.* 885-887.

Ergonomics Analysis and Design research Consortium. 2003. *User's Manual for Multimedia Video Task Analysis™ (MVTA™).* Wisconsin: Wisconsin Alumni Research Foundation (WARF).

Federal Register. 2000. *29 CFR Part 1910, Ergonomics Program; Final Rule.*US Department of Labor, Occupational Safety and Health Administration, 65(220).

Fleiss, J. L., and Cohen, J. 1973. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measure of Reliability. *Educational and Psycholgical Measurement, 33:* 613-619.

Forsman, M., Hansson, G.A., Medbo, L., Asterland, P., and Engstrom, T. 2002. A method for evaluation of manual work using synchronized video recordings and physiological measurements. *Applied Ergonomics*, 33: 533-540.

Fransson-Hall, C., Gloria, R., Kilbom, A., Winkel, J., Karlqvist, L. and Wiktorin, C. 1995. A Portable Ergonomic Observation Method (PEO) for Computerized On-line Recording of Postures and Manual Handling. *Applied Ergonomics,* 26: 93-100.

Genaidy, A. M., Simmons, R. J., Guo, L., and Hidalgo, J. A. 1993. Can Visual Perception be used to Estimate Body Part Angles?. *Ergonomics,* 36(4): 323-329.

Gold, J.E., Park, J.S., Punnett, L. Work routinization and implications for ergonomic exposure assessment. *Ergonomics,* 1: 12-27.

Hansson, G.A., Asterland, P., Holmer, N.G., and Skerfving, S. (2001). Validity and reliability of triaxial accelerometers for inclinometry in posture analysis. Medical and Biological Engineering and Computing, 39: 405-413.

Hansson, G.A., Arvidsson, I., Ohlsson, K., Nordander, C., Mathiassen, S.E., Skerfving, S., and Balogh, I. 2006. Precision of measurements of physical workload during standardized manual handling. Part II: inclinometry of head, upper back, neck, and upper arms. *Journal of Electromyography and Kinesiology*, 16: 125-136.

Hartley, H.O. 1950. The use of range in analysis of variance. *Biometrika,* 37: 271-280.

Hignett, S., and McAtamney, L. 2000. Rapid Entire Body Assessment (REBA). *Applied Ergonomics,* 31*:* 201-205.

Hoffman, G. 1981. Tendinitis and bursitis. *American Family Practice.* 23: 103-110.

Jansen, J.P., Burdorf, A., and Steyerberg, E. 2001. A novel approach for evaluating level, frequency and duration of lumbar posture simultaneously during work. *Scandinavian Journal of Work and Environmental Health*, 27(6): 373-380.

Juul-Kristensen, B., Fallentin, N., and Ekdahl, C. 1997. Criteria for classification of posture in repetitive work by observation methods: A Review. *International Journal of Industrial Ergonomics,* 19: 397-411.

Juul-Kristensen, B., Hansson, G. Å, Fallentin, N., Andersen, J. H., Ekdahl, C. 2001. Assessment of Work Postures and Movements usnign a Video-Based Observation Method and Direct Technical Measurements. *Applied Ergonomics*, 32*:* 517-524.

Karhu, O., Kansi, P., and Kuorinka, I. 1977. Correcting Working Postures in Industry: a practical method for analysis. *Applied Ergonomics,* 8(4): 199-201.

Ketola, R., Toivonen, R., and 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.

Keyserling, W. M. 1986. Postural Analysis of the Trunk and Shoulders in Simulated Real Time. *Ergonomics,* 29(4): 569-583.

Kilbom, A. and Persson, J. 1987. Work technique and its consequences for musculoskeletal disorders. *Ergonomics,* 30: 273–279.

Kilbom, A. 1994. Assessment of Physical Exposure in Relation to Work-Related Musculoskeletal Disorders-what information can be obtained from systematic observations. *The Scandinavian Journal of Work and Environmental Health,* 20*:* 30-45.

Kraemer, H. C., and Bloch, D. A. 1988. Kappa Coefficients in Epidemiology: An Appraisal of a Reappraisal. *Journal of Clinical Epidemiology,* 41(10): 959-968.

Leskinen, T., Hall, C., Rauas, S., Ulin, S., Tönnes, M., Miikari-Juntura, E., and Takala, E. 1997. Validation of Portable Ergonomic Obersvation (PEO) Method using Optoelectronic and Video Recordings. *Applied Ergonomics,* 28(2): 75-83.

Levene, H. 1960. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press, 278-292.

Li, G., and Buckle, P. 1999. Current Techniques for Assessing Physical Exposure to Work-Related Musculoskeletal Risks, with Emphasis on Posture-Based Methods. *Ergonomics,* 42(5): 676-695.

Lievens, F., Sanchez, J. I., and de Corte, W. 2004. Easing the Inferential Leap in Competency Modeling: The Effects of Task-Related Information and Subject Matter Expertise. *Personnel Psychology,* 57: 881-904.

Loomis, D. and Kromhout, H. Exposure variability concepts and applications in occupational epidemiology. *American Journal of Industrial Medicine*, 45: 113-122.

Lowe, B. D. 2004a. Accuracy and Validity of Observational Estimates of Wrist and Forearm Posture. *Ergonomics,* 47(5): 527-554.

Lowe, B. 2004b. Accuracy and validity of observational estimates of wrist and forearm posture. *Ergonomics*, 47(5): 527-554.

Maclure, M., and Willett, W. C. 1987. Misinterpretation and Misuse of the Kappa Statistic. *American Journal of Epidemiology,* 126(2): 161-169.

Marras, W.S., Lavender, S.A., Leurgans, S.E., Fathallah, S.E., Ferguson, S.A., Allread, W.G., and Rajulu, S.L. 1995. Biomechanical risk factors for occupationally related low back disorders. *Ergonomics*, 38: 377-410.

Masse, L. C. and Heesch, K. C. 2002. Raters' Objectivity in Using the Compendum of Physical Activities to Code Physical Activity Diaries. *Measurement in Physical Education and Exercise Science,* 6(4): 207-224.

Mathiassen, S.E., and Winkel, J. 1991. Quantifying variation in physical load using exposure-vs-time data. *Ergonomisc,* 34(12): 1455-1468.

Mathiassen, S.E., Moller, T., and Forsman, M. 2003. Variability in mechanical exposure within and between individuals performing a highly constrained industrial work task. *Ergonomics,* 46(8): 800-824.

McAtamney, L., and Corlett, E. N. 1993. RULA: A Survey Method for the Investigation of Work-Related Upper Limb Disorders. *Applied Ergonomics,* 24(2): 91-99.

Moller, T., Mathiassen, S.E., Franzon, H., and Kihlberg, S. Job enlargement and mechanical exposure variability in cyclic assembly work. Ergonomics, 47: 19-40.

Moore, S. and 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: 443-458.

Müller, R., and Büttner, P. 1994. A Critical Discussion of Intraclass Correlation Coefficients. *Statistics in Medicine,* 13: 2465-2476.

National Institute for Occupational Safety and Health. 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, Public Health Service, Centers for Disease

National Institute for Occupational Safety and Health. 2001. *National Occupational Research Agenda for Musculoskeletal Disorders: Research Topics for the Next Decade, A Report by the NORA Musculoskeletal Disorders Team.* DHHS (NIOSH) Publication No. 2001-117. www.cdc.gov/niosh/2001.117

Control and Prevention, and National Institute for Occupational Safety and Health. DHHS (NIOSH) Publication No. 97-141.

Norkin, C. C., and White, D. J. 1987. *Measurement of Joint Motion: A Guide to Goniometery.* Philadelphia: F. A. Davis Company.

O'Brien, R.G. 1981. A simple test for variance effects in experimental designs. *Psychological Bulletin,* 89: 570-474.

Ortiz, D., Marcus, M., Gerr, F., Jones, W., and Cohen, S. Measurement variability in upper extremity posture among VDT users. *Applied Ergonomics*, 28(2): 139-143.

Palmerud, G., Forsman, M., Sporrong, H., Herberts, P., and Kadefors, R. 2000. Intramuscular pressure of the infra- and supraspinatus muscles in relation to hand load and arm posture. *European Journal of Applied Physiology*, 83: 223-230.

Pan, C. S., Gardner, L. I., Landsittel D. P., Hendricks, S. A., Chiou, S. S., and Punnett, L. 1999. Ergonomic Exposure Assessment: an application of the PATH systematic observation method to retail workers. Postures, Activities, Tools, and Handling. *International Journal of Occupational and Environmental Health, 5:* 79-87.

Panel on Musculoskeletal Disorders and the Workplace Commission on Behavioral and Social Sciences and Education. 2003.National Research Council and Institute of Medicine. *Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities.* Washington D.C.: National Academy Press.

Paquet, V.L., Punnett, L., Buchholz, B. 2001 Validity of fixed-interval observations for postural assessment in construction work. *Applied Ergonomics,* 32(3): 215-224.

Peretz, C., Goldberg, P., Kahan, E., Grady, S. and Goren, A. 1997. The variability of exposure over time: a prospective longitudinal study. *Annals of Occupational Hygiene*, 41(4): 485-500.

Priel, V. Z. (1974). A Numerical Definition of Posture. *Human Factors,* 16(6): 576-584.

Punnett, L. Fine, L.J., Keyserling, W.M, Herrin, G.D., and Chaffin, D.B. 1991. Back disorders and nonneutral trunk postures of automobile assembly workers. *Scandinavian Journal of Work and Environmental Health*, 17: 337-346.

Punnett, L., Fine, L.J., Keyserling, W.M., Herrin, G.D., and Chaffin, D.B. Shoulder disorders and postural stress in automobile assembly work. *Scandinavian Journal of Work and Environmental Health*, 26: 283-291.

Shavelson, R. J., and Webb, N. M. (1991). *Generalizability Theory.* California: Sage Publications, Inc.

Shoukri, M.M., Asyali, M.H., and Donner, A. 2004. Sample size requirements for the design of reliability study: review and new results. *Statistical Methods in Medical Research*, 13: 251-271.

Shrout, P. E. and Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin,* 86(2): 420-428.

Silverstein, B.A., Bao, S.S., Fan, Z.J., Howard, N., Smith, C., Spielholz, P., Bonauto, D., and Viikari-Juntura, E. 2008. Rotator cuff syndrome: personal, work-related psychosocial, and physical load factors. *Journal of Occupational and Environmental Medicine*, 50: 1062-1076.

Spielholz, P., Silverstein, B., Morgan, M., Checkoway, H., and Kaufman, J. (2001). Comparison of Self-Report, Video Observation and Direct Measurement Methods ofor Upper Extremity Musculoskeletal Disorder Physical Risk Factors. *Ergonomics,* 44(6): 588-613.

Stetson, D. S., Keyserling, W. M., Siverstein, B. A., and Leonard, J. A. (1991). Observational Analysis of the Hand and Wrist: A Pilot Study. *Applied Occupational Environmental Hygiene,* 6(11): 927-937.

Stevens, E. M. Jr., Vos, G. A., Stephens, J.P., and Moore, J. S. (2004). Inter-Rater reliability of the Strain Index. *Journal of Occupational and Environmental Hygiene, 1,* 745-751.

Svendsen, S.W., Mathiassen, S.E., and Bonde, J.P. 2005. Task based exposure assessment in ergonomic epidemiology: a study of upper arm elevation in the jobs of machinists, car mechanics, and house painters. *Occupational and Environmental Medicine*, 62: 18-26.

Tichauer, E.R. 1966. Some aspects of stress on forearm and hand in industry. *Journal of Occupational Medicine,* 8: 63-71.

Trask C., Koehoorn, M., Village, J., Johnson, P.W., and Teschke, K. 2006a. Modeling determinants of working exposures and exposure variability. Proceedings of the 16th Congress of the International Ergonomics Association. Maastricht, Netherlands: Abstract number 699.

Trask C., Koehoorn, M., Village, J., Morrison, J., Teschke, K., and Johnson, P.W. 2006b. Evaluating full-shift low back EMG and posture measurements for epidemiologic studies. Proceedings of the 16th Congress of the International Ergonomics Association. Maastricht, Netherlands: Abstract number 130.

Trask, C., Teschke, K., Village, J., Chow, Y., Johnson, P. Luong, N., and Koehoorn, M. 2007. Measuring low back injury risk factors in challenging work environments: an evaluation of cost and feasibility. *American Journal of Industrial Medicine*, 50: 687-696.

Trask, C., Teschke, K., Morrison, J., Johnson, P., Village, J., and Koehoorn, M. 2008. How long is long enough? Evaluating sampling durations for low back EMG assessment. *Journal of Occupational and Environmental Health*, 5(10): 664-670.

Van der Beek, A. J., van Gaalen, L. C., and Frings-Dresen, M. H. W. (1992). Working Postures and Activities of Lorry Drivers: A Reliability Study of On-Site Observation and Recording on a Pocket Computer. *Applied Ergonomics,* 23(5): 331-336.

Van Eerd, D., Hogg-Johnson, S., Mazumder, A., Cole, D., Wells, R., and Moore, A. 2009. Task exposures in an office environment: a comparison of methods. *Ergonomics*, 52(10): 1248-1258.

VanLeeuwen, D. M. (1997). Assessing Reliability of measurements with Generalizability Theory: An Application to Inter-Rater Reliability. *Journal of Agricultural Education,* 38(3): 36-42.

Walter, S.D., Eliasziw, M., and Donner, A. 1998. Sample size and optimal designs for reliability studies. *Statistics in Medicine*, 17: 101-110.

Wells, R., Mathiassen, S.E., Medbo, L., and Winkel, J. 2007. Time-a key issue for musculoskeletal health and manufacturing. *Applied Ergonomics*, 38: 733-744.

Westgard, R.H. and Winkel, J. Ergonomic intervention research for improved musculoskeletal health: a critical review. *International Journal of Industrial Ergonomics*, 20: 463-500.

Wiktorin, C., Mortimer, M., Ekenvall, L., Kilbom, Å., and Hjelm, E. W. (1995). HARBO, A Simple Computer-Aided Observation Method for Recording Work Postures. *Scandinavian Journal of Work and Environmental Health,* 21: 440-449.

Winkel, J. and Mathiassen, S.E. 1994. Assessment of physical work load in epidemiologic studies: concepts, issues, and operational considerations. *Ergonomics*, 37(6): 979-988.

Yen, T. Y., and Radwin, R. G. 2002. A Comparsion between Analysis Time and Inter-Analyst Reliability using Spectral Analysis of Kinematic Data and Posture Classification. *Applied Ergonomics*, 33*:* 85-93.

APPENDICES

APPENDIX A

Study I and II Special Analysis Procedures

These procedures were specific to the analysis of Study I as well as the larger prospective cohort study. These procedures include how postures of the neck, shoulder, and wrist were measured or estimated and any rules that were used in the analysis process. These procedures were used in the rater training and used as guidelines during analysis.

Procedures for the Measurement of Neck, Shoulder, and Wrist Postures

The following procedures and rules for measurement applied generally to all of the body parts assessed in this study. Due to the difficulties of video recording worker tasks in a manufacturing environment, rules of measurement had to be created to aid with postural estimation and to minimize error and limitations as much as possible. They were as follows:

1) When the view of the body landmarks to measure the neck, shoulder, or wrist were completely obstructed by equipment, other subjects, or the subject herself/ himself, the "missing data" category was assigned. These body landmarks included the nose, chin, ears, metacarpals, and elbow.
2) When a subject's neck/head, shoulder/arm, or hand/wrist were located inside equipment (refrigerator, etc) where the view of important body landmarks were obstructed, the "missing data" category was assigned.
3) When a subject's neck/head, shoulder/arm, or hand/wrist were not present in the camera views provided, the "missing data" category was assigned.

Measurement of Neck Posture

Neutral neck posture was defined as the range of motion between 20° extension and 45° flexion. Non-neutral neck postures were those neck positions greater than 20° extension or greater than 45° flexion. Flexion and extension of the cervical spine occur in the sagittal plane around a coronal axis. Neck postures are difficult to estimate from video. In manual goniometry, neck postures are measured using vertebra in the cervical region of the spinal column in contrast to the rest of the spinal column. During pilot video analysis, it was discovered that this type of measurement would be too difficult for video observation due to varying camera angles and differing body types.

After research, several trial methods, and some creativity, a neck measurement method found in Norkin and White's (1985) "Measurement of Joint Motion: Goniometry" was decided to be the best available method for video observation. In this method, the fulcrum of a goniometer is centered over the external auditory meatus. The proximal arm of a goniometer is aligned so that it is perpendicular or parallel to the ground. The distal arm is aligned with the base of the nares. Estimation of neck posture was aided by using angle transparencies discussed in the general procedures. To aid in neck posture estimation, one axis of the angle was aligned parallel to the ground or a representative object with the intersection of the two axes placed on the external auditory meatus of the ear. If the base of the nares fell below or rose above the other angle axis, the neck was marked as non-neutral flexion or extension.

Rules of Neck Measurement

Due to the difficulties of video recording worker tasks in a manufacturing environment, rules of neck measurement had to be created to aid with postural estimation and to minimize error and limitations as much as possible. They are as follows:

1) When a subject's neck/head was positioned in a neutral posture observed from one sagittal view, then turned away and was viewed from the other sagittal view without any noticeable flexion or extension, the "neutral" category was assigned.
2) When a subject's neck/head was positioned in a neutral posture observed from a sagittal view, then turned away with some movement in the up or down direction, the "missing data" category was assigned unless the neck/head was obviously greater than 45° flexion or greater than 20° extension.
3) Parallax error occurs when the camera views are not optimal. When the sagittal view of a subject was not adequate to estimate posture, subjective measurements had to be utilized. The posture of the neck could be more severe than it actually is. The "neutral" category was assigned for slight flexion or extension. The "flexion" category was assigned if flexion greater than 45° was obvious. In all other circumstances, the "missing data" category was assigned.
4) If a subject was side-bending and only a posterior view of the neck/head was available, the "missing data" category was assigned. If a subject was side-bending and a sagittal view of the neck/head is available, the "missing data" category was assigned unless the posture was obviously neutral or extreme flexion/extension.
5) When a subject was leaning forward at the waist, the measurement of flexion and extension changed slightly. The raters had to use a reference perpendicular to the

ground. One axis of the angle transparency was placed perpendicular to the ground with the intersection of the axes placed at the external auditory meatus. If the base of the nares exceeded the other axis, either the "flexion" or "extension" category was assigned.

6) The ear was a primary landmark used to estimate neck posture. Many subjects have hair that covers the ear when not pulled back. If hair was covering the ear, raters had to estimate the location of the ear and assign the appropriate posture category. Estimation was performed by using other body landmarks such as the eyes and chin.
7) When only a posterior or anterior view was available, the "missing data" category was assigned unless the neck/head was obviously in a neutral or extreme posture. When only a posterior view available, one could still observe neutral neck posture when the subject was looking straight ahead. However, as soon as the neck/head moved in an up or down direction, the category of "missing data" was assigned. If the neck/head flexed or extended greater than 45° flexion or 20° extension, the categories of "flexion" or "extension" were assigned.
8) If the primary landmarks (ear and nose) were obstructed, raters used the chin, eyes, and glasses as an alternative to estimate posture if appropriate.

Measurement of Shoulder Posture

Neutral shoulder posture was defined as shoulder flexion/abduction less than or equal to 60°. Estimation of glenohumeral motion was attempted. Flexion occurs in the sagittal plane around a coronal axis. Abduction occurs in the frontal plane around an anterior-posterior axis. Mild shoulder flexion/abduction was defined as flexion/abduction greater than 60° but less than or equal to 90°. To aid in estimation of mild shoulder flexion in MVTA, one axis of a 60° angle was aligned on the lateral midline of the thorax. The intersection of the two axes was placed at the glenohumeral joint. If the elbow rose above the other axis, it was denoted as mild flexion. To aid estimation of mild shoulder abduction when the camera view was anterior to a subject, one axis of a 60° angle was aligned with the sternum. The vertex of the two lines was placed at the glenohumeral joint. If the elbow rose above the other axis, it was denoted as mild abduction. To aid estimation of mild shoulder abduction when the camera view was posterior to a subject, one axis of the 60 degree angle was aligned with the vertebral column. Severe shoulder flexion/abduction was defined as flexion/abduction greater than 90°. Severe flexion/abduction was estimated in the same manner as mild flexion/abduction, except that a 90° angle was used instead of a 60° angle to aid in estimation.

Rules of Shoulder Measurement

Due to the difficulties of video recording worker tasks in a manufacturing environment, rules of shoulder measurement had to be created to aid with postural estimation and to minimize error and limitations as much as possible. They are as follows:

1) If part of the arm was visible such as the biceps, posture was estimated and the appropriate category was assigned.
2) If the subject was laterally leaning at the waist, shoulder abduction was estimated and assigned the proper category when the view was posterior or anterior to the

subject. However, if the view was sagittal to the subject, shoulder flexion was estimated and the appropriate category was assigned.

3) If the subject was leaning forward at the waist and only a posterior or anterior view is available, only shoulder abduction was estimated. If the view was sagittal to the subject, only shoulder flexion was estimated. Otherwise, "missing data" was assigned.

4) If only a posterior view was available and the subject turned in a manner where the shoulder of concern is on the far side of the subject's body, posture was estimated if the arm is visible or obviously in a neutral or extreme posture. However, if the subject moved their arm (flexion or abduction) while turning and the arm was visible, the "missing data" category shall be assigned.

5) If only a sagittal view was available and the shoulder of concern was on the opposite side of the subject's body, neutral and flexion positions were estimated and the appropriate category was assigned. However, if the shoulder was abducting, neutral and extreme (near 90°) postures were estimated, but the transition from neutral to extreme posture was categorized as "missing data."

Measurement of Wrist Posture

Neutral wrist posture was defined as wrist flexion and extension less than or equal to 30°. Flexion and extension of the wrist occurs in the sagittal plane around a coronal axis. Non-neutral wrist posture was defined as wrist flexion and extension greater than 30°. To aid estimation of wrist flexion greater than 30°, one axis of a 30° angle was aligned along the ulna with the intersection of the two axes at the lateral aspect of the carpal bones. If the metacarpals rose above or fell below the other angle axis, the posture was denoted as non-neutral.

Rules of Wrist Measurement

The wrist was extremely difficult to analyze due to many factors. Lighting, distance, and the views available played an important role in the ability to estimate wrist posture. Wrist movements were more subtle than the gross body movements of the shoulder and neck, therefore the transparency angles were rarely used to aid in estimation because they do not add much more certainty. Wrist posture designations were made by educated estimations. Due to the difficulties of video recording worker tasks in a manufacturing environment, rules of wrist measurement had to be created to aid with postural estimation and to minimize error and limitations as much as possible. They are as follows:

1) If the view of the wrist/hand was blocked for a short time and no obvious movement of the wrist occurred, the "neutral" category was assigned.
2) Pneumatic tools were often used by the subjects. When a subject used a pneumatic tool, the "neutral" posture category was assigned unless the wrist was obviously in flexion or extension greater than 30°.
3) When the camera view was located anterior to the subject and the hands were palms down or facing each other, posture was estimated and the appropriate category was assigned.
4) When the camera view was located posterior to the subject and the subject reached out to the side of their body making the wrist visible, posture was estimated and the appropriate category was assigned.

5) When the camera view was located sagittal to the subject and the hands were palms down, the posture was estimated and the appropriate category was assigned. When the hands were positioned with palms facing each other, the posture was estimated if the wrists were obviously neutral or in extreme flexion/extension. If the posture was questionable, the “missing data” category was assigned.

6) If the subject was relaxing in between task cycles and the arm was hanging down to the side of the body, the “neutral” category was assigned. When the wrist was not visible for short periods of time during this period, the posture was still designated as “neutral.”

7) If lighting was inadequate so that the raters could not see the wrist/hand, the “missing data” category was assigned.

8) If the view of the wrist was from a distance so great that it was improbable to estimate wrist posture adequately, the “missing data” category was assigned.