Applied Ergonomics 60 (2017) 275–281

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Observer variability in posture assessment from video recordings: The effect of partly visible periods

Catherine Trask [a, b, *], Svend Erik Mathiassen [a], Mehdi Rostami [c], Marina Heiden [a]

[a] Centre for Musculoskeletal Research, Department of Occupational and Public Health Sciences, University of Gävle, SE – 801 76 Gävle, Sweden
[b] Centre for Health and Safety in Agriculture, College of Medicine, University of Saskatchewan, 104 Clinic Place PO Box 23, Saskatoon, Saskatchewan, S7N 2Z4, Canada
[c] Collaborative Biostatistics Program, School of Public Health, University of Saskatchewan, Health Sciences building, 107 Wiggins Road, Saskatoon, Saskatchewan, S7N 5E5, Canada

A R T I C L E I N F O

Article history:
Received 29 February 2016
Received in revised form
4 November 2016
Accepted 16 December 2016

Keywords:
Observation
Inter-rater agreement
Posture
Exposure assessment
Pulp mill

A B S T R A C T

Observers rank partly visible postures on video frames differently than fully visible postures, but it's not clear if this is due to differences in observer perception. This study investigated the effect of posture visibility on between-observer variability in assessments of trunk and arm posture. Trained observers assessed trunk and arm postures from video recordings of 84 pulp mill shifts using a work sampling approach; postures were also categorized as 'fully' or 'partly' visible. Between-worker, between-day, and between-observer variance components and corresponding confidence intervals were calculated. Although no consistent gradient was seen for the trunk, right upper arm posture showed smaller between-observer variance when all observers rated a posture as fully visible. This suggests that, partly-visible data, especially when observers disagree as to the level of visibility, introduces more between-observer variability when compared to fully visible data. Some previously-identified differences in daily posture summaries may be related to this phenomenon.

1. Background

Observation is a common exposure assessment tool for musculoskeletal risk factors due to its accessibility, portability, and the range of exposures that may be included (Trask et al., 2007; Takala et al., 2010). However, studies show that observations may be associated with uncertainty related to the camera angle (Sutherland et al, 2007, 2008; Qu et al., 2012); the size of body parts and range of motion observed (Bao et al., 2009); and poor or insufficient lighting (Bao et al., 2009). Observers may also vary in their assessment of the same image. For example, between-observer variance can account for up to half of the total exposure variance in arm posture assessment from video recordings of real work, even when within-observer variance is accounted for (Rezagholi et al., 2012). Controlled experimental studies have shown some minor differences in posture estimates depending on the degree of experience and expertise of observers (Weir et al., 2007; Andrews et al., 2008). The influence of experience has also been demonstrated in studies showing that focused training of observers may lead to better agreement when rating postures during real work (van der Beek et al., 1992). However, the effects of observer training and experience alone seem unlikely to account for the large variability between observers reported in previous studies of working postures in real occupational settings (Bao et al., 2009; Rezagholi et al., 2012; Mathiassen and Paquet, 2010; Dartt et al., 2009).

Part of the uncertainty in observation measures may be related to incomplete visibility of postures that require approximation or inference based on the visible parts of the body and the occupational context in the image. A recent study of aircraft baggage handlers showed that almost 38% of video frames were rated as 'invisible' in terms of arm posture, with 43% deemed 'partly visible' and only 19% 'fully visible' (Trask et al., 2015). Daily posture summary variables were 20–45% lower for fully visible postures than partly visible ones. Essentially, this previous study showed that partly visible posture can occur frequently, which raised a concern about whether observer performance differs between partly and

* Corresponding author.Centre for Musculoskeletal Research, Department of Occupational and Public Health Sciences, University of Gävle, SE – 801 76 Gävle, Sweden.
E-mail addresses: catherine.trask@usask.ca (C. Trask), svenderik.mathiassen@hig.se (S.E. Mathiassen), mer521@mail.usask.ca (M. Rostami), marina.heiden@hig.se (M. Heiden).

http://dx.doi.org/10.1016/j.apergo.2016.12.009
0003-6870/

fully visible postures. Unfortunately, that dataset didn't include frames analyzed by multiple observers, so the question of between-observer differences remained unanswered.

Observation is inherently subjective, and the observer will rely on previous experiences when rating postures. Observers differ in ratings even when observing fully visible postures under ideal viewing conditions, and this variability can be expected to increase when additional cognitive elements are added to the observation process, such as having to base a rating on previous experiences of body configurations when parts of the body are not fully visible. Also, within-observer variability can be expected to differ depending on viewing conditions, since the individual observer may be more uncertain about ratings if they involve “guess-timating” to a larger extent. If these effects are substantial, this could inform observer training and decision-making rules or procedures for partly-visible situations, which could, in turn, lead to a better statistical performance of posture observation methods and thus less demands on resources invested in the observation process *per se* (Trask et al., 2014; Mathiassen et al., 2013).

The purpose of this study was to determine the extent to which visibility of postures in still images of paper mill workers influence summary statistics and between-observer variance of trunk and arm posture ratings. Considering the previously-reported factors influencing between-observer variance, we hypothesize that between-observer variability will be higher when postures are partly visible.

2. Methods

2.1. Study population and sampling

Twenty-eight workers (26 men and 2 women) at a pulp mill and paper products facility were recruited to participate in this study. Their mean age was 45 years (range 21–59 years), height 1.77 m (range 1.65–2.07 m), and weight 87 kg (range 61–170 kg). They had worked at the facility for, on average, 25 years (range 1–36 years). Each worker was visited 3 times for a total of 84 work shifts. The shifts were selected to obtain a balance between shift types occurring at the facility (*i.e.*, morning/afternoon/night and weekend day/night). During each shift, the workers were video recorded for up to 11 h using a single camera while performing their regular work tasks throughout their workplace. In total, the participating workers were video recorded for 604 h during work. The study was conducted in accordance with the declaration of Helsinki and approved by the Regional Ethical Review Board in Uppsala (2011/026).

2.2. Observers

Three observers analyzed the video using a customized software program (ViSPA) described previously (Trask et al., 2013, 2015). The observers included 2 men and one woman without specific coursework or training in ergonomics or anatomy prior to this study; collectively their academic backgrounds included language, technical writing, and political science with professional experience in equestrian training. This was deemed appropriate since prior knowledge has been shown to have negligible impact on observation error (Weir et al., 2011). The observers were trained using a standardized protocol that involved 30 h of instruction and observation practice with group discussions to build consensus and facilitate convergence of observation definitions.

2.3. Observation data collection and processing

Once trained, all three observers analyzed each of the 84 edited video recordings independently of one another. Still frames were selected at 135-second intervals, yielding up to 287 unique frames per shift.

For each frame, observers categorized the degree of visibility for the trunk and upper right arm as: completely invisible, partly visible, or fully visible. Partial visibility occurred when workers' body parts were partly occluded by other workers, equipment, or materials, or when part of the worker was cut off by the camera frame. In these cases, observers made an inference about the position of the body part using other body parts or cues about the task. Observers recorded partly and fully visible posture angles at a self-selected pace using software with an angle-matching graphical user interface as previously reported (Trask et al., 2013; Bao et al., 2007); postures were determined and stored by the software with 1-degree resolution relative to gravity for the trunk ($-180°$ (i.e. extension) to $+180°$) and upper arm ($0°$ to $+180°$). Observers also noted if the trunk was twisted more than 20°. The amount of time between displaying a still frame and the observer submitting the rating was automatically recorded by the computer.

Once observers had reviewed and recorded angles and visibility for the each video still frame, the ratings for each worker's work day were summarized into exposure summary metrics. For trunk flexion/extension, mean and 90th percentile angle were calculated, as well as time flexed greater than 45° and 60°. The percent time the trunk was laterally flexed more than 20° or twisted more than 20° were also calculated. Right upper arm posture was summarized into mean and 90th percentile inclination angle, as well as time elevated above 45° and 60°. Similar metrics have been used in previous publications describing occupational exposure (Wahlstrom et al., 2010; Palmerud et al., 2012; Kazmierczak et al., 2005) and can be considered ‘typical’ exposure metrics of interest to ergonomists and occupational health and safety professionals.

2.4. Statistical analysis

Although 16 156 unique frames were reviewed by observers, only those frames which were observed by all three observers (i.e., for which postures were not characterized by even one observer as ‘completely invisible’) were used in the analyses. When analysing the trunk, 10 877 frames were considered partly or fully visible by all three observers; 8684 frames for the right upper arm. For each frame, visibility was categorized into 4 possibilities: 1) all observers agreed that the frame is fully visible (FFF), 2) two observers said fully, one said partly (FFP), 3) one said fully, two said partly (FPP), and 4) all said partly (PPP). Posture summary estimates were calculated for frames in each of these four visibility categories. Thus, four sets of 84 daily summary estimates of each of the posture variables, distributed among 28 workers, were available from each of the three observers as a basis for further analysis. The 95% confidence intervals of daily summary estimates within each visibility category were calculated using Wald-based methods, except for confidence intervals of daily 90th percentile values, which were calculated using Woodruff's method (Francisco and Fuller, 1991; Särndal et al., 2003; Dorfman and Valliant, 1993).

Posture variance components associated with worker, shift and observer were then calculated using a model containing both hierarchical and crossed effects. Three shifts were available for each worker, and thus the effect of shift is nested within worker, but since each frame was analyzed by three observers, the observer level is not nested, but rather crossed with the other levels (Raudenbush, 1993):

$$h\left(y_{ijkl}\right) = \beta + b_i + b_{ij} + b_k + \varepsilon_{ijkl}, \quad (1)$$