Inter-rater reliability of PATH observations for assessment of ergonomic risk factors in hospital work

Jung-Keun Park[a,b,*], **Jon Boyer**[b], **Jamie Tessler**[b], **Jeffrey Casey**[b,c], **Linda Schemm**[b], **Rebecca Gore**[b], **Laura Punnett**[b], and **Promoting Healthy and Safe Employment (PHASE) in Healthcare Project Team**[b]

[a]Occupational Safety and Health Research Institute, Korea Occupational Safety and Health Agency, Incheon, South Korea

[b]Department of Work Environment, University of Massachusetts Lowell, Lowell, MA, USA

[c]Occupational Health Clinics for Ontario Workers Inc, Windsor, Ontario N8X 4G1, Canada

Abstract

This study examined the inter-rater reliability of expert observations of ergonomic risk factors by four analysts. Ten jobs were observed at a hospital using a newly expanded version of the PATH method (Buchholz *et al.* 1996), to which selected upper extremity exposures had been added. Two of the four raters simultaneously observed each worker onsite for a total of 443 observation pairs containing 18 categorical exposure items each. For most exposure items, kappa coefficients were 0.4 or higher. For some items, agreement was higher both for the jobs with less rapid hand activity and for the analysts with a higher level of ergonomic job analysis experience. These upper extremity exposures could be characterised reliably with real-time observation, given adequate experience and training of the observers. The revised version of PATH is applicable to the analysis of jobs where upper extremity musculoskeletal strain is of concern.

Keywords

inter-rater reliability; PATH method; ergonomic risk factor; hospital work

1. Introduction

Exposure to ergonomic risk factors for prolonged periods can lead to a variety of potentially disabling injuries and disorders of musculoskeletal tissues and peripheral nerves (Bernard 1997, National Research Council and Institute of Medicine 2001). Hospital work by its very nature is performed by employees with extremely diverse jobs such as, for example, administrators, physicians, nurses, laboratory technicians, clerical workers, food service, laundry and maintenance workers. Despite this great variety of work types, hospital employees as a group are at high risk for musculoskeletal disorders (MSDs) or injuries (Fuortes *et al.* 1994, Camerino *et al.* 2001, Gillen *et al.* 2007) due to the demands of their work, including repetitive or prolonged motions, awkward postures, forceful manual

exertions and handling of patients and heavy objects (Bru *et al.* 1994, Fuortes *et al.* 1994, Hignett and Richardson 1995, Hernandez *et al.* 1998, Lagerström *et al.* 1998, Messing *et al.* 1998, Elford *et al.* 2000, Owen *et al.* 2002, Janowitz *et al.* 2006, Park 2006).

A variety of observational techniques have been useful for characterising and evaluating ergonomic exposure to risk factors (e.g. Karhu *et al.* 1977, Stetson *et al.* 1991, McAtamney and Corlett 1993, Fransson-Hall *et al.* 1995, Kemmlert 1995, Wiktorin *et al.* 1995, Colombini 1998, Messing *et al.* 1998, Hignett and McAtamney 2000, Janowitz *et al.* 2006, Cann *et al.* 2008, David *et al.* 2008, Baker *et al.* 2009). Many of these approaches either assume regular work cycles as the basis of time sampling or do not explicitly provide a sampling protocol.

In visual job observation, agreement among analysts is believed to vary with the specific method used, the exposures being observed, the skill and training of the analysts and perhaps characteristics of the job as well (Paquet *et al.* 2001, Voskuijl and van Sliedregt 2002). Evaluations of inter-rater reliability (IRR) for ergonomic risk factors have been reported for some exposure assessment methods in various sectors including healthcare (Johnsson *et al.* 2004, Warming *et al.* 2004), educational services (Cann *et al.* 2008, Baker *et al.* 2009), general industry (Keyserling 1986, van der Beek *et al.* 1992, Burt and Punnett 1999) and construction (Buchholz *et al.* 1996, 2003). However, these studies were conducted mostly for limited sets of job titles and thus did not permit systematic examination of the factors that might influence agreement among analysts.

Posture, activity, tools and handling (PATH) is a direct observation method that was developed originally for observation of construction workers (Buchholz *et al.* 1996) because their jobs are highly variable over multiple timescales. The method has potential application to any type of non-routine work and has now also been used to analyse risk factors for hip and knee disorders in dairy farming (Howard 1997), manual materials handling (MMH) in retail stores (Pan *et al.* 1999) and apple harvesting (Fulmer *et al.* 2002). In the healthcare sector, it has also been utilised to analyse the work of home healthcare aides (Dybel 2000) and nursing home employees (Rockefeller 2002). Each adaptation involved modification of the template (exposure sampling form) to customise it for the specific content (tasks, tools, etc.) of that job or industrial sector.

In 2003, the PATH method was further adapted for a large exposure assessment study of multiple jobs in the healthcare industry. The revisions involved the addition of selected distal upper extremity risk factors, such as wrist and hand postures, that were not included in the original version.

The PATH method has acceptable validity and reliability for trunk, shoulder and leg postures and manual materials handling in the analysis of construction work (Buchholz *et al.* 1996, Paquet *et al.* 2001). However, because the upper extremity can move so much more quickly than the larger body masses previously characterised, it was not certain that these new exposure items could be recorded accurately in real time. Thus, the present study sought to examine the reliability of the newly revised PATH instrument and of the observers who were to analyse ergonomic exposure in hospital work. The specific goals were to examine inter-

rater agreement of PATH observations across a convenience sample of 10 jobs in one hospital and to evaluate whether agreement was higher with either rater's prior ergonomics experience or with slower hand motion speed in the task(s) observed.

2. Methods

2.1. Study site and subjects

This study was carried out at a hospital in north-eastern Massachusetts, USA. It was part of the exposure assessment effort for the epidemiological study conducted by the Promoting Healthy and Safe Employment (PHASE) project team at the University of Massachusetts Lowell (UML) (d'Errico *et al.* 2007, Cifuentes *et al.* 2008). For training purposes, UML employees were observed. In the final dataset, 12 hospital workers in 10 jobs were observed: nuclear medical technician, radiology technician, ultrasound staff, human resources staff and receptionist. These jobs were selected from the first hospital departments that agreed to participate in the initial stages of the study. The participants were nine female and three male workers; their ages ranged from 18 to 62 years. Each subject was approached for permission after his/her supervisor gave permission, agreed to participate voluntarily and signed an informed consent form approved by the Institutional Review Board at UML.

2.2. Materials

2.2.1. The expanded PATH instrument—Unlike the original PATH method (Buchholz *et al.* 1996), the new version consisted of three templates, which were completed in a fixed sequence and spacing; each observation took 90 s, with intervals of 45, 30 and 15 s for the three component templates. The first, 'whole body', template was very similar to the original PATH checklist, with nine items on trunk, leg, shoulder and elbow postures; tasks and activities (weight handled, MMH (yes/no) and MMH action); and other work conditions (noise level and vibration level) at one point in time (Table 1). The second, 'hand/forearm', template (eight items) covered hand/forearm risk factors such as type of hand grasp, wrist/forearm deviation, keyboard use, contact stress, vibration, weight in hands and also at a single point in time, at a fixed time interval after the first set of observations. The third, 'hand activity', template was included to determine hand activity level (HAL) (Latko *et al.* 1997) after continuous observation of a 15-s work period immediately following the second set. The HAL was scored as an integer from 0 to 10.

In addition, identifying information recorded about each job and shift observed included level of work routinisation: 1) single routine task; 2) multiple routine tasks; 3) single variable task; 4) multiple mixed tasks; 5) multiple variable tasks (Park 2000, Gold *et al.* 2006). The new items were selected on the basis of literature review (e.g. McAtamney and Corlett 1993, Kilbom 1994b, Latko *et al.* 1997, Dybel 2000, Rockefeller 2002) and discussion with a senior researcher (L.P.) who was involved since the beginning of PATH development in the mid-1990s. Operational definitions were developed for each item of each instrument and were reviewed iteratively during the analysts' training (see below). During the training period, and before beginning formal data collection, the PATH templates were iteratively revised as needed to eliminate ambiguity and correct identifiable sources of disagreement.

2.2.2. Computer and timer—The checklists for each job were converted into electronic templates with InspectWrite[TM] software (Penfact Inc., USA) and uploaded into two hand-held computers or personal digital assistants (PDAs) (Toshiba Pocket PC e310, using the Pocket PC 2002 operating system). Each PDA was fully recharged and the clock was checked before data collection. A desktop personal computer was used as an authoring workstation for transferring and storing exposure data from the PDAs. During the data collection for evaluation of inter-rater agreement, a stopwatch or electronic timer was used to standardise the time intervals and synchronise the observations of the two raters.

2.3. Rater training

Four raters, who were graduate student research assistants in the PHASE project, completed a 30-h training program, as recommended by the original PATH developers (Buchholz *et al.* 1996). The training program included explanation of the exposure assessment methods and instruments, procedures for data management and cleaning and practice data collection with feedback on agreement among observers. Raters A, B and C each had a master's degree in Occupational Ergonomics and at least 18 months of experience prior to training, while rater D had a master's degree in Industrial Hygiene and no previous experience in ergonomic job analysis.

As part of the training process, the four raters conducted IRR pilot-tests, analysing the activities of various workers from video recordings as well as in real time at worksites in the university. All raters were trained on the PDAs after successful observation practice with paper and pencil. The goal of evaluating agreement was to qualify each rater to collect independent ergonomic exposure data in the hospital. Different versions of the newly revised PATH instrument were used over the training period. These 'training data' were analysed as they were collected and areas of disagreement were explored in order to identify needed revisions in variable definitions or instructions to the analyst. Items 17 and 18, for example, were added in the last versions of the revised instrument.

2.4. PATH data collection

A walkthrough was conducted in each department prior to collection of PATH data. A brief interview was held with the volunteer worker and identifying information was recorded for the job and the worker.

For IRR evaluation, two of the four raters collected exposure data on each worker at the same time. All IRR datasets were collected during a 1-month period. A stopwatch or electronic timer was used to standardise the time intervals and synchronise the observations of the two observers. One rater was designated to generate vocal cues for the other: (1) at the beginning of the observation for the whole body template; (2) after 45 s to begin observation for the hand/forearm template; (3) after 30 s to start observation of HAL; (4) at the end of the 15-s observation to record the HAL, input data into the third template, and then immediately to commence the next cycle of observations.

2.5. Data analysis

2.5.1. Data management—Field data were taken back to the university and transferred from the PDA to the authoring workstation, managed for data cleaning and stored for future data analyses. The job and shift identification data were manually entered into a Microsoft Excel spreadsheet (Microsoft Corporation, Redmond, WA, USA).

2.5.2. Statistical analysis—Agreement between each pair of raters was evaluated for each item using percent agreement, as well as the kappa coefficient, either unweighted (dichotomous variables) or weighted (ordinal variables with three or more possible categories), with 95% CI (Fleiss and Cohen 1973, Fleiss 1981). Percent agreement was defined as the proportion of the total number of observations in which two raters recorded the same category (Cohen 1960). Percent agreement of 80% or higher was considered satisfactory (Warming *et al*. 2004).

Kappa is the proportion of agreement after chance-expected agreement is removed (Cohen 1960, Fleiss 1981). Each kappa or weighted kappa coefficient was interpreted on a three-degree scale: poor agreement for <0.4; fair to good agreement for 0.4–0.75; and excellent agreement for >0.75 (Fleiss, 1981). The kappa or weighted kappa coefficient was not computed for those items where the sample size did not meet the formula $2k^2$ for items ranging between $3 \le k \le 10$ categories (Cicchetti 1981).

For the HAL variable, the weighted kappa coefficient was calculated after classifying into three categories: low = 0–<3.3, medium = 3.3–<6.7 and high = 6.7–10. Intraclass correlation coefficient (ICC) was also considered for analysis of HAL data but could not be used because the HAL data were not normally distributed.

In order to compare the IRR between jobs with different levels of hand activity, the median HAL value was computed for each job and used for categorisation (low, medium, or high) as above. In the case of observations of multiple workers with the same job title, all HAL data for that job title were combined to compute the median. Among the 10 jobs, five were classified as low HAL and the rest of them as medium HAL. It was hypothesised that the IRR would be higher in observations collected from the lower HAL jobs rather than the higher HAL jobs.

For comparing IRRs among analysts, their experience in ergonomics and job analysis was classified into two levels: I: professional experience with ergonomic job analysis (raters A, B and C); and II: limited experience with short training time (rater D). Thus, rater pairs AB, AC and BC (group A) were all pairs of experienced raters while rater pairs AD, BD and CD (group B) were pairs in which one rater was inexperienced. It was hypothesised that the IRR would be higher in observations made by group A (i.e. rater pair with same experience level) rather than group B (i.e. rater pair with different experience levels).

The category of 'not observed or not sure' in each item was treated as a missing value in data analyses. Data analyses were conducted using SAS 9.1 (SAS Institute Inc. 2003).

3. Results

A total of 443 observation pairs were collected for 10 jobs in the hospital (Table 2). Eight jobs were observed by one pair of raters and two jobs were observed by two or three rater pairs. The observation duration for each job ranged from 27 min to 163 min. Rater B conducted the most PATH field observations (41% of the total field dataset), followed by rater D (29%), rater C (21%) and rater A (9%). All 10 jobs were multiple mixed tasks (work routinisation level = 4), which seems typical of many jobs in the healthcare setting.

3.1. Inter-rater reliability

The percent agreement levels ranged from 42.5 to 100 (Table 3). They were 80% or higher for 12 items. The kappa coefficients were 'excellent' for leg posture, keyboarding, hand/forearm vibration (item no. 15) and hand observed (right/left), 'fair to good' for trunk posture, shoulder/arm elevation, elbow posture, weight in hands (no. 5), neutral/gross grasp, wrist/forearm deviation, pinch grip, hand/forearm contact stress, weight in hands (no. 16) and HAL, 'poor' for MMH action, noise level and vibration (no. 9). The kappa coefficient was not defined for MMH.

The percent agreement levels were 80 or higher for six items in each of the whole body, hand/forearm and hand activity templates. The kappa coefficients were 0.44 or higher for all items in the distal upper extremity templates but 0.40 or higher for only five out of nine items in the whole body template.

3.2. Agreement by job level of hand activity

The percent agreement in the low HAL jobs was higher for each of 10 items when compared to that in the medium HAL jobs (Table 4). It was substantially lower for noise level and equal or negligibly lower for the remaining items. The kappa coefficient in the low HAL jobs was higher for each of eight items and lower for six items. The former eight items were among those 10 with higher percent agreement. Most differences in the kappa coefficients were small with overlapping confidence intervals. The kappa coefficients were undefined for hand/forearm vibration (no. 15) and MMH.

3.3. Agreement by ergonomic experience level of the raters

The percent agreement in group A was higher than that in group B for 13 items (Table 5) and lower only for two items, namely, elbow posture and weight in hands (no. 16). Percent agreement was equal between the groups for three items: MMH; hand/forearm vibration (no. 15); and hand observed. The kappa coefficient of group A was higher than that of group B for each of nine items while it was lower only for five items. The first nine items were among those 13 items with higher percent agreement. Differences in kappa values between the two groups varied. For noise level, there was a markedly large difference in the kappa coefficients and their confidence intervals between the two groups. Agreement could not be compared for vibration (no. 9), hand/forearm vibration (no. 15) and MMH since the kappa coefficients were not defined.

4. Discussion

In a convenience sample of 10 hospital jobs, an expanded version of the PATH instrument had fair to good IRR for 14 (78%) of the 18 ergonomic risk factors evaluated. Overall, agreement was at least as good for the new upper extremity items as for the original PATH variables. Exposures to the upper extremity were generally rated with slightly higher agreement in the lower HAL job groups and by the observers with more ergonomics experience. Overall, it proved feasible to utilise PATH observations to characterise ergonomic exposure such as non-neutral postures, repetitive hand motion, contact stress and weight in the hands. Thus, it appears possible to assess risk of work-related hand/forearm MSDs in handintensive jobs such as laboratory technician or clerical assistant.

4.1. Statistical methods for analyzing agreement

Inter-rater agreement is one type of reliability or reproducibility measure and can be evaluated using a variety of statistical tests. Although percent agreement is still widely utilised for assessment of IRR (e.g. Johnsson *et al.* 2004, Warming *et al.* 2004), it has been criticised on the grounds that it does not account for chance agreement. ICC has been used for quantitative variables but it is influenced by the distribution, in that it assumes a normal distribution of an underlying continuous variable. Since the HAL data in this study were not normally distributed, the ICC was not a suitable IRR measure for the HAL variable.

The kappa statistic has been widely used for categorical variables, but it has the disadvantage of being highly sensitive to the marginal distributions of the ratings (Feinstein and Cicchetti 1990). In this study, kappa had certain limitations for assessment of IRR, depending on how data were distributed in contingency tables. For example, where all of the MMH data were located in one cell of the contingency table, the kappa coefficient was undefined. There were three such items for which kappa statistics could not be computed and thus this study could only rely on percent agreement to evaluate IRR.

4.2. Sources of variability in agreement

The reliability of an observational technique may strongly depend on the experience level of the raters (Paquet *et al.* 2001, Voskuijl and van Sliedregt 2002), the number of exposure items to be recorded (van der Beek *et al.* 1992, Paquet *et al.* 2001), clear definition of the variables (Kilbom 1994a, Burt and Punnett 1999) and the nature of the work being observed, such as motion speed or the predictability of work activities. In general, these factors have been discussed anecdotally but rarely examined formally. The level of experience in exposure assessment, both within and between raters, is a factor that has been neglected in most exposure studies (Noyes 1994). Voskuijl and van Sliedregt (2002) reported that job information type (e.g. behaviour or work-oriented elements) and rater experience in job analysis, among other sources of variability, were highly significant moderators of IRR. The work nature, e.g. the amount of active or dynamic work, has been documented to affect the reliability of observations for postural assessment (Burdorf *et al.* 1992).

Based on iterative evaluations throughout the observer training period, revisions of the newly expanded PATH version were undertaken precisely to resolve sources of discordance.

In particular, agreement levels were markedly improved for each of the postures, among items, in the field dataset. The improvement in IRRs after the training period was reassuring in that potential sources of error, such as unclear posture definitions, were remedied by the training program and revisions in variable definitions.

The measures of IRR were compared with regard to job characteristics (slower vs. faster hand motions) and level of observer experience in ergonomic job analysis. Both IRR measures were higher in the low HAL jobs than in the medium for a majority of these items, as expected. Agreement was also better for the pairs of raters in which both had prior ergonomics experience.

Both the number of items to record per unit observation and the number of categories within items would be other potential sources of rater disagreement (van der Beek *et al.* 1992, Paquet *et al.* 2001). In this study, the IRR results were mixed across items with regard to the number of categories. For instance, the kappa statistics for leg posture, with five categories of postures observed, were 0.92 or higher, while those for elbow posture with three categories were 0.45 or less (Tables 3, 4, 5). Thus, these data did not support the hypothesis that fewer categories would produce higher agreement. However, discrete leg postures may be easier to distinguish than elbow angles in a single plane. More vs. fewer categories for the same exposure were not directly compared, so the hypothesis has also not been disproven.

The kappa coefficients were at least 'fair to good' for all items in the distal upper extremity templates but only five of nine items in the whole body template, showing that there were markedly different patterns of agreement levels across the templates. The higher reliability for the distal upper extremity exposures was likely attributed to the template design. The distal upper extremity templates were composed of simpler items (i.e. seven dichotomous and two other items) so that it might be easier to judge and record an observed event for those dichotomous variables. This appears to support that a direct observational instrument with fewer items produces less error in observation.

IRR reflects the amount of random and systematic error inherent in an observational method (Gardiner *et al.* 2002). If IRR among observers is low, the usefulness of the observations is severely limited (Fleiss 1981). Thus, in such a case, it would be desirable to search for and rectify sources of disagreement (Dunn 1989). Noise level had both low percent agreement and negative kappa (Tables 3, 4, 5), except in observations made by raters of group A (Table 5). The results and post-hoc investigation showed that the observer with no prior ergonomics experience rated the noise level differently in a systematic manner. Because negative kappa shows worse than chance-expected agreement, the noise level data collected by this rater will be dropped from future data analyses.

4.3. Comparison with other analyses of inter-rater reliability

A very broad range of inter-rater agreement levels was obtained in this study, from 42.5% to 100% of observations. Inter-rater agreement has previously been reported with other versions of the PATH method. Buchholz *et al.* (1996) examined IRR in analysis of construction work for several of the original PATH codes (e.g. body postures, activity and grasp type), which generally correspond to the whole body template items reported here.

Percent agreement ranged from 54 to 99 during observation of two workers in a pipe-laying operation. Pan *et al.* (1999) used a modification of the PATH method in retail store work, where the kappa coefficients were 0.50 to 0.63 for items such as body postures, weight handled and material-handling classification. These values improved markedly with further experience, with a range from 81 to 100% in four construction job tasks for postures (trunk, arms, legs), activities, tools used and load handled (Buchholz *et al.* 2003). (In a separate project, agreement between paper-and-pencil and PDA recoding was quite high.)

The IRR levels of this study were equivalent to or higher than those of other studies in which different observation methods were used. Burt and Punnett (1999) evaluated 18 postures in 70 jobs in an automotive manufacturing company, with percent agreement ranging from 26 to 99 and kappa coefficients from 0 to 0.55. In a study using a direct observational instrument for 45 nursing patient transfers in hospital wards (Johnsson *et al.* 2004), the percent agreement levels were 51 to 93 and the kappa coefficients were 0.16 to 0.77 for 16 items covering three phases of a transfer.

4.4. Study limitations

The IRR data were obtained from 10 jobs, which were identified in the earliest stage of the study. At that time, access had not yet been obtained to any high HAL jobs in the hospital. Given that agreement was inversely related to worker hand activity, agreement among observers would presumably be lower for jobs with even higher hand speed.

Confidence intervals were broad for some kappa coefficients, reflecting limited statistical power, especially in the comparisons by HAL level and observer experience.

External validity of the original PATH items, such as trunk, arm and leg postures, has previously been demonstrated (Buchholz *et al.* 1996, Paquet *et al.* 2001). Validity of the new upper extremity items should also be evaluated.

5. Conclusions

A new expanded version of the PATH method permits characterisation of exposure to upper extremity ergonomic risk factors for MSDs. The new items were developed with iterative refinements of item definitions and classification criteria; additional training also improved reliability between raters. Most of the new items could be observed with fair to good agreement between raters. For some items, IRR was higher both for the jobs with slower hand activity and for raters with more prior ergonomics experience. The results show that it is feasible to observe ergonomic risk factors for MSDs of the distal upper extremity in real time, at least in jobs that are not performed very rapidly. The revised instrument may be useful for ergonomists or clinicians who need to conduct an ergonomic intervention as well as assess exposure to risk factors for MSDs in hospital employees.

Acknowledgments

References

Baker NA, Cook JR, Redfern MS. 2008; Rater reliability and concurrent validity of the Keyboard Personal Computer Style instrument (K-PeCS). Applied Ergonomics. 40:136–144. [PubMed: 18275932]

Bernard, BP, editor. Musculoskeletal disorders and workplace factors: A critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back. Cincinnati, OH: National Institute of Occupational Safety and Health; 1997. Publication no. 97–141

Bru E, Mykletun RJ, Svebak S. 1994; Assessment of musculoskeletal and other health complaints in female hospital staff. Applied Ergonomics. 25:101–105. [PubMed: 15676956]

Buchholz B, et al. 1996; PATH: A work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work. Applied Ergonomics. 27:177–187. [PubMed: 15677058]

Buchholz B, et al. 2003; Quantification of ergonomic hazards for ironworkers performing concrete reinforcement tasks during heavy highway construction. AIHA Journal. 64:243–250. [PubMed: 12688849]

Burdorf A, et al. 1992; Measurement of trunk bending during by direct observation and continuous measurement. Applied Ergonomics. 23:263–267. [PubMed: 15676874]

Burt S, Punnett L. 1999; Evaluation of interrater reliability for posture observations in a field study. Applied Ergonomics. 30:121–135. [PubMed: 10098805]

Camerino D, et al. 2001; Job strain and musculoskeletal disorders of Italian nurses. Occupational Ergonomics. 2:215–223.

Cann AP, et al. 2008; Inter-rater reliability of output measures for a posture matching assessment approach: a pilot study with food service workers. Ergonomics. 51(4):556–572. [PubMed: 18357541]

Cicchetti DV. 1981; Testing the normal approximation and minimal sample size requirements of weighted kappa when the number of categories is large. Applied Psychological Measurement. 5:101–104.

Cifuentes M, et al. 2008; Job strain predicts survey response in healthcare industry workers. American Journal of Industrial Medicine. 51(4):281–289. [PubMed: 18247321]

Cohen J. 1960; A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 20:37–46.

Colombini D. 1998; An observational method for classifying exposure to repetitive movements of the upper limbs. Ergonomics. 41:1261–1289. [PubMed: 9754031]

David G, et al. 2008; The development of the Quick Exposure Check (QEC) for assessing exposure to risk factors for work-related musculoskeletal disorders. Applied Ergonomics. 39:57–69. [PubMed: 17512492]

d'Errico A, et al. 2007; Hospital injury rates in relation to socioeconomic status and working conditions. Occupational and Environmental Medicine. 64(5):325–333. [PubMed: 17182643]

Dunn, G. Design and analysis of reliability studies: The statistical evaluation of measurement errors. New York: Oxford University Press; 1989.

Dybel, G. Thesis (Doctor of Science). University of Massachusetts Lowell; 2000. Ergonomic evaluation of work as a home health care aide: Descriptive and epidemiological analysis.

Elford W, Straker L, Strauss G. 2000; Patient handling with and without slings: an analysis of the risk of injury to the lumbar spine. Applied Ergonomics. 31:185–200. [PubMed: 10711981]

Feinstein AR, Cicchetti DV. 1990; High agreement but low kappa: I. the problems of two paradoxes. Journal of Clinical Epidemiology. 43:543–549. [PubMed: 2348207]

Fleiss, JL. Statistical methods for rates and proportions. 2. New York: John Wiley & Sons, Inc; 1981.

Fleiss JL, Cohen J. 1973; The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and Psychological Measurement. 33:613–619.

Fransson-Hall C, et al. 1995; A portable ergonomic observation method (PEO) for computerized on-line recording of postures and manual handling. Applied Ergonomics. 26:93–100. [PubMed: 15677005]

Fulmer S, et al. 2002; Ergonomic exposure in apple harvesting: Preliminary observations. American Journal of Industrial Medicine. 42:3–9.

Fuortes LJ, et al. 1994; Epidemiology of back injury in university hospital nurses from review of workers' compensation records and a casecontrol survey. Journal of Occupational Medicine. 36:1022–1026. [PubMed: 7823214]

Gardiner MD, Faux S, Jones LE. 2002; Interobserver reliability of clinical outcome measures in a lower limb amputee population. Disability and Rehabilitation. 24:219–225. [PubMed: 11926262]

Gillen M, et al. 2007; The association of socioeconomic status and psychosocial and physical workplace factors with musculoskeletal injury in hospital workers. American Journal of Industrial Medicine. 50:245–260. [PubMed: 17311255]

Gold JE, Park JS, Punnett L. 2006; Work routinization and implications for ergonomic exposure assessment. Ergonomics. 49:12–27. [PubMed: 16393801]

Hernandez L, et al. 1998; A study of musculoskeletal strain experienced by nurses. Occupational Ergonomics. 1:123–133.

Hignett S, McAtamney L. 2000; Rapid Entire Body Assessment (REBA). Applied Ergonomics. 31:201–205. [PubMed: 10711982]

Hignett S, Richardson B. 1995; Manual handling human loads in a hospital: An exploratory study to identify nurses' perceptions. Applied Ergonomics. 26:221–226. [PubMed: 15677021]

Howard, NL. Thesis (Master of Science). University of Massachusetts Lowell; 1997. The development of exposure assessment models for ergonomic stressors to the hip and knee in dairy farming.

Janowitz IL, et al. 2006; Measuring the physical demands of work in hospital settings: Design and implementation of an ergonomics assessment. Applied Ergonomics. 37:641–658. [PubMed: 16226213]

Johnsson C, et al. 2004; A direct observation instrument for assessment of nurses' patient transfer technique (DINO). Applied Ergonomics. 35:591–601. [PubMed: 15374767]

Karhu O, Kansi P, Kuorinka I. 1977; Correcting working postures in industry: A practical method for analysis. Applied Ergonomics. 8:199–201. [PubMed: 15677243]

Kemmlert K. 1995; A method assigned for the identification of ergonomic hazards – PLIBEL. Applied Ergonomics. 26(3):199–211. [PubMed: 15677019]

Keyserling WM. 1986; Postural analysis of the trunk and shoulders in simulated real time. Ergonomics. 29:569–583. [PubMed: 3709510]

Kilbom A. 1994a; Assessment of physical exposure in relation to work-related musculoskeletal disorders – what information can be obtained from systematic observations? Scandinavian Journal of Work Environment and Health. 20:30–45.

Kilbom A. 1994b; Repetitive work of the upper extremity: part II-the scientific basis (knowledge base) for the guide. International Journal of Industrial Ergonomics. 14:59–86.

Lagerström M, Hansson T, Hagberg M. 1998; Work-related low-back problems in nursing. Scandinavian Journal of Work Environment and Health. 24:449–464.

Latko WA, et al. 1997; Development and evaluation of an observational method for assessing repetition in hand tasks. AIHA Journal. 58:278–285. [PubMed: 9115085]

McAtamney L, Corlett EN. 1993; RULA: a survey method for the investigation of work-related upper limb disorders. Applied Ergonomics. 24(2):91–99. [PubMed: 15676903]

Messing K, Chatigny C, Courville J. 1998; 'Light' and 'heavy' work in the housekeeping service of a hospital. Applied Ergonomics. 29:451–459. [PubMed: 9796791]

National Research Council and Institute of Medicine. Musculoskeletal disorders and the workplace: Low back and upper extremities. Washington DC: National Academy Press; 2001.

Noyes B. 1994; Inter-rater reliability: Regaining credibility with your staff and financial officer while meeting JCHO standards. Journal of Nursing Administration. 24:7–8.

Owen BD, Keene K, Olson S. 2002; An ergonomic approach to reducing back/shoulder stress in hospital nursing personnel: A five year follow up. International Journal of Nursing Studies. 39:295–302. [PubMed: 11864653]

Pan CS, et al. 1999; Ergonomic exposure assessment: An application of the PATH systematic observation method to retail workers. International Journal of Occupational and Environmental Health. 5:79–87. [PubMed: 10330506]

Paquet V, Punnett L, Buchholz B. 2001; Validity of fixed-interval observations for postural assessment in construction work. Applied Ergonomics. 32:215–224. [PubMed: 11394462]

Park, J-K. Thesis (Doctor of Science). University of Massachusetts Lowell; 2006. Exposure assessment and musculoskeletal disorder risk factors in hospital laboratories.

Park, J-S. Thesis (Doctor of Science). University of Massachusetts Lowell; 2000. Ergonomic exposure assessment and musculoskeletal disorders in automobile manufacturing.

Rockefeller, KA. Thesis (Doctor of Science). University of Massachusetts Lowell; 2002. Evaluation of an ergonomic intervention in Washington State nursing homes.

SAS Institute Inc. Statistical analysis software, version 9.1 for Windows. Cary, NC: SAS Institute Inc; 2003.

Stetson DS, et al. 1991; Observational analysis of the hand and wrist: a pilot study. Applied Occupational and Environmental Hygiene. 6:927–937.

van der Beek A, van Gaalen LC, Frings-Dresen HW. 1992; Working postures and activities of lorry drivers: A reliability study of on-site observation and recording on a pocket computer. Applied Ergonomics. 23:331–336. [PubMed: 15676879]

Voskuijl OF, van Sliedregt T. 2002; Determinants of interrater reliability of job analysis: A meta-analysis. European Journal of Psychological Assessment. 18:52–62.

Warming S, et al. 2004; An observation instrument for the description and evaluation of patient transfer technique. Applied Ergonomics. 35:603–614. [PubMed: 15374768]

Wiktorin C, et al. 1995; HARBO, a simple computer-aided observation method for recording work postures. Scandinavian Journal of Work Environment and Health. 21:440–449.

Table 1

Item and category by template in the revised posture, activity, tools and handling instrument.

Item	Category*
Whole body template	
1. Trunk posture	1 = Neutral <20°
	2 = Moderate flexion 20°–<45°
	3 = Severe flexion 45°
	4 = Lateral bent/twist flexed
	5 = Lateral bent/twist neutral
2. Leg posture	1 = Stand (flexion <35°)
	2 = Walking/running
	3 = Sitting
	4 = Kneeling (one or both knees)
	5 = Squat (both knees 80°)
	6 = Lunge (one knee 35°)
	7 = Crawl
	8 = Stand on one foot
3. Shoulder/arm elevation	1 = Both arms <60°
	2 = one arm 60°
	3 = two arms 60°
4. Elbow posture	1 = Neutral (30°–150°)
	2 = Extension (>150°)
	3 = Extreme flexion (<30°)
5. Weight in hands	1 = <10 lbs
	2 = 10–<50 lbs
	3 = 50–<150 lbs
	4 = 150 lbs
6. Manual materials handling (MMH)	1 = No MMH
	2 = one hand
	3 = two hands
7. MMH action 1 = No MMH action	2 = Carry/hold
	3 = Push/pull/drag lift
	4 = Lower
8. Noise level	1 = Person nearby can be heard in a normal voice
	2 = Person nearby must raise the voice to be heard
	3 = Person nearby must shout to be heard
9. Vibration	1 = None
	2 = Segmental
	3 = Whole-body
Hand/forearm template	
10. Neutral/gross grasp	1 = No
	2 = Yes

Item	Category[*]
11. Wrist/forearm deviation	1 = No
	2 = Yes
12. Pinch grip	1 = No
	2 = Yes
13. Keyboarding	1 = No
	2 = Yes
14. Hand/forearm contact stress	1 = No
	2 = Yes
15. Hand/forearm vibration	1 = No
	2 = Yes
16. Weight in hands	1 = <10 lbs
	2 = 10–<50 lbs
	3 = 50–<150 lbs
	4 = 150 lbs
17. Hand observed	1 = Right
	2 = Left
Hand activity template	
18. Hand activity level	Eleven categories 0 to 10 with verbal anchors:
	0 = hands idle most of the time; no regular exertions
	2 = consistent, conspicuous, long pauses
	4 = slow steady motion; frequent pauses
	6 = steady motion; infrequent pauses
	8 = rapid steady motion; no regular pauses
	10 = rapid steady motion; difficulty keeping up

[*] The category of 'Not observed/not sure' in each item is omitted.

Table 2

Job title and rater pair in posture, activity, tools and handling observations.

Job title	Rater pair	Number of observation pairs	Subject
Cat scan supervisor	BC	22	S1
Receptionist	BC, BD	86	S2, S3, S4
Radiology technician I	CD	27	S5
Radiology technician II	AB, AC, BC	35	S6
Human resources assistant	AD	39	S7
Nuclear medical supervisor	BD	60	S8
Nuclear medical technician	BC	79	S9
Ultrasound supervisor	BD	26	S10
Ultrasound clinical coordinator	BD	54	S11
Benefits specialist	AB	15	S12
Total		443	

Table 3

Inter-rater reliability for each item in posture, activity, tools and handling observations.

| | Inter-rater reliability (n = 443 observation pairs) | | | |
| | | Kappa | | |
Item	% agreement	Coefficient*	95% CI	Classification
1. Trunk posture	74.5	0.53	0.46–0.60	Fair to good
2. Leg posture	96.3	0.94	0.91–0.97	Excellent
3. Shoulder/arm elevation †	87.8	0.66	0.55–0.76	Fair to good
4. Elbow posture †	79.7	0.40	0.27–0.53	Fair to good
5. Weight in hands †	99.3	0.40	(–)0.15–0.94	Fair to good
6. Manual materials handling (MMH)	100	_§	–	–
7. MMH action	98.8	0.37	0.01–0.73	Poor
8. Noise level †	42.5	(–)0.19	(–)0.26–(–)0.11	Poor
9. Vibration	99.5	0	–	Poor
10. Neutral/gross grasp	72.2	0.44	0.36–0.53	Fair to good
11. Wrist/forearm deviation	74.8	0.50	0.41–0.58	Fair to good
12. Pinch grip	87.3	0.72	0.65–0.79	Fair to good
13. Keyboarding	95.9	0.83	0.75–0.91	Excellent
14. Hand/forearm contact stress	87.2	0.60	0.51–0.70	Fair to good
15. Hand/forearm vibration	100	1.0	–	Excellent
16. Weight in hands †	98.8	0.44	0.01–0.88	Fair to good
17. Hand observed	100	1.0	–	Excellent
18. Hand activity level †	75.3	0.65	0.58–0.71	Fair to good

Note: 95% CI and significance are shown for kappa.

*
Significance: $p < 0.0001$ for each kappa coefficient shown, except for vibration (no. 9, $p > 0.99$).

†
Ordinal variable for which the weighted kappa coefficient was calculated.

§
Undefined (zero in the denominator).

Table 4

Comparison of inter-rater reliability between jobs with low and medium hand activity levels (HAL).

Item	Low HAL job (n = 259 observation pairs)			Medium HAL job (n = 184 observation pairs)		
	% agreement	Kappa		% agreement	Kappa	
		Coeff.*	95% CI		Coeff.*	95% CI
1. Trunk posture	77.0	0.54	0.45–0.64	71.0	0.51	0.4–0.61
2. Leg posture	94.8	0.92	0.88–0.96	98.3	0.97	0.93–1.0
3. Shoulder/arm elevation†	90.0	0.69	0.55–0.83	84.7	0.62	0.46–0.77
4. Elbow posture†	79.5	0.39	0.19–0.58	79.9	0.41	0.22–0.59
5. Weight in hands†	100	1.0	–	98.3	(–)0.01	–
6. Manual materials handling (MMH)	100	–§	–	100	–§	–
7. MMH action	99.6	0.75	0.40–1.0	97.7	(–)0.01	–
8. Noise level†	32.5	(–)0.45	(–)0.53–(–)0.37	56.4	0.17	0.07–0.27
9. Vibration	99.6	0	–	99.4	0	–
10. Neutral/gross grasp	75.9	0.50	0.39–0.61	66.7	0.34	0.21–0.47
11. Wrist/forearm deviation	76.6	0.52	0.41–0.63	72.1	0.42	0.29–0.56
12. Pinch grip	86.8	0.70	0.61–0.80	88.0	0.74	0.64–0.85
13. Keyboarding	97.9	0.75	0.54–0.96	92.9	0.82	0.72–0.92
14. Hand/forearm contact stress	91.7	0.67	0.53–0.80	80.6	0.53	0.40–0.67
15. Hand/forearm vibration	100	1.0	–	100	–§	–
16. Weight in hands†	99.6	0.80	0.798–0.801	97.6	(–)0.01	–
17. Hand observed	100	1.0	–	100	1.0	–
18. Hand activity level†	75.2	0.58	0.48–0.69	75.4	0.63	0.54–0.73

Note: 95% CI and significance are shown for kappa.

*
$p < 0.0001$ for each kappa coefficient shown, except for noise level ($p = 0.0015$) in the medium HAL job and others.

($p > 0.88$ each) such as vibration (no. 9) in the low HAL job, weight in hands (no. 5 and no. 6), MMH action and vibration (no. 9) in the medium HAL job.

†
Ordinal variable for which the weighted kappa coefficient was calculated.

§
Undefined (zero in the denominator).

Table 5

Comparison of inter-rater reliability between rater pairs with experienced raters in each pair (group A) and with one inexperienced rater in each pair (group B).

Item	Group A (n = 185 observation pairs)			Group B (n = 258 observation pairs)		
	% agreement	Kappa Coeff.*	95% CI	% agreement	Kappa Coeff.*	95% CI
1. Trunk posture	82.3	0.68	0.58–0.77	68.8	0.42	0.33–0.52
2. Leg posture	97.8	0.96	0.93–1.0	95.2	0.92	0.88–0.96
3. Shoulder/arm elevation†	90.0	0.75	0.61–0.88	86.2	0.56	0.41–0.71
4. Elbow posture†	77.7	0.33	0.14–0.53	81.2	0.45	0.27–0.62
5. Weight in hands†	99.4	0	–	99.2	0.50	(–)0.1–1.0
6. Manual materials handling (MMH)	100	–§	–	100	–§	–
7. MMH action	98.9	0	–	98.8	0.50	0.1–0.90
8. Noise level†	96.7	0.86	0.74–0.97	3.2	(–)0.59	(–)0.73–(–)0.46
9. Vibration	100	–§	–	99.2	0	–
10. Neutral/gross grasp	74.4	0.49	0.36–0.62	70.5	0.40	0.28–0.52
11. Wrist/forearm deviation	79.1	0.59	0.47–0.70	71.6	0.43	0.32–0.55
12. Pinch grip	93.7	0.87	0.79–0.94	82.5	0.60	0.49–0.71
13. Keyboarding	98.3	0.93	0.85–1.0	94.0	0.74	0.61–0.87
14. Hand/forearm contact stress	88.4	0.59	0.43–0.75	86.3	0.61	0.49–0.73
15. Hand/forearm vibration	100	–§	–	100	1.0	–
16. Weight in hands†	98.3	0	–	99.2	0.67	0.36–0.97
17. Hand observed	100	1.0	–	100	1.0	–
18. Hand activity level†	84.2	0.79	0.71–0.86	68.9	0.51	0.41–0.60

Note: 95% CI and significance are shown for kappa.

*
Significance: $p < 0.0001$ for each kappa coefficient shown, except for items ($p > 0.99$ each) such as weight in hands (no. 5 and no. 16), MMH action in group A and vibration (no. 9) in group B.

†
Ordinal variable for which the weighted kappa coefficient was calculated.

§
Undefined (zero in the denominator).