Ergonomics
Vol. 51, No. 8, August 2008, 1219–1237

The Cochran–Weiss–Shanteau performance index as an indicator of upper limb risk assessment expertise

Claire A. Williams[a]*, Roger A. Haslam[b] and David J. Weiss[c]

[a]*COPE Ltd., Technology Drive, Beeston, Nottinghamshire, NG9 2ND, UK;* [b]*Department of Human Sciences, Loughborough University, Leicestershire, LE11 3TU, UK;* [c]*Department of Psychology, California State University, Los Angeles, CA 90032, USA*

Ergonomists and many other professionals apply ergonomics principles to musculoskeletal health problems. This study examines whether there are differences when it comes to judgement expertise concerning upper limb disorders (ULDs) between ergonomists and those with less ergonomics training. The Cochran–Weiss–Shanteau (CWS) performance index combines judgement consistency with discrimination into one CWS index. Fifty-eight professionals working in the musculoskeletal health area, from four different professions, judged the likelihood of staff complaining of ULDs in a number of written work scenarios containing ULD risk factors. A student group (n = 148) taking an introductory ergonomics module was used as a reference. The ergonomists scored higher on the CWS index than all of the other groups, performing significantly better than all but the occupational health advisors. Performance improved with increased training level but not with experience. This study suggests that ergonomists are quantifiably different from other ergonomics advisors in their judgement performance in this context. Given the global cost of musculoskeletal disorders, assessing the expertise of those giving ergonomics advice for the management of musculoskeletal health is of great significance. This study presents a method for assessing judgement performance in ULD risk assessment, an important part of musculoskeletal health management.

Keywords: musculoskeletal disorders; competence; professional practice; risk assessment

1. Introduction

The importance of musculoskeletal disorders (MSDs) in terms of their impact on the working population is well established. According to the Health and Safety Executive (HSE), MSDs are the most common work-related illness in Great Britain, affecting 1.1 million people per year. An estimated 11.8 million working days per year are lost to work-related MSDs in Great Britain, with each sufferer having taken around 19.4 days off in that 12 month period (Health and Safety Executive 2005). In the USA in 2005, MSDs accounted for 375,540 cases, or 30% of the injuries and illnesses with days away from work (Bureau of Labour Statistics 2006), whilst across Europe as a whole MSDs are the most common work-related problem, with almost 24% of the EU's workers reporting backache and 22% reporting muscular pains (European Foundation for the Improvement of Living and Working Conditions 2007).

*Corresponding author. Email: claire.williams@copeohs.com

ISSN 0014-0139 print/ISSN 1366-5847 online

DOI: 10.1080/00140130802087094
http://www.informaworld.com

It is widely accepted that ergonomics has a prominent part to play in the prevention of MSDs (Buckle 2005, David 2005, Dempsey 2007). Ergonomists and other professionals alike are working in what has been described by Westgaard and Winkel (1996) as School-III of ergonomics, namely, the arena of musculoskeletal health. In doing so, these professionals are applying ergonomics principles in an effort to prevent MSDs from occurring and to ameliorate situations when they do.

In this paper, non-ergonomist professionals working in the field of musculoskeletal health who apply ergonomics principles are termed 'ergonomics advisors'. In the UK, ergonomics advisors might include health and safety advisors (HSAs), occupational health advisors (OHAs) and physicians, physiotherapists, occupational therapists and specialist furniture suppliers. In the USA, they might include industrial engineers, occupational medicine nurses and physicians, industrial hygienists, safety engineers and physical therapists. The purpose of this study was to examine the judgement performance of some of these professional groups in one aspect of MSD management, namely, upper limb disorder (ULD) risk assessment.

1.1. *Expertise and ergonomics*

Expertise in others, particularly in decision making, is a subject of interest to ergonomists as part of their practice (Farrington-Darby and Wilson 2006, Piegorsch *et al.* 2006). However, expertise in ergonomists themselves is also of importance for the profession, in spite of the paucity of studies in this area (Piegorsch *et al.* 2006). Some of the few studies that have examined expertise amongst ergonomics practitioners have looked at expert–novice differences when undertaking different types of ergonomics work.

Weston and Haslam (1992) investigated the abilities of those with and without formal ergonomics training to recognise and apply ergonomics principles in design and evaluation activities. They found that ergonomics performance significantly improved with ergonomics training when those with and without training were compared with ergonomics 'experts' on two different tasks. In effect, this study compared novices, experts and what others have termed 'advanced beginners' (those with some experience/training in a field) (Dreyfus *et al.* 1986) and found performance to improve with training level.

Jones *et al.* (1999) similarly compared trained and untrained non-ergonomists with ergonomists for a materials handling risk assessment task. They found that both non-ergonomist groups were able to identify hazards, though they were less able to prioritise tasks for assessment. There was a trend indicating that the trained group carried out 'better' assessments than the untrained group (both qualitatively and quantitatively for hazards identified and solutions proposed), though the difference was not statistically significant. Both the non-ergonomist groups proposed fewer solutions than the ergonomist.

A later study by Winnemuller *et al.* (2004) compared an ergonomist's assessments of work-related musculoskeletal risk factors with those of supervisors and workers. In assessing the absence or presence of risk factors, the supervisors agreed with the ergonomist 81% of the time and the workers agreed 77% of the time. Where there was disagreement, the non-ergonomists overestimated the risks.

Stanton and Young (2003) examined the reliability and validity of different ergonomics methods when employed by novices. They reported that novices could reliably apply certain ergonomics techniques with validity; these techniques included checklists and questionnaires. The more structured technique, keystroke level model (Card *et al.* 1983), proved the most valid and reliable tool in the hands of novices when carrying out product evaluation and usability tasks. Previous work by the same research team (Stanton and Stevenage 1998) had demonstrated that performance by novices and experts on ergonomics evaluation tasks was linked with the complexity of the device being evaluated; when evaluating a simpler device than that in the 2003 study, the 1998 novices performed better than the 2003 novices, while experts, on the other hand, performed better than the 2003 novices, even when evaluating a more complex device (Baber and Stanton 1996). Clearly, the complexity of the task being undertaken matters when examining the performance of experts and novices applying ergonomics principles.

Other work has analysed the objectivity of ergonomics experts' decisions and the consensus amongst them. Keyserling and Wittig (1988) studied five ergonomics experts (university researchers) and compared their assessment decisions regarding ergonomics stressors and also compared their decisions with objective tools such as the National Institute for Occupational Safety and Health (NIOSH) *Work Practices Guide for Manual Lifting* (National Institute for Occupational Safety and Health 1981). They found that there was good consensus between their experts and with the NIOSH guide. Where there was disagreement, the experts rated the stressors higher than the objective tool.

Further decision-making work examined the differences between ergonomists from different backgrounds (Piegorsch *et al.* 2006). These authors identified the decision-making processes used by ergonomists from industrial engineering and physical therapy backgrounds when generating recommendations to prevent and control low back pain. Their aim was to identify the schemata (the abstract form in which people organise and store information from previous experiences) that both kinds of ergonomists used and incorporate these into a conceptual framework if they were sufficiently similar. They found that the background of the ergonomists had less influence on decision making than individual personality traits and the constraints of the practice environment. Their framework model adequately described the decision making of ergonomists with backgrounds in either discipline.

So, previous work has examined a variety of different ergonomics activities in order to look for differences between experts and novices and to examine the behaviours of experts with different backgrounds. With this previous work in mind, the ergonomics activity of interest for the present study is discussed below.

1.2. Upper limb disorder risk assessment

ULD risk assessment forms part of the tasks that ergonomists and ergonomics advisors commonly undertake. Identifying jobs with risk factors linked with the development of ULDs as well as other MSDs is relatively commonplace (Keyserling and Wittig 1988, Keyserling *et al.* 1993, Dempsey *et al.* 2005, Piegorsch *et al.* 2006).

In brief, the process involves having knowledge of the potential risk factors, observing work tasks and workplaces for evidence of these risk factors, quantifying them and finally making a judgement on the likely outcome. This process is known as risk assessment and is aided by a relatively strong evidence base for the risk factors associated with ULDs as well as by a number of ULD risk assessment tools.

The literature has established associations between the development of ULDs and certain risk factors via reporting schemes (e.g. Cherry *et al.* 2001), epidemiological reviews (Bernard 1997) and population surveys (Jones 1998, Health and Safety Executive 2005, Bureau of Labour Statistics 2006, European Foundation for the Improvement of Living and Working Conditions 2007). These known risks have then been incorporated into risk assessment tools such as rapid upper limb assessment (RULA) (McAtamney and Corlett

1993), occupational repetitive actions index (OCRA) (Colombini 1998, Colombini *et al.* 2002, Occhipinti and Colombini 2007) and quick exposure checklist (QEC) (Li and Buckle 1999) as well as into governmental guidance such as the Washington State Ergonomics Rule (WAC) (Washington State Department of Labor and Industries 2000) and *Upper Limb Disorders in the Workplace* (HSG60) (Health and Safety Executive 2002).

This last set of risk assessment guidance, HSG60 (as described in Graves *et al.* 2004), incorporates the evidence from many of the other sources as a statement of best practice in ULD risk management for non-specialists. These sources include WAC (Washington State Department of Labor and Industries 2000), the OSHA screening tool (Occupational Safety and Health Administration 2000), QEC (Li and Buckle 1999), the 'Upper Extremity Checklist' (Keyserling *et al.* 1993), RULA (McAtamney and Corlett 1993), OCRA (Colombini 1998) and the 'TUC Guide to Assessing WRULDs Risks' (Buckle and Hoffman 1994).

Seven risk areas are outlined in HSG60 as being evidentially linked with the development of ULDs. These are repetition, force, duration of exposure, awkward posture, psychosocial factors, individual differences (issues such as age, experience, etc.) and working environment (issues such as temperature, lighting and vibration). These areas form the basis of what any ergonomics advisor or ergonomist would investigate when assessing a workplace for ULD risk and HSG60 has the advantage for this study of being targeted at non-specialists.

1.3. The Cochran–Weiss–Shanteau performance index

In the extensive literature on expertise outside the ergonomics discipline, the ability to evaluate domain-specific situations correctly has been held to be an important indicator of expertise. An expert should identify and discriminate between different 'stimuli' with which they are faced (diagnosis) and do so in a consistent way (Weiss and Shanteau 2003). They should also be consistent in the predictions (prognosis) they make from their diagnoses. Customarily, empirical measures of this ability compare responses made by those being assessed with those of established experts. In essence, this approach presumes that the judgements of the established experts are correct. In turn, the established experts are certified because their judgements matched those of a previous generation of experts. Weiss and Shanteau (2003) proposed a way to break through the circularity in this reasoning by measuring two necessary, observable properties of expert judgement. Just as these properties are necessary in a mechanical measuring instrument, they ought to be exhibited within the judgements of a candidate expert. An expert should respond differently to different stimuli and should respond similarly to similar stimuli. Weiss and Shanteau (2003) labelled these properties discrimination and consistency respectively. Furthermore, they proposed a performance index that combines the properties into a ratio called the Cochran–Weiss–Shanteau (CWS) performance index. The consistency property is 'reverse-scored' (i.e. as inconsistency) because it serves as the denominator of the ratio. The motivation for combining the properties is that someone without true ability can adopt simple strategies for maximising one property at the expense of the other (for example, vary responses widely or hardly at all), but only someone who is an expert within the domain can achieve both simultaneously. A key feature of the CWS index is that the analyst can assess the degree of exhibited judgemental expertise without presuming to know the true values of the objects being judged.

In practice, the variance among responses to different stimuli is used as the estimate of discrimination and the variance among responses to the same stimulus is used as the measure of inconsistency. All that is needed to fuel these computations is an individual's dataset containing repeated responses to a variety of stimuli. The true values of the stimuli do not matter. Variances, with their heavy weighting of large discrepancies, have traditionally been used by statisticians to capture precision of measurement (Grubbs 1973), with a ratio form as the usual arrangement for comparison.

The CWS performance index is defined as:

$$\frac{\text{Discrimination}}{\text{Inconsistency}}$$

Where discrimination is high and inconsistency is low (prerequisites for expertise though not sufficient for it as a judge may be both discriminating and consistent but wrong), the CWS index will be high. Conversely, where discrimination is low or inconsistency is high, the CWS will be low. While high CWS is not sufficient to guarantee expertise, it is necessary. No one whose CWS is low can be said to be judging expertly. The *caveat* is that because validity is not presumed, there is no certain way to ensure that the judge is attending to the correct aspects of the stimuli. Of course, no assessment approach that is unwilling to make the strong assumption that a gold standard of truth is available can do any better. A major advantage of this index is that it is objective, in the sense that it can be calculated from the judgements themselves without requiring an expression of opinion. Thus, CWS obviates the need to use the decisions of an 'expert' as an unproven gold standard against which to compare the decisions of the other study participants (for example, as used in the studies outlined in section 1.1.).

As discussed above, one of the more common tasks that ergonomics advisors undertake is risk assessment. To do this, advisors examine the workplace for evidence of specific hazards (diagnosis) and make a judgement based on that evidence about the likely outcome (prognosis). In this way, they undertake similar activities to clinicians, auditors and other judges. Indeed, the CWS index has been successfully used in a number of different situations including: to discern between expert and novice financial auditors and personnel selectors (Shanteau *et al.* 2002); to demonstrate the improvement in performance, with training, of occupational therapists (Weiss *et al.* 2006); and to highlight the differences in discernment and consistency between medical doctors diagnosing heart disease (Weiss and Shanteau 2003). It was proposed, therefore, that the CWS index might be an appropriate, objective method with which to assess ergonomics judgements.

Some of the other attributes linked with expertise in addition to high levels of discernment and consistency are lengthier experience (James 2007), higher certification (Shanteau *et al.* 2002), superior knowledge (Shanteau *et al.* 2002, James 2007) and certain behavioural characteristics such as greater confidence (Shanteau 1988, Abdolmohammadi and Shanteau 1992, Shanteau *et al.* 2002).

1.4. Study aims

This study examined how discriminating and consistent different ergonomics advisors are in their 'diagnoses' and 'prognoses' when it comes to ULD risk assessment. In addition, the training level and duration of experience of the participants were also recorded, so as to look at these potential predictors of expertise alongside the CWS method. Behavioural characteristics were not examined.

At the outset, there were three specific aims for this work:

(1) To identify which individuals/groups giving ergonomics advice are the most 'expert' judges in terms of their consistency and discrimination.
(2) To identify which characteristics of the participants were associated with better judgement performance.
(3) To assess whether CWS is a useful method for evaluating judgement expertise in the context of ergonomics.

2. Methods

2.1. Judgement scenarios

In order to employ the CWS index, repeated, measurable judgements needed to be gathered from participants in an area of ergonomics with which they were familiar. Evaluating different work environments and making judgements about the likelihood of staff complaining of ULDs was the chosen judgement task, and a method of presenting situations about which judgements could be made was required. Other studies examining clinical reasoning by physical therapists concluded that work carried out 'live' in the clinical environment (i.e. 'on the shop floor' in this context) was too variable to allow for cross subject comparison (James 2001, 2007). A more controlled method of presenting judgement situations was therefore required.

The situations needed to be presented in a manner that would have face validity to the participants and could be administered conveniently, without recourse to complicated technology. They also needed to be sufficiently difficult to allow the experts to demonstrate their expertise, whilst being within the understanding of all of the participants. In other words, both 'floor' and 'ceiling' effects should be avoided (Farrington-Darby and Wilson 2006).

With all these issues in mind, short written scenarios were produced about which judgements could be made, describing real environments encountered during the professional experience of one of the authors. This approach of providing controlled scenarios, rather than following individuals as they worked 'live', had been endorsed previously by the creators of the CWS index as 'reproducible success in controlled settings predicts success in real-world applications' (Weiss and Shanteau 2003). Written scenarios were chosen over 'video' clips because of their relative ease in both generation and administration. The written scenarios allowed participants to review the information as much as was required in order to make their judgements, whereas anything missed when first viewing a video clip would be lost. Furthermore, the psychosocial risk factor would have been difficult to represent in video form.

2.2. Risk factors

The scenarios needed to have different 'risk' stimuli embedded in them which were evidence based. To this end, the HSE's HSG60 document (Health and Safety Executive 2002) was used as a starting point for defining the stimuli, as it outlines both the risk factors and, where known, the particular levels of the risk factors that should be cause for concern. It is also useful in including psychosocial risk factors, which are implicated in the reporting of ULDs (Graves *et al.* 2004).

For the purpose of this study, the number of risk factors included in the scenarios was intentionally limited in order to restrict the number of scenarios generated. The reason for this was that all the risk factors used were to be incorporated into scenarios in all of the different combination options of 'present' or 'absent'. Therefore, if there were n risk factors, there would need to be 2^n scenarios, as there were two 'states' for each risk factor ('present' or 'absent'). As a consequence, it was decided to limit the number of risk factors to five, as 32 scenarios (2^5) would be a practicable number to generate, whereas 64 (2^6) or more was not.

To this end, five of the seven risk areas were selected from those outlined in HSG60. Those selected were repetition, force, duration of exposure, awkward posture and psychosocial factors. Effectively, this only omitted 'individual differences' (issues such as age, experience, etc.) and 'working environment' (issues such as temperature, lighting and vibration).

2.3. *Incorporating the risk factors in the scenarios*

The risk factors chosen (or 'cues') had to be represented in the scenarios at a level that was generally accepted as 'present' or 'absent'. Therefore, the guidance outlined in HSG60, as well as the accompanying evidence (Graves *et al*. 2004), was used to determine these levels wherever possible. An example of a scenario is shown in Figure 1. The following sections outline each of the risk factors and explain how they were incorporated in the scenarios.

2.3.1. Repetition

HSG60 describes repetition with the following definers:

- 'the same movements' are 'repeated every few seconds';
- 'a cycle or sequence repeated twice per minute'; or
- 'more than 50% of the task involves performing a repetitive sequence of motions'.

2.3.2. Force

HSG60 descriptors for force are:

- 'For the hand/wrist, high force tasks are those with estimated average individual hand force requirements of 4 kg or above'.
- 'Pinching an unsupported object weighing 0.9 kg (2 lbs) or more per hand, or using a similar pinching force (e.g. holding a small binder clip open)'.

2.3.3. Awkward postures

Awkward postures are described in HSG60 for different parts of the upper limb. The description of the postures is often in combination with other risk factors such as repetition or long duration. In the scenarios generated for this study, awkward postures were described independently of the duration of exposure or repetition, which were present or absent in their own right. All body parts were considered together such that awkward posture would be described for either the hands/wrists, arms/shoulders or head/neck. The following outlines the awkward postures from HSG60 that have the potential to cause harm:

- For the wrist, deviations from neutral either 'up or down' or 'to either side' are highlighted as problematic where they are held or occur repetitively, where 'the greater the deviation from a neutral position, the greater the risk'.

Example

In order to place the car windscreens into the storage racks, operators have to adopt a posture with arms fully outstretched, to reach the edges of the glass. Each windscreen weighs upwards of 5.5 kg and is placed into racks at about waist height.

The nature of the task involves lone-working for the 30 min that staff work in the racks during their shift. They have to replace 120 windscreens in the session, otherwise they receive reduced pay and a reprimand from the supervisor.

'How likely do you think it is that staff will complain of an upper limb disorder as a result of this task? Please mark on the scale below'.

Very unlikely ——————————————————— Very likely

Key

Force risk	Present	'Each windscreen weighs upwards of 5.5 kg'
Repetition risk	Present	'They have to replace 120 windscreens in the session'
Postural risk	Present	'operators have to adopt a posture with arms fully outstretched'
Duration of exposure risk	Absent	'the 30 min that staff work in the racks during their shift'
Psychosocial risk	Present	'lone-working'; 'otherwise they receive reduced pay and a reprimand from the supervisor'

Figure 1. Example of a scenario with risk factors identified.

- Other postures highlighted as problematic are repetitively turned or twisted hands, where the palms face up or downwards; holding the hands with the palms facing downwards; using a wide finger and/or hand span to grip, hold or manipulate.
- Postures caused by poorly fitting hand tools in terms of size, shape or handedness.
- Awkward postures were also described for the arms and shoulders; working above the head or with elbows above the shoulders as well as with arms repeatedly moving or held out to the side was cited as risky.
- Static postures of the shoulder or elbow are mentioned along with any other awkward forwards, sideways, backwards or 'across the body' reaching.
- For the 'head and neck' repetitive or held bending or twisting are highlighted as issues, along with awkward postures caused by the visual and/or environmental demands of the work.

2.3.4. Duration of exposure

Duration of exposure could be of interest in terms of the length of time an individual does a task or job as a whole, or in terms of the duration of exposure to a particular, known risk factor. For example, a repetitive, forceful task might be of concern after a shorter duration of exposure than one that is neither of these things. However, even a non-repetitive, 'light' task might be cause for concern if the duration of exposure is extensive. Both represent situations where insufficient recovery time is available for the fatigued musculoskeletal and/or mental system. In the first situation, it is simply that fatigue is reached more quickly. HSG60 gives guidance on the duration of exposure in a number of ways but the focus is on risks (postures, repetitions or forces) to which workers are exposed for 'more than two hours total per work day'.

2.3.5. Psychosocial factors

The following outlines the psychosocial factors flagged in HSG60 as having the potential to cause harm:

- Aspects of the work setup, such as paced work, piecework discouraging breaks, frequent, tight deadlines, sudden changes in workload and unsupervised overtime.
- Aspects of the workers, such as whether they find it difficult to keep up, feel they have a lack of support, have insufficient training/information and have little control.
- A task-specific psychosocial issue, which is whether the job requires high levels of attention and concentration.

2.3.6. Scenarios

All five risk factors were present or absent in all possible combinations in the 32 scenarios that were generated. Some examples of each of the risk factors as used in the scenarios are presented in Table 1. An example of a scenario is represented in Figure 1, with the risk factor 'key' identifying which factors are present and which absent. This key was obviously not included in the scenarios presented to the participants.

2.4. Participants

In order to compare and contrast the judgements of individuals and groups, participants were recruited from a number of different professions known to give ergonomics advice pertaining to musculoskeletal issues. In order to provide sufficient power to discriminate between groups, a minimum of 10 participants was sought from each professional group. To recruit this number of individuals, purposive sampling from locally available professional groups was undertaken. To be included, participants had to attest to being involved with using ergonomics to deal with musculoskeletal health issues in the workplace. They were asked to record the number of years experience they had in workplace health and safety and what ergonomics training, if any, they had undertaken.

As a reference group, first year Loughborough University students taking an introductory ergonomics module were also invited to participate. This group was selected because they had some awareness of MSDs and workplaces from a theoretical standpoint and would be able to relate to the scenarios. They had not, however, had practical experience of dealing with MSDs and/or their associated factors in the workplace. The final numbers of participants are given in Table 3.

2.5. Administering the scenarios

As described above, 32 scenarios were generated in total, incorporating all of the five risk areas (see section 2.3) in all combinations from none up to five. In informal pilot studies, it was found that it took too long to make 32 judgements twice (two judgements on each scenario being required to test for consistency) and participants lost concentration. Therefore, 16 of the scenarios were randomly selected and collated in an A5-sized booklet. The risk factors contained in these 16 scenarios are outlined in Table 2, although the content of the judgements themselves was not of interest in this study.

All of these scenarios were administered twice to the participants so that they had to repeat their judgements for all the cases. The order of the scenarios was randomised using a random number generator and differed between the first and second booklets.

The scenarios were administered with consideration of two conflicting requirements. The first was the need to repeat the scenarios relatively quickly after the first administration, in order to be pragmatic about participants' limited time availability. The second issue was the desire to ensure that this was a test of expertise rather than of memory. It was important to avoid participants recalling their judgements from the first time and simply repeating them. This predicated against a rapid repetition in

Table 1. Examples of the risk factor cues from the scenarios.

Risk factor	Cue absent	Cue present
Repetition	'Customer demand is such that staff retrieve one every 2 or 3 min'	'Operators involved in this task lift single paper clips from a tray and put them in boxes. They count 60, one at a time into each box and fill a box every minute'
Force	'This task involves placing a thin sheet of plastic weighing about 100 g on to the top of each jar as it comes down the line'	'Staff in the deli have to retrieve meat joints weighing around 5 kg from the chiller cabinet. They use one hand only for the retrieval'
Duration of exposure	'150 paper cups are dropped one at a time into a stack by operators during the 25 min period they work here'	'About 60 breadsticks per hour are placed on the line by each operator, and they carry out this task for the majority of their 8 h shift'
Psychosocial	'They can therefore interact easily and freely and enjoy good relationships with the management'	'The atmosphere created by the staff and management in the deli is negative. There is low morale, no team spirit and very little interaction between staff as they work'
Awkward posture	'They can keep a neutral posture due to the specially developed ''meat handles'', which are attached to each joint before they are put in the chiller'	'Because of the line's orientation, operators must bend their wrists to place the breadsticks on the line'

Table 2. Risk factors contained in the scenarios.

Scenario number	Number of risk factors	Specifics
1	None	–
2	1	Repetition
3	1	Duration
4	2	Force and Repetition
5	2	Force and Duration
6	2	Repetition and Posture
7	2	Repetition and Psychosocial
8	2	Posture and Psychosocial
9	3	Force, Posture and Repetition
10	3	Duration, Posture and Psychosocial
11	3	Force, Duration and Psychosocial
12	3	Repetition, Duration and Psychosocial
13	3	Repetition, Force and Duration
14	4	Repetition, Force, Duration and Posture
15	4	Repetition, Force, Posture and Psychosocial
16	4	Repetition, Force, Duration and Psychosocial

favour of a longer time interval between the two administrations. In order to accommodate these two factors, the participants were given a brief explanation about the study, after which they signed consent forms. They then worked through the first scenario booklet, marking the rating scale with a vertical line in the place that best represented their opinion (see example in Figure 1). At this point, nothing was said about the need to repeat the exercise nor about the fact that this would be an expertise 'measure', so as to avoid the participants endeavouring to recall their judgements.

Having carried out this task, participants put their booklets into an envelope and engaged in another activity for between 60 and 90 min. This intervening activity varied from participating in a focus group discussion about ergonomics (HSAs, OHAs, ergonomists) to a lecture about design (students), to a training course on the treatment of shoulders (physiotherapists), to getting on with their own work (ergonomists, physiotherapists). Participants were then given the second booklet and asked to repeat the judgement task, without reference to their initial booklet.

The whole procedure (with 16 scenarios repeated twice) was pilot tested with five individuals, to ensure that the instructions were clear and that the procedure was manageable. This being the case, these data were discarded and recruitment began for the main study.

2.6. *Analysis*

The point at which each participant marked the 100 mm line was measured in mm, treated as % likelihood and input into the CWS calculator software (available at http://www.k-state.edu/psych/cws/software.htm). A total of 10 participants who failed to record a judgement for both repeats of a scenario were excluded from the analysis and are not reported here.

The software generated a CWS index for each participant. Having generated this index, group means were calculated for each profession (group of participants), each training level and years of experience group, using the method suggested by Weiss and

Edwards (2005). The mean CWS for a group is calculated from square of the sum of the square roots of the individual CWSs by:

$$\overline{CWS} = \left(\frac{\sum \sqrt{CWS_i}}{n} \right)^2$$

where $\overline{CWS}$ = group mean CWS; CWS_i = individual i's CWS index; n = number of individuals in the group.

CI were calculated in order to ascertain whether the groups differed statistically in CWS performance.

$$CI = \overline{CWS} \pm t \cdot s_{\overline{x}}$$

where t is used rather than z because group sizes are relatively small so that the sample's standard deviation rather than the population's standard deviation to calculate the CI, avoiding the assumption of normality. $S_{\bar{x}}$ = standard error of the group's CWS index. According to Payton *et al.* (2003), when 84.3% CI overlap, the difference between the groups cannot be said to be statistically significant at the 0.05 level.

3. Results

3.1. Sample characteristics

In total, 208 individuals took part, with 198 producing complete, usable responses. Descriptive statistics in terms of experience and training level are shown in Table 3. In order to rationalise their responses, the participants' ergonomics training levels were coded as described in the table footnote.

The majority of the participants had no additional ergonomics training beyond their professional qualification, with the students having none at all. By contrast, the ergonomists' median was level 4, indicating that a high proportion held a BSc, MSc or PhD degree in ergonomics.

The physiotherapists and OHAs had the most experience practising in the field of occupational health, averaging just over 12 years. The HSAs and ergonomists followed, with 6 and 7 years respectively, while the students had 0 years of experience.

Table 3. Participant characteristics and Cochran–Weiss–Shanteau (CWS) indices of the five groups of participants.

	n	Experience (years) Mean (SD)	Training level (Median)	Group mean CWS index (with 84.3% CI)
Ergonomists	11	7.9 (7.4)	4.0	16.2 (CI 8.3 – 26.7)
OHAs	22	12.4 (7.1)	2.0	7.1 (CI 5.1 – 9.4)
HSAs	11	6.0 (3.7)	2.0	5.9 (CI 4.2 – 7.9)
Physiotherapists	14	12.2 (11.8)	2.0	5.4 (CI 3.4 – 7.8)
Students	140	0	1.0	5.6 (CI 5.0 – 6.2)

OHA = occupational health advisor; HSA = health and safety advisor.

Note: Training levels: 1 = none; 2 = ergonomics training as part of their professional qualification; 3 = ergonomics short course, ergonomics certificate or ergonomics diploma; 4 = ergonomics BSc, MSc or PhD degree.

3.2. *Discrimination and consistency*

Because the CWS is derived from two parameters, consistency and discrimination, there are four different combinations available, if each parameter is considered as low or high. These are: high consistency with high discrimination; low consistency with low discrimination; low consistency with high discrimination; and high consistency with low discrimination. Examples from the dataset of the most and least expert combinations are represented in Figures 2 and 3.

In Figure 2, a highly consistent and discriminating (the most 'expert' combination) participant's judgements are recorded. The ULD likelihoods are plotted for each time this participant made a judgement on each of the 16 scenarios (series 1 being the first time the judgement was made and series 2 being the second). This individual exhibited high consistency (as shown by the similarity of the two judgements at each data point), as well as high discrimination (as demonstrated by the spread of the judgements up the y-axis – from 5% up to over 90% likelihood rating).

In Figure 3, the participant showed low discrimination, with most responses being placed between 20% and 70% likelihood rating. Furthermore, consistency was also low, as demonstrated by the great difference between the two judgements for the same scenario.

3.3. *Effect of profession, training and years of experience on Cochran–Weiss–Shanteau*

According to the group mean CWS indices (Table 3), where a high value of the index indicates a more 'expert' judgement, ergonomists are significantly more 'expert' when making the judgements required of them by these scenarios (mean CWS = 16.2) than their physiotherapist (mean CWS = 5.4) or their health and safety colleagues (mean CWS = 5.9), and than the student group (mean CWS = 5.6). These are statistically significant results, indicted by the non-overlapping CI. Although the ergonomists also had a higher group mean CWS index than their OHA colleagues (mean CWS = 7.1), this was not a statistically significant difference.

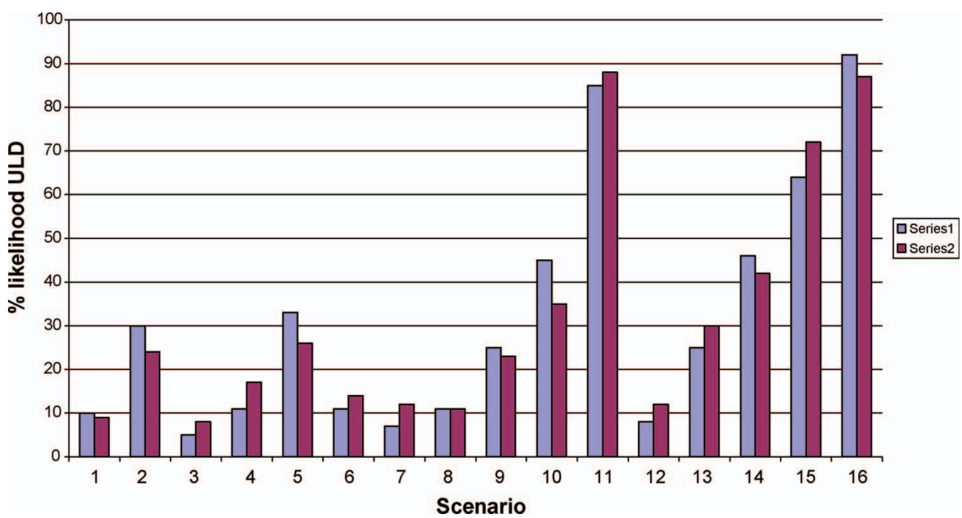

Figure 2. Example of a highly consistent and highly discriminating participant. ULD = upper limb disorder.

From Table 4 it is clear that, as training level increases, so does the mean CWS index. Participants with an ergonomics BSc, MSc or PhD degree performed significantly better than those with no ergonomics training at all, or none beyond their initial professional qualification. The difference between those who had ergonomics degrees and those with short courses/certificates or diplomas is not statistically significant.

Because the whole of the student group had 0 years of experience it was removed from the sample to calculate the correlation coefficient between CWS and years of experience. This was not found to be statistically significant (Pearson's R = 0.034, n = 58).

4. Discussion

At the outset, there were three specific aims for this work, as outlined in section 1.4. Each of these aims will be discussed in turn.

4.1. Who are the most 'expert' judges?

When using the CWS index of performance as a measure of expertise in this study, ergonomists performed significantly more expertly than their physiotherapist or health

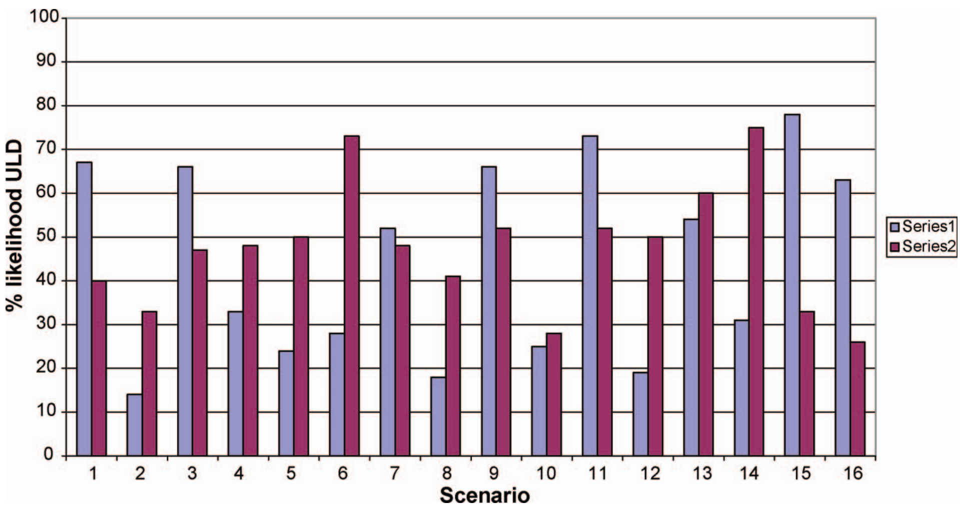

Figure 3. Example of an inconsistent and undiscriminating participant. ULD = upper limb disorder.

Table 4. Training level group mean Cochran–Weiss–Shanteau (CWS) and CI (higher index = better performance).

Training	n	Group mean CWS index	CI
None	140	5.64	5.0–6.2
No training beyond own professional qualification	37	5.71	4.63–6.90
Short course, certificate or diploma in ergonomics	10	8.66	4.7–13.8
Ergonomics BSc, MSc or PhD degree	11	16.2	8.3–26.7

Note: Groups whose CI do not overlap have mean CWS significantly different at the 0.05 level.

and safety colleagues and significantly better than the students. Although their mean CWS index was higher than that of their OHA colleagues, this was not a statistically significant difference. However, the ergonomists were the only group whose CI did not overlap with the students. So having established that ergonomists do perform more 'expertly' when using the CWS index as the metric, what are the available explanations from the data?

4.2. *Factors linked with better performance*

4.2.1. *Level of ergonomics training*

In this study, CWS index varied with level of training, with the mean CWS increasing as training level increased. All of the ergonomists had 'degree level or higher' qualifications in ergonomics. The majority of the other professionals taking part in this study had only the ergonomics training that formed part of their own professional qualification.

The OHAs were the professional group with the largest number of individuals with some additional ergonomics training; 27% of their number (six individuals) had taken short courses or certificates in ergonomics. They were also the highest scoring 'non-ergonomist' professional group in terms of mean CWS index.

Despite the relatively small samples in this study, it is clear that having an ergonomics degree or higher is linked with a higher CWS index compared to having no training at all or to only that which forms part of one's professional qualification. Larger samples will be required to see whether there is a significant difference between the degree level and the short course/certificate/diploma level of training when it comes to judgement performance in these ULD scenarios. What the training findings suggest, however, is that ergonomists perform better because they have had higher level training. This result concurs with the findings of Weston and Haslam (1992), in that increased training is positively linked with performance and extends their findings to include a risk assessment task in concurrence with Jones *et al.* (1999).

4.2.2. *Years of experience*

Having more years of experience has sometimes been linked with expertise (James 2007) and sometimes not (Shanteau *et al.* 2002). In the present study, CWS index did not significantly correlate with years of experience, in line with Shanteau *et al.*'s proposition that experience is a poor predictor of expertise. This issue may be confounded because MSDs have become such a key issue in the last 5–10 years. Consequently, someone with 5 years of experience may have spent all 5 years dealing only with MSDs and could perhaps have honed their risk assessment judgement skills by their focus on this. Their performance may be as good as or better than someone with much more experience who has not focused exclusively on this one skill. This importance for expertise development of the opportunity to practice a specific skill has been noted by other authors studying expertise in fields such as weather forecasting and firefighting (Pliske *et al.* 2004).

4.3. *Usefulness of Cochran–Weiss–Shanteau in the context of ergonomics*

In the study, the CWS index differentiated between those who might be deemed experts (ergonomists) in line with the findings of previous studies (Shanteau *et al.* 2002, Weiss and Shanteau 2003, Weiss *et al.* 2006). Clearly, whilst this study has suggested there are differences in the mental models and levels of expertise amongst the groups, there is much that cannot be assumed and would benefit from further investigation. Undertaking a think-aloud protocol analysis study or questionnaire study alongside measuring the CWS index (Backlund *et al.* 2004, Skånér *et al.* 2005, James 2007) would help to acquire more detail about how and why the participants make the decisions they make. Investigation of the decision-making strategies and schemata in line with other previous work (Piegorsch *et al.* 2006) for all the groups represented here would also be of interest.

4.3.1. Expertise from a decision aid

HSG60 (Health and Safety Executive 2002) was devised as a decision aid for the non-expert, to facilitate workplace analysis and risk assessment for ULDs. It would be an interesting further study to take, for example, a sample of the student group and allow them to use HSG60 as they make their decisions about each scenario. In this manner, it would follow the studies of Stanton and Young (2003) more closely and enable conclusions to be drawn about the validity and reliability of a tool like HSG60 in the hands of novices. It would be of interest then to compare them with another student group who made their decisions without it, as well as seeing how their performance compared to the 'experts' (ergonomists in this study). This would allow for the comparison of high level training (the ergonomists) with live use of a decision aid by those with limited training (the other groups) to ascertain the impact on judgement expertise. This would be helpful in elucidating how expert the higher level performance of the ergonomist actually is. Of further interest might be to see how and if the ergonomist improves by using the decision aid.

4.3.2. Extrapolation from controlled environment to the workplace

A final area of further study would be to see whether superior performance in this judgement study translates into greater success in making judgements in the workplace. Following participants carrying out risk assessments in the field would help to clarify this.

Furthermore, as Cornford and Athanasou (1995) point out: 'What separates the expert from merely the competent performer is that the expert can also tell you how to fix those faults and get things working once more'. However, this is a multi-faceted issue, based not only on the ability to make sound judgements but on a whole host of other factors such as employee and employer willingness to change (Whysall *et al.* 2004).

4.4. Potential limitations of this study

4.4.1. Language and wording

Care was taken to base the scenarios on the criteria outlined in HSG60. However, the scenarios were generated by an ergonomist and potentially described the work situations in a way that favoured the experiential expertise of the ergonomists. That notwithstanding, it was the cues themselves rather than the whole picture that should have informed the decisions.

In addition, the intended cue distribution may have been confounded by the fact that there may be strong risk associations with some specific work environments used in the scenarios. For example 'delicatessens' and 'production lines' are environments historically associated with ULDs and these might be cues in and of themselves when used in a scenario. That said, this effect should remain the same 'within' an individual's repeated responses and, whilst it might affect the discrimination index, it should not impact on their consistency.

4.4.2. Sample size

Practical issues often make it difficult to secure large samples of working professionals. In the present study, it was perhaps fortunate that a sample size of 11 experts was sufficient for the ergonomists to separate themselves from the other professionals. Larger samples would generate smaller CI, which might in turn make it possible to make more fine distinctions between the other groups.

4.4.3. Judgement content

Having examined the nature of the judgements made, as one measure of expertise, further understanding of the expertise of the participants could be gained from examining whether or not their assessment of risk increases in line with the number of risk factors. This is the proposed next step for this research.

5. Concluding remarks

The combination of judgement consistency and discrimination into one index affords the conclusion that ergonomists are quantifiably different from other ergonomics advisors in their judgement performance in the specific context of ULDs. Whilst not the whole picture, this index is one objective measure of expertise. In this study, higher CWS was linked with higher ergonomics training level, but not with longer experience in occupational health and safety. Further study will examine the judgements made with reference to the risk factors contained in the scenarios, to see whether participants increased their judgement of likelihood of staff complaining of an ULD in line with the increase in risk factors.

Acknowledgements

We would like to thank COPE Occupational Health and Ergonomic Services Ltd. for their sponsorship, and the participants for their time.

References

Abdolmohammadi, M.J. and Shanteau, J., 1992. Personal attributes of expert auditors. *Organizational Behavior and Human Decision Processes*, 53, 158–172.

Baber, C. and Stanton, N.A., 1996. Human error identification techniques applied to public technology: predictions compared with observed use. *Applied Ergonomics*, 27, 119–131.

Backlund, L., *et al.*, 2004. GPs' decisions on drug treatment for patients with high cholesterol values: a think-aloud study. *BMC Medical Informatics and Decision Making*, 4, 23.

Bernard, B.P. ed., 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*. Washington, DC: US Department of Health and Human Services.

Buckle, P. and Hoffman, A., 1994. *TUC Guide to assessing WRULDs risks*. London: College Hill Press.

Bureau of Labour Statistics, 2006. *Nonfatal occupational injuries and illnesses requiring days away from work, 2005* [online]. Available from: http://www.bls.gov/news.release/pdf/osh2.pdf [Accessed 8 September 2007].

Card, S.K., Moran, T.P., and Newell, A., 1983. *The psychology of human–computer interaction*. Hillsdale, NJ: Erlbaum.

Cherry, N.M., *et al.*, 2001. The reported incidence of work-related musculoskeletal disease in the UK: MOSS 1997–2000. *Occupational Medicine*, 51, 450–455.
Colombini, D., 1998. An observational method for classifying exposure to repetitive movements of the upper limbs. *Ergonomics*, 41, 1261–1289.
Colombini, D., Occhipinti, E., and Grieco, A., 2002. *Risk assessment and management of repetitive movements and exertions of upper limbs: Job analysis, OCRA risk index, prevention strategies and design principles*. Amsterdam: Elsevier Science.
Cornford, I. and Athanasou, J., 1995. Developing expertise through training. *Industrial and Commercial Training*, 27, 10–18.
David, G.C., 2005. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational Medicine*, 55, 190–199.
Dempsey, P.G., 2007. Effectiveness of ergonomics interventions to prevent musculoskeletal disorders: Beware of what you ask. *International Journal of Industrial Ergonomics*, 37, 169–173.
Dempsey, P.G., McGorry, R.W., and Maynard, W.S., 2005. A survey of tools and methods used by certified professional ergonomists. *Applied Ergonomics*, 36, 489–503.
Dreyfus, H.L., Dreyfus, S.E., and Athanasiou, T., 1986. *Mind over machine: the power of human intuition and expertise in the era of the computer*. New York: Free Press.
European Foundation for the Improvement of Living and Working Conditions, 2007. *Fourth European working conditions survey*. Dublin: European Foundation for the Improvement of Living and Working Conditions.
Farrington-Darby, T. and Wilson, J.R., 2006. The nature of expertise: A review. *Applied Ergonomics*, 37, 17–32.
Graves, R.J., *et al.*, 2004. Development of risk filter and risk assessment worksheets for HSE guidance –'Upper Limb Disorders in the Workplace' 2002. *Applied Ergonomics*, 35, 475–484.
Grubbs, F.E., 1973. Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics*, 15, 53–66.
Health and Safety Executive, 2002. Upper limb disorders in the workplace. HSG60(rev). 2nd ed. Sudbury: HSE Books.
Health and Safety Executive, 2005. *Health and Safety Statistics 2004/05*. London: HSE.
James, G.A., 2001. Clinical reasoning in novices: refining a research question. *British Journal of Therapy and Rehabilitation*, 286–293.
James, G.A., 2007. Modelling diagnosis in physical therapy: a blackboard framework and models of experts and novices. *Ergonomics*, 50, 335–351.
Jones, J.R., 1998. *Self-reported work-related illness in 1995: results from a household survey*. Sudbury: HSE Books.
Jones, J.R., Cockcroft, A., and Richardson, B., 1999. The ability of non-ergonomists in the health care setting to take manual handling risk assessments and implement changes. *Applied Ergonomics*, 30, 159–166.
Keyserling, W.M., Stetson, D.S., and Silverstein, B.A., 1993. A checklist for evaluating ergonomic risk factors associated with upper extremity cumulative trauma disorders. *Ergonomics*, 36, 807–831.
Keyserling, W.M. and Wittig, S.J., 1988. An analysis of experts' ratings of ergonomic stress. *International Journal of Industrial Ergonomics*, 2, 291–304.
Li, G. and Buckle, P., 1999. *Evaluating change in exposure to risk for musculoskeletal disorders: a practical tool*. Sheffield: Health and Safety Executive.
McAtamney, L. and Corlett, E., 1993. RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24, 91–99.
National Institute for Occupational Safety and Health, 1981. *A work practices guide for manual lifting*. Cincinnati, OH: US Department of Health and Human Services (NIOSH).
Occhipinti, E. and Colombini, D., 2007. Updating reference values and predictive models of the OCRA method in the risk assessment of work-related musculoskeletal disorders of the upper limbs. *Ergonomics*, 50 (11), 1727–1739.
Payton, M.E., Greenstone, M.H., and Schenker, N., 2003. Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science*, 3:34 [online]. Available from: insectscience.org/3:34 [Accessed 10 September 2007].
Piegorsch, K.M., *et al.*, 2006. Ergonomic decision-making: A conceptual framework for experienced practitioners from backgrounds in industrial engineering and physical therapy. *Applied Ergonomics*, 37, 587–598.

Pliske, R.M., Crandall, B., and Klein, G., 2004. Competence in weather forecasting. *In*: K. Smith, J. Shanteau, and P. Johnson, eds. *Psychological investigations of competence in decision making*. Cambridge: Cambridge University Press, 40–68.

Shanteau, J., 1988. Psychological characteristics and strategies of expert decision makers. *Acta Psychologica*, 68, 203–215.

Shanteau, J., *et al.*, 2002. Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, 136, 253–263.

Skånér, Y., *et al.*, 2005. General practitioners' reasoning when considering the diagnosis heart failure: a think-aloud study. *BMC Family Practice*, 6, 4.

Stanton, N.A. and Stevenage, S., 1998. Learning to predict human error: issues of acceptability, reliability and validity. *Ergonomics*, 41, 1737–1756.

Stanton, N.A. and Young, M.S., 2003. Giving ergonomics away? The application of ergonomics methods by novices. *Applied Ergonomics*, 34, 479–490.

Washington State Department of Labor and Industries, 2000. *Ergonomics rule, (WAC) WAC296–62–051, Ergonomics* [online]. Available from: http://www.lni.wa.gov/wisha/ergo/rule_docs/ces/Title,%20TOC%20and%20Exec%20Summary.PDF [Accessed 10 September 2007].

Weiss, D.J. and Edwards, W., 2005. A mean for all seasons. *Behavior Research Methods*, 37, 677–683.

Weiss, D.J. and Shanteau, J., 2003. Empirical assessment of expertise. *Human Factors*, 45, 104–116.

Weiss, D.J., Shanteau, J., and Harries, P., 2006. People who judge people. *Journal of Behavioural Decision Making*, 19, 441–454.

Westgaard, R.H. and Winkel, J., 1996. Guidelines for occupational musculoskeletal load as a basis for intervention: a critical review. *Applied Ergonomics*, 27, 79–88.

Weston, J.D. and Haslam, R.A., 1992. How successfully can non-ergonomists recognise and apply ergonomic principles? *In*: E.J. Lovesay, ed. *Contemporary Ergonomics*. Birmingham: Taylor & Francis, Ergonomics Society, 334–339.

Whysall, Z.J., Haslam, R.A., and Haslam, C., 2004. Processes, barriers, and outcomes described by ergonomics consultants in preventing work-related musculoskeletal disorders. *Applied Ergonomics*, 35, 343–351.

Winnemuller, L., *et al.*, 2004. Comparison of ergonomist, supervisor and worker assessments of work related musculoskeletal risk factors. *Journal of Occupational and Environmental Hygiene*, 1, 414–422.