Validity of the APHIRM Toolkit Hazard and Risk Assessment Survey

PURPOSE AND CONTEXT OF APHIRM SURVEY USE

This survey is one of a set of risk management procedures in the APHIRM Toolkit. It is designed **to assess hazard levels and enable identification of those that present the greatest risk of work-related musculoskeletal disorders (MSDs) and of stress-related mental health problems (MHPs)**. Toolkit users are guided through each phase of the conventional OHS risk management cycle. In addition to the survey, the toolkit includes procedures to identify work-related *causes* of identified hazards, and procedures to select risk control actions in accord with the hierarchy of risk control. Its ongoing use promotes continual improvement, with effectiveness of risk control actions being evaluated when workers are surveyed again during the following risk management cycle. Toolkit development was informed by our collaborative work with the World Health Organisation network of Collaborating Centres in Occupational Health to develop a generic format for risk management 'toolkits' suitable for workplace use (Macdonald, 2012) and by 'implementation science' principles and evidence (Meyers et al, 2012).

The APHIRM survey includes the following components, which assess levels of:

- **work-related hazards:** biomechanical hazards that may affect MSD risk, and psychosocial hazards that may affect both MSD and MHP risk
- **musculoskeletal discomfort/pain** (scored out of 60), indicating MSD risk
- **mental stress** (scored out of 48), indicating risk of mental health and other stress-related related problems.

Because both MSD and MHP risk can be affected by very large numbers of hazards, it is not 'reasonably practicable' (SafeWork Australia, 2013) for workplaces to control risk from all of them during the same risk management cycle. The toolkit therefore includes an algorithm which identifies the *worst* hazards *at that time point*, based on rated hazard exposure levels (severity, frequency/duration) and relationships between hazard exposure ratings and scores on each outcome measure (discomfort/pain, stress), in accord with research on methods to identify hazards that present the greatest risk (Clarke & Cooper, 2004).

HOW CAN 'VALIDITY' OF THE SURVEY BE EVALUATED?

At its most basic level, 'validity' refers to:

> "… how well the assessment tool actually measures the underlying outcome of interest. Validity is not a property of the tool itself, but rather of the interpretation or specific purpose of the assessment tool" (Sullivan, 2011)

Validity must be evaluated in relation to its intended purpose, the context in which it is used and results interpreted (Knetka et al, 2019).

Research on the validity of risk assessment methods such as the APHIRM survey have typically considered the following three aspects of validity: *construct* validity, *content* validity, and *criterion* validity – both *concurrent* and *predictive* (Takala et al., 2010; Stanton, 2016; Macdonald & Oakman, 2022). Consistent with these domain-specific practices, the more general *Standards for Educational and Psychological Testing* (AERA, APA and NCME, 2014) identify five different types of evidence that may be relevant to assessing validity, including:

1. Evidence that assessment scores are related as expected to scores from other valid indicators of the same or related constructs – that is, evidence of **criterion validity**.
2. Evidence that relationships between assessment items is consistent with structure of the construct being assessed – that is, evidence of **construct validity.** In the present context, this aspect of validity has been defined

as "assessment of "hazard constructs identified as important by an evidence-based theoretical model of factors affecting musculoskeletal and/or mental health disorders" (Macdonald & Oakman, 2022).

3. Evidence of a relationship between content of the assessment tool and the construct it is intended to measure – that is, evidence of **content validity.** This requires that survey items assess manifestations of the construct that are most relevant in the particular case (Macdonald & Oakman, 2022). This requires that items are meaningful in relation to respondents' particular job characteristics and working environments, and also that they can be accurately interpreted which means that respondents' literacy requires consideration. For example, survey items suitable for use with tertiary-educated respondents in professional or managerial roles may require some adaptation to ensure content validity when used with 'blue collar' workers.

The other two types of 'validity evidence' identified by those *Standards* are:

4. Evidence that survey respondents are motivated to provide valid responses to assessment items, and
5. Evidence that assessment consequences are in accord with the intended purpose of assessment.

These both refer to the context in which assessment occurs, and there is some debate concerning whether consequences should be seen as part of validity (Knetka et al, 2019). However, both of these factors are significant considerations in the context of workplace risk management and are addressed in the design of various APHIRM Toolkit components, including the information provided to survey respondents which assures them that their responses will be entirely anonymous, they will have the opportunity to comment on survey results and that results will be acted on to improve workplace risk management.

Using criteria 1 to 3 above, evidence of the survey's validity for its intended purpose is outlined below, separately for each of the survey components: hazards, musculoskeletal discomfort/pain, mental stress.

VALIDITY OF THE APHIRM SURVEY'S ASSESSMENT OF HAZARD LEVELS

Construct validity. The survey is intended to assess levels of all physical (biomechanical) hazards that may influence MSD risk, and all psychosocial hazards that may influence risks of MSDs and MHPs. It is also designed to be brief so that the time required for workers to complete the survey during working time is as short as possible – which is essential to enable its widespread adoption and sustainability as part of routine workplace risk management. The very large number of psychosocial hazards that have been identified as potentially affecting risk therefore presents a challenge. Kop et al (2016) noted the absence of any agreed reference taxonomy of the psychosocial characteristics of workplace environments. They created such a taxonomy by categorizing factors present in the main theoretical models of the occupational health domain and then used this to analyse content of the 17 'work environment' questionnaires most frequently cited in scientific literature. This analysis revealed major differences in the content of these questionnaires, with none providing comprehensive coverage.

Based on their analysis of these surveys, Kop et al suggested that users should "select a questionnaire on the basis of how well its content corresponds to their needs in relation to their problem", since coverage of all possible topics would require a questionnaire that was much too long. In accordance with this, content of the APHIRM survey has focused on developing content shown to be most appropriate for the purpose and context of its use. Selection of constructs has taken account of evidence that the relative impacts on risk of some hazards can vary between occupational groups such as blue-collar versus white-collar jobs (Oakman and Chan, 2015; Metzler and Bellingrath, 2017; Berthelsen et al., 2020).

Initial development was based on an extensive review of international literature (Macdonald & Evans, 2006) followed by exploratory research in four large Australian organisations where MSD risk was recognised as problematic (Macdonald et al, 2008). Its development was also informed by our previous research on stress-related issues among 'blue collar' workers in jobs where work rates were largely or entirely determined by assembly line speed or by stringent production targets and deadlines (Macdonald 2000, 2003a, 2003b). The Work Organization Assessment Questionnaire (WOAQ; Griffiths et al, 2006) was an important initial source of psychosocial hazard items because it was developed for use with both white- and blue-collar workers in manufacturing industry and similar environments, so content validity of its items for workers in jobs where MSD risk is high was better than that of most psychosocial hazard surveys. WOAQ items relate to five factors: workload issues, reward and recognition, quality of relationships with management, quality of relationships with colleagues, and quality of physical environment.

Over more recent years this APHIRM survey has been amended based on current research evidence on the work-related determinants of MSD risk, occupational stress and associated health and safety problems. The Copenhagen Psychosocial Questionnaire (COPSOQ; Burr et al, 2019) has provided a particularly useful source of items to assess stress-related health problems because it is based on multiple theoretical frameworks, has been updated in accord with new research evidence and validated in a wide variety of contexts (see https://www.copsoq-network.org/validation-studies/). Table 1 below shows both the WOAQ factors and COPSOQ scales addressed by this APHIRM survey.

Poor job security is not included because job insecurity is typically influenced strongly by factors beyond the workplace including the 'job market' and overall economic climate, whereas the primary purpose of this survey is to support workplace risk management and at workplace level there is little if any capacity to eliminate or effectively reduce risk from such factors. Also, early trialling showed that its association with stress was very variable and never among the strongest correlates of stress or discomfort/pain.

Table 1. Psychosocial hazard items in the APHIRM toolkit survey, grouped according to the most closely corresponding COPSOQ scales and WOAQ factor labels.

COPSOQ scales	WOAQ factors	Number of APHIRM survey items
Quantitative demands	1. Workload issues	3
Work pace	"	2
Quality of work	NA	1
Emotional demands	NA	2
Influence at work	NA	1
Possibilities for development	2. Reward and recognition	4
Variation of work	"	1
Meaning of work	NA	2
Predictability	3. Quality of relationships with management	1
Recognition	"	1
Role clarity	"	2
Quality of leadership	"	2
Social support from supervisor	"	3
Vertical trust	NA	1
Role conflicts	NA	2
Illegitimate tasks	NA	1
Social support from colleagues	4. Quality of relationships with colleagues	2
Sense of community at work		
Organisational justice	NA	3
NA	5. Quality of physical environment	6

When surveys such as this are used for research purposes, scores for groups of items representing higher-level constructs or factors would normally be used. However, the factor structure of the APHIRM survey is not relevant when the survey is used for workplace risk management. Translation of survey results into effective risk control actions for people *in a specific job* requires use of item-level scores rather than factor scores; the toolkit's online algorithm identifies the 'top 10' hazard items to be prioritised for risk control actions and surveyed workers are asked to provide feedback on the likely sources or causes of each hazard. This requires that hazards are identified in words that are clearly interpretable in relation to potential risk control actions. Labels given to factors representing *groups* of items (i.e. factors) are necessarily much broader in meaning and too imprecise for this purpose, whereas the wording of individual hazard items is sufficiently specific to elicit useful feedback from the workers concerned. Workplace reporting of survey results in terms of higher-level constructs would not achieve this essential purpose of the APHIRM survey – that is, translation of risk assessment into effective risk control.

Nevertheless, factor analysis (PCA) of results from 2,327 respondents to the current 52-item online version of the APHIRM survey has been performed for research purposes, since factor scores are more useful than item scores in

analyses to demonstrate the criterion validity of hazard assessment in relation to musculoskeletal discomfort/pain. This analysis demonstrated the existence of 10 distinctly separate and theoretically meaningful factors: three biomechanical and seven psychosocial; results are shown in Table 2 below. Eleven items fell into two *Physical Task Demands* factors unique to APHIRM, and 40 fell into one of seven psychosocial factors. The remaining item (prolonged sitting) was accepted as a tenth (single-item) factor since there is ample evidence that lack of physical activity during work can negatively affect health, including MSD risk (Coenen et al, 2018).

The only item with an unusually low factor loading (0.33) was "Influence on decisions about your work (e.g. what you do, how you do it, how much of it)" in Factor 1B *Job Development Opportunities*. Such an item is not part of WOAQ, which was developed for workers in similar occupations to those in the present study; importantly, items for that questionnaire were selected based on discussions with stakeholders "to test their current relevance to the manufacturing industry"(Griffiths et al, 2006). Our own previous research on factors affecting the health of workers performing highly repetitive work in Australian manufacturing industry suggested that 'having a say' (e.g. via genuinely two-way consultation) might be a more valid construct in such workplace contexts (Macdonald, 2003b). However, 'Influence' was retained within the APHIRM survey because of its broader theoretical significance, along with WOAQ items 'Consultation about changes in your job' and 'Senior management attitudes'. It is noteworthy that the latter two items fall within the APHIRM factor *Quality of Management and Leadership* rather than with the 'Influence' item in *Job Development Opportunities*.

As shown in Table 2, the seven APHIRM psychosocial factors mapped very clearly onto WOAQ factors (Griffiths et al, 2006). COPSOQ scales appeared either as directly corresponding to a APHIRM factor such as *Co-worker relationships*, or combined within a single APHIRM factor such as COPSOQ 'quantitative demands' and 'work pace' scales together constituting the APHIRM *Workload: Quantity & Pace* factor. Table 2 also shows that Cronbach alpha values for each factor are uniformly acceptable. Overall, the factor structure of the APHIRM survey was found to have excellent construct validity for its intended purpose.

Table 2. APHIRM survey hazard factors identified by Principal Components Analysis. For items where there is an equivalent WOAQ item, numbers in the WOAQ column refer to the WOAQ factors shown in Table S1 above.

APHIRM Factor	WOAQ	APHIRM toolkit item	Factor loading	Mean (SD)	Cronbach α
1A. QSL: Quality of Supervision & Leadership (10 items)		Arguments/ problems are sorted out in a fair way	0.80	2.86 (1.05)	0.92
		People here are treated fairly	0.78	2.78 (1.12)	
		Employees can trust information that comes from management	0.73	2.90 (1.13)	
		Work is shared out fairly between people	0.67	2.90 (1.07)	
	1	Support from supervisor	0.67	2.49 (1.20)	
	1	Your work is noticed and appreciated by your supervisor or manager	0.64	2.97 (1.16)	
	1	Senior management attitudes	0.64	3.07 (1.21)	
	1	Communication with supervisor	0.60	2.54 (1.15)	
	1	Feedback on performance	0.58	2.86 (1.16)	
	2	Consultation about changes in your job	0.54	3.00 (1.07)	
1B. JDO: Job Development Opportunities (7 items)	2	Opportunities for learning new skills	0.79	3.01 (1.16)	0.85
	2	Opportunities to use your skills	0.78	2.72 (1.02)	
	2	Amount of variety in the work you do	0.72	2.76 (1.01)	
	2	Opportunities for promotion	0.71	3.32 (1.06)	
	2	Sufficient training for the job	0.61	2.83 (1.15)	
	2	Flexibility of working hours	0.57	2.79 (1.15)	
		Influence on decisions about your work (e.g. what you do, how you do it, how much of it)	0.33	3.00 (1.12)	
2. PTDWB: Physical Task Demands		Push or pull things	0.79	2.97 (1.21)	0.88
		Squat or kneel while you work	0.79	2.97 (1.27)	
		Lift or carry things that are heavy	0.76	2.84 (1.16)	

Whole Body (8 items)		Work in twisted or awkward postures	0.76	2.71 (1.19)	
		Work with arms raised above shoulder level	0.73	2.20 (1.11)	
		Work so hard or fast that you get a little out of breath	0.68	2.09 (1.01)	
		Work with your body bent forward	0.64	2.82 (1.15)	
		Work standing in one position	0.53	2.30 (1.13)	
3. WQP: Workload: Quantity & Pace (6 items)	3	Have to work at a fast pace for the whole shift	0.76	2.80 (1.03)	0.85
	3	Too much work to do in the available time	0.76	2.79 (1.12)	
	3	Have to go faster to meet deadlines or target quotas	0.75	2.84 (1.09)	
	3	Have to work very fast	0.74	3.14 (1.03)	
	3	Get behind with your work	0.64	2.32 (1.07)	
	3	Have enough time to complete all your work well	0.61	2.31 (1.07)	
4. RCED: Role Conflicts & Emotional Demands (5 items)	1	People disagree about the correct way to do some things	0.54	2.75 (1.00)	0.71
		As part of your work, you have to help people who are upset or unhappy	0.51	3.09 (1.08)	
		Some parts of your job seem unnecessary or a waste of time	0.49	2.55 (1.06)	
		Your work puts you in emotionally disturbing or upsetting situations	0.45	2.44 (1.03)	
		People take short cuts to get things done, rather than use correct procedures	0.45	2.58 (1.11)	
5. MCSW : Meaningful & Clearly Specified Work (4 items)		The work you do is important	0.76	2.02 (0.95)	0.78
		Your work is meaningful … doing it well makes a difference to people	0.75	2.30 (1.07)	
	1	You know exactly what work you are expected to do and how to do it	0.51	2.09 (0.82)	
	1	Your work goals and responsibilities are clear	0.46	2.36 (0.91)	
6.PE: Physical Environment (6 items)	5	Physical environment hazards (noise, light, temperature, etc)	0.70	3.03 (1.05)	0.83
	5	Facilities for breaks, meals	0.68	2.81 (1.18)	
	5	Exposure to physical danger	0.58	3.07 (1.00)	
	5	Work stations and work space	0.49	2.92 (1.00)	
	5	Equipment, tools, IT or software	0.42	3.14 (1.11)	
	5	Health & safety	0.41	2.69 (1.13)	
7. PTDHF: Physical Task Demands Hands/Fingers (3 items)		Use your hands or fingers to hold or grip things	0.66	3.83 (1.21)	0.67
		Keep repeating the same movements, every minute or so	0.66	3.48 (1.13)	
		Have to make very precise movements to place things accurately	0.62	3.21 (1.21)	
8. CR: Co-worker Relationships (2 items)	4	How well you work with your co-workers	0.82	2.19 (1.05)	0.87
	4	How well you get on with your co-workers	0.82	2.14 (1.01)	
9. PS: Prolonged Sitting (1 item)		Work sitting still without moving around	0.72	2.45 (1.31)	NA

More recently, a new item has been added to Construct 4 RCED *Role Conflicts & Emotional Demands*: "There are unpleasant arguments or conflicts at your workplace". Using a small set of data from manufacturing workplaces, Cronbach alpha for this scale after addition of the new item was found to be 0.78, compared with the previously determined 0.71 (Table 2 above).

In addition, three new hazard assessment items address sexual harassment and bullying (Never, A Few times, Monthly, Weekly, Daily). Each of these items includes a brief explanation of key words in the question; wording of

these questions and associated explanations was based on review of research and existing questionnaires that specifically address this topic. The questions are:

- *Have you OR someone else at your workplace experienced workplace aggression or violence by a co-worker, manager, or member of the public?*
  Workplace aggression or violence includes threats of violence; angry shouting or finger-pointing or invasion of personal space; intimidating new/young workers (e.g. hazing); and actual physical attacks, including by coughing or spitting.
- *Have you OR someone else at your workplace been bullied by a co-worker, manager, or member of the public?*
  Bullying includes repeated unpleasant teasing, or repeated actions or words that harass, humiliate, or unfairly target or exclude someone – in person or via social media, texts, etc.
- *Have you OR someone else at your workplace been sexually harassed by a co-worker, manager, or member of the public?*
  Sexual harassment is unwelcome sexual comments or actions - in person or via social media, texts, etc.

It is assumed that these three items constitute a separate, additional scale; when a larger data set is available, further factor analyses will be conducted to confirm this. Based on the existing small data set, Cronbach Alpha of this 3-item scale is 0.74.

Content validity. As described above, content validity requires that survey items are appropriate for the ways in which the targeted construct may be manifested in that type of job, and for the likely range of literacy levels of workers in such jobs. When there were existing WOAQ items that addressed a targeted construct, these were usually preferred over others because of their greater content validity for the present purpose. However, WOAQ does not cover all relevant constructs so item development often entailed amendments to wording of items from existing scales to enhance their content validity for the present purpose. For example, the two core items in the COPSOQ *Emotional Demands* scale were amended from: "Do you have to deal with other people's personal problems as part of your work" to "As part of your work, you have to help people who are upset or unhappy" and from "Is your work emotionally demanding" to: "Your work puts you in emotionally disturbing or upsetting situations".

Criterion validity – Hazards scores in relation to Discomfort/Pain score. Effects on workers' musculoskeletal discomfort/pain levels of exposures to physical (biomechanical) hazards are long established (National Research Council, 2001), and effects of exposures to psychosocial hazards are now very well established also (National Research Council, 2001; Marras et al., 2009; Hauke et al., 2011; Eatough et al., 2012; Lang et al., 2012; Widanarko et al., 2015; Neupane et al., 2017; Candan et al., 2019; Bodin et al., 2020; Dianat et al., 2020; Tang, 2020; Graveling et al., 2021; Taibi et al., 2021; Liu et al., 2020). Our own research in developing the APHIRM survey has confirmed such evidence. Using data from earlier versions of the APHIRM survey, collected in a wide range of industry sectors where MSD risk is problematic, we have found that both physical and psychosocial hazard measures were strongly associated with workers' discomfort/pain levels (Macdonald et al, 2008; Oakman, 2014; Oakman et al, 2014; Oakman and Chan, 2015). (Survey items used to generate the APHIRM discomfort/pain score have remained constant throughout the survey development process; validity of this score is discussed below.)

In common with most other researchers on this topic, our earlier research used multiple regression analysis to demonstrate relationships between hazards and discomfort/pain. However, that form of analysis aims to identify variance in MSD risk that is attributable uniquely to each separate hazard, which can be problematic when many of these hazards interact with each other in their effects on risk (National Research Council, 2001; Marras et al, 2000; Widanarko et al, 2015). It has been pointed out that: "The typical, real-life impact of co-occurring work factors may be obscured when they are studied separately or mutually adjusted. If a factor is typically a part of a factor complex with a substantial net impact on pain, studying that factor separately may conceal its contribution" (Christensen et al, 2018). This problem can be addressed by use of Latent Profile Analysis (LPA), which identifies subgroups of workers who typically share particular combinations of hazard levels that can be described as representing typical 'work situations'.

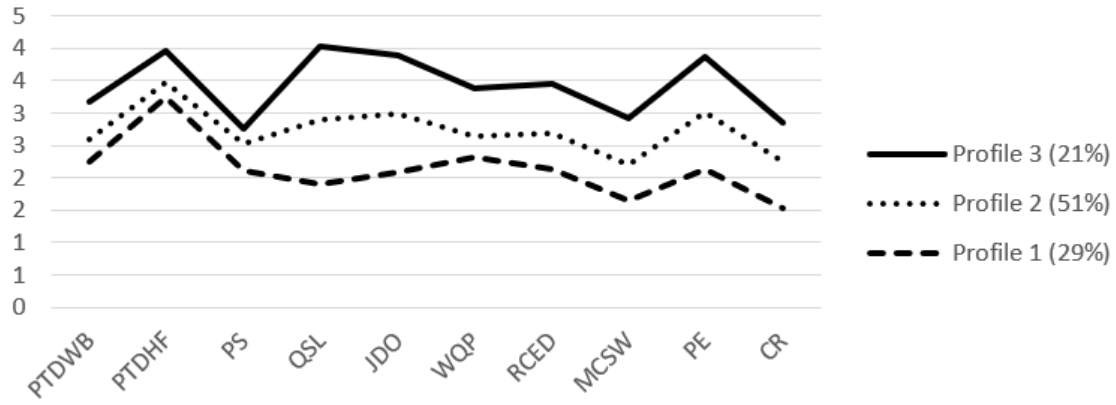

Figure 1. Estimated scores for hazard factors: PTDWB Physical Task Demands Whole Body, PTDHF Physical Task Demands Hands/fingers, PS Prolonged sitting, QSL Quality of Supervision & Leadership, JDO Job Development Opportunities, WQP Workload Quantity & Pace, RCED Role Conflicts & Emotional Demands, MCSW Meaningful & Clearly Specified Work, PE Physical Environment, CR Co-worker Relationships.

Using scores for each of the APHIRM hazard factors identified in Table 2 above, we applied LPA to data from 2,329 respondents to the current online version of the APHIRM survey. Three groups of workers with significantly different hazard profiles were identified, as shown in Figure 1. Mean discomfort/pain scores for the three groups were substantially different and highly significant. For Profiles 1 to 3 respectively, mean(SD) values were: 6.69 (7.24), 10.74 (9.31) and 17.46 (11.8). That is, the mean MSP score (out of 60) of people in Profile 3 was almost 11 points higher than that of people in Profile 1. This clearly demonstrates that the APHIRM hazard survey has excellent validity in relation to the discomfort/pain score criterion.

Criterion validity – Hazard scores in relation to Stress score. The initial 2007 formulation of this survey included a measure of stress: the Exhaustion scale from the General Well-Being Questionnaire. This was developed by Cox and colleagues (Cox & Griffiths, 2005; Cox et al, 2006), and was used as a criterion measure during development of WOAQ (Griffiths et al, 2006) which was an important source of this APHIRM survey's hazard items. To minimise survey completion time that stress scale was subsequently deleted from the present survey, but a similar scale to measure stress has recently been added (see section below on Validity of the APHIRM Toolkit's Stress Score) so that toolkit users now have the choice of focusing on management of MSD risk *or* risk of stress-related mental health disorders *or* both together.

A very large body of research evidence demonstrates effects of workplace hazard exposures such as those assessed by this survey on workers' stress levels and associated health risks (e.g. Leka & Jain, 2010; Harvey et al, 2017; Metzler & Bellingrath, 2017; Chen et al, 2018). Our own earlier research on relationships between hazard scales and the stress measure used in the initial version of this survey found substantial overlap between the hazards most strongly associated with stress and those associated with discomfort/pain (Macdonald et al, 2008), consistent with the well-documented role of stress in the aetiology of both musculoskeletal and mental health disorders. Research on relationships between the toolkit's current set of hazard factors and the new stress scale is underway and will be published when a larger data set is available.

VALIDITY OF THE APHIRM TOOLKIT'S MUSCULOSKELETAL DISCOMFORT/PAIN SCORE

Construct and content validity. A high proportion of work-related diagnosed MSD cases are non-specific – that is, they do not fit criteria for diagnosing a particular type of MSD (Punnett & Wegman, 2004; Roquelaure et al, 2006; Hagberg et al, 2012). In this workplace context the construct being assessed is musculoskeletal discomfort and pain, often in more than one body region, that is either *caused or exacerbated by a wide range of work-related hazard exposures.* This differs from clinical research contexts where the construct of interest is a particular type of disordered *individual* functioning*,* which has implications for methods used to assess the validity of MSD symptom surveys. Researchers have used response item analysis to evaluate the validity of a survey of musculoskeletal symptoms (Norton et al, 2019), but that form of analysis assumes that a single latent construct is being assessed by a

multi-item scale (Yang & Kao, 2014). Such an assumption may be defensible when the symptoms assessed are viewed as arising from a single over-arching cause, such as inflammatory arthritis in a clinical population (which was the case in the analysis by Norton and colleagues). However, when the targeted construct is *work-related* musculoskeletal discomfort/pain arising from a diverse range of workplace hazard exposures rather than from a particular medical condition, there are not multiple items assessing a single construct so item response theory is inapplicable.

Having considered such complexities, a committee of occupational physicians representing the International Commission on Occupational Health (ICOH) concluded, based on extensive consultation with other physicians and researchers, that workplace risk management should focus on assessing musculoskeletal discomfort or pain (Hagberg et al, 2012). Importantly, they recommended that "measures must include aspects such as the severity, frequency, and intensity of pain". Consistent with these recommendations, this APHIRM survey includes a discomfort/pain score which serves as the primary indicator of *work-related* MSD risk; survey items elicit ratings of both frequency and severity for each of five body regions, which are combined to score overall severity of the discomfort/pain. The construct validity of this assessment method is therefore clear. And since it has been used successfully since 2007 with large numbers of workers in a very diverse range of jobs, its content validity is also well established.

Criterion validity. From the viewpoint of workplace risk managers, probably the most important evidence of the validity of a discomfort/pain measure is its relationship with MSD-related workplace costs arising from associated lost time, workers' injury compensation claims, absenteeism, and so on. Unfortunately, we have found that such relationships are difficult to establish because workplace records typically do not include sufficient information to document details of the work and job history of survey respondents, nor of individuals involved in recorded incidents/claims, and there is insufficient information to enable calculation of incident rates per person-hours worked separately for different work groups over a sufficiently long period. Nevertheless, Macdonald et al (2008) analysed records from a large organisation with unusually good records and were able to use data from a very large work group to identify a close relationship between distribution of discomfort/pain scores across the five body regions and distribution of documented MSD incidents across those five body regions, which suggests that this discomfort/pain score would be predictive of claims rates. They also found that discomfort/pain score was strongly predictive of 'yes' responses to the question: "Have you ever taken time off work because of this discomfort/pain?" If possible, however, a large longitudinal research project to document much more fully the relationships between survey responses and MSD-related workplace costs would be helpful in promoting to workplace managers the value of managing all types of hazards that affect musculoskeletal discomfort/pain levels.

VALIDITY OF THE APHIRM TOOLKIT'S STRESS SCORE

Construct and content validity. Occupational or work-related stress is a complex, multidimensional phenomenon defined by the World Health Organization as "the response people may have when presented with work demands and pressures that are not matched to their knowledge and abilities and which challenge their ability to cope" (World Health Organization, 2020). This WHO definition of stress reflects a large body of research which has concluded that: "The experience of stress is therefore defined by, first, the person's realisation that they are having difficulty coping with demands and threats to their well-being, and, second, that coping is important and the difficulty in coping worries or depresses them." (Cox et al, 2000, p.42). It has also been clearly established that chronically high stress increases the risk of 'burnout' with associated work performance deficiencies and health problems including MSDs (Edu-Valsania et al, 2022).

The new APHIRM survey scale to assess 'stress' was formulated to assess the construct as outlined above. All except one of its 12 items are from the following COPSOQ 'Health and Well-being' domain scales:

- *Stress* (3 items: problems relaxing, been tense, been irritable)
- *Cognitive Stress* (4 items: difficulty with thinking clearly, concentrating, making decisions, remembering)
- *Burnout* (4 items: worn out, physically exhausted, tired, emotionally exhausted).

The twelfth APHIRM survey item is "Had difficulty in falling or staying asleep", which is from the *General Well Being Questionnaire* (GWBQ) that was used by Cox and colleagues to evaluate effects on workers of using WOAQ to improve psychosocial hazard management. This item was selected instead of the 4-item COPSOQ 'Sleeping Troubles' scale, aiming to keep survey length to a minimum. GWBQ (and WOAQ) items were developed for use in workplace

risk management so their content validity was assumed to be adequate – confirmed by process evaluation during trialling in Australian manufacturing workplaces.

Factor analysis of the responses currently available from trialling this new APHIRM scale in manufacturing workplaces confirm its construct validity. As shown in Table 3, results identified three factors corresponding to the above three COPSOQ scales. The additional 'difficulty sleeping' item fell within the factor corresponding to COPSOQ *Cognitive Stress*, which is consistent with the known effects of insufficient sleep on cognitive capacities and performance.

Table 3. Factors identified by analysis of the APHIRM survey 'Stress' scale.

Factor	Survey item	Factor loading	Mean (SD)	Cronbach α
COPSOQ *Cognitive Stress* scale plus GWBQ *sleeping difficulties* item	Had difficulty in making decisions	.838	0.81 (0.901)	.895
	Had difficulty with remembering	.825	0.80 (1.003)	
	Found it difficult to think clearly	.824	0.92 (0.989)	
	Had problems concentrating	.766	1.09 (1.029)	
	Had difficulty falling or staying asleep	.536	1.14 (1.122)	
COPSOQ *Burnout* scale	Been physically exhausted	.890	1.59 (1.163)	.915
	Felt worn out	.872	1.66 (1.181)	
	Felt tired	.761	2.09 (1.104)	
	Been emotionally exhausted	.743	1.60 (1.199)	
COPSOQ *Stress* scale	Been irritable	.819	1.33 (1.185)	.900
	Been tense	.728	1.41 (1.170)	
	Had problems relaxing	.694	1.49 (1.167)	

Criterion validity. As discussed above in relation to the APHIRM discomfort/pain measure, it would be helpful to investigate criterion validity in relation to workplace records of mental health-related incidents or claims records, since such evidence would be seen as highly relevant by workplace managers. Previous research on the criterion validity of the COPSOQ scales used in this APHIRM scale have demonstrated validity in terms of this kind of evidence: for example, the stress scale was found to be predictive of long-term sickness absence (Pejtersen et al, 2014). Measures of burnout have also been found to predict sickness absence (e.g. Peterson et al, 2011).

CONCLUSION

There is now ample evidence that the APHIRM Toolkit Hazard and Risk Assessment Survey is valid for use in managing risks of work-related musculoskeletal disorders and other stress-related physical and mental health problems, including the detrimental effects of stress on workers' performance capacities.

REFERENCES

AERA, APA, and NCME: American Educational Research Association, American Psychological Association, and National Council for Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC.

Berthelsen H, Westerlund H, Bergström G, Burr H. (2020). Validation of the Copenhagen psychosocial questionnaire version III and establishment of benchmarks for psychosocial risk management in Sweden. *International Journal of Environmental Research and Public Health*. 17(9):3179.

Bodin, J., Garlant´ezec, R., Costet, N., et al. (2020). Shoulder pain among male industrial workers: validation of a conceptual model in two independent French working populations. *Appl. Ergon*. 85, 103075.

Burr, H., Berthelsen, H., Moncada, S., Nubling, M., Dupret, E., Demiral, Y., Oukyk, HJ., Kristensen, T.S. et al. (2019). The third version of the Copenhagen Psychosocial Questionnaire, *Safety and Health at Work* 10, 482-503.

Candan, S.A., Sahin, U.K., Akoglu, S. (2019). The investigation of work-related musculoskeletal disorders among female workers in a hazelnut factory: prevalence, working posture, work-related and psychosocial factors. *Int. J. Ind. Ergon*. 74, 102838.

Chen, W-L., Wang, C-C., Chiang, S-T., Wang, Y-C., Sun, Y-S., Wu, W-T., & Liou, S-H. (2018). The impact of occupational psychological hazards and metabolic syndrome on the 8-year risk of cardiovascular diseases – A longitudinal study. PLoS ONE 13(8): e0202977.

Christensen JO, Nielsen MB, Finne LB, Knardahl S. (2018). Comprehensive profiles of psychological and social work factors as predictors of site-specific and multi-site pain. *Scandinavian Journal of Work, Environment & Health*. 44(3):291-302.

Clarke, S., Cooper, C., (2004). A risk management approach to occupational stress. In: Clarke, S., Cooper, C. (Eds.), *Managing the Risk of Workplace Stress*. Routledge, London, pp. 102–118.

Coenen P, Healy GN, Winkler EA, Dunstan DW, Owen N, Moodie M, et al. (2018). Associations of office workers' objectively assessed occupational sitting, standing and stepping time with musculoskeletal symptoms. *Ergonomics*. 61(9):1187-95.

Cox, T., & Griffiths, A. (2005). The nature and measurement of work stress: theory and practice. In J. R. Wilson, and E. N. Corlett (Eds), *Evaluation of Human Work: A practical ergonomics methodology* (3rd ed., pp. 553-571). Taylor and Francis: London.

Cox, T., Griffiths, A. & Rial-Gonzalez, E. (2000). *Research on Work-related Stress.* European Agency of Safety & Health at Work, Luxembourg.

Cox, T., Oliver, A., Rial-González, E., Tomás, J.M., Griffiths, A.,& Thompson, L. (2006). The development of a Spanish language version of the Worn Out scale of the General Well-Being Questionnaire (GWBQ). *The Spanish Journal of Psychology,* 9:1, 94-102.

Dianat, I., Afshari, D., Sarmasti, N., et al. (2020). Work posture, working conditions and musculoskeletal outcomes in agricultural workers. *Int. J. Ind. Ergon*. 77, 102941.

Eatough E, Way J, Chang C. (2012). Understanding the link between psychosocial work stressors and work-related musculoskeletal complaints. *Appl Ergon.* 43(3):554–63.

Edu-Valsania, S., Laguia, A. & Mioriano, J.A. (2022). Burnout: A review of theory and measurement. *International Journal of Environmental Research and Public Health.* 19, 1780.

Graveling, R., Smith, A., Hanson, M. (2021). *Musculoskeletal Disorders: Association with Psychosocial Risk Factors at Work*. European Agency for Safety and Health at Work, Luxembourg.

Griffiths A, Cox T, Karanika M, Khan S, Tomas JM. (2006). Work design and management in the manufacturing sector: development and validation of the Work Organisation Assessment Questionnaire. *Occupational & Environmental Medicine*. 63(10):669 –675.

Hagberg M, Violante FS, Bonfiglioni R, Descatha A, Gold J, Evanoff B, et al. (2012). Prevention of musculoskeletal disorders in workers: classification and health surveillance. Statement of the scientific committee on musculoskeletal disorders of the International Commission on Occupational Health. *BMC Musculoskelet Disord*. 13(1):109.

Harvey, S.B., Modini, M., Joyce, S., Milligan-Saville, J.S., Tan, L., Mykletun, A., Bryant, R.A., Christensen, H. & Mitchell, P.B. (2017). Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occup Environ Med*. 74:301–310.

Hauke A, et al. (2001). The impact of work-related psychosocial stressors on the onset of musculoskeletal disorders in specific body regions: a review and meta-analysis of 54 longitudinal studies. *Work Stress*. 25(3):243–56.

Knetka, E., Runyon, C., Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE—Life Sciences Education*. 18:rm1, 1–17, Spring 2019.

Kop J, Althaus V, Formet-Robert N, Grosjean V. (2016). Systematic comparative content analysis of 17 psychosocial work environment questionnaires using a new taxonomy. *Int J Occup Environ Health*. 22(2):128–41.

Lang J, Ochsmann E, Krasus T, Lang JW. (2012). Psychosocial work stressors as antecedents of musculoskeletal problems: a systematic review and meta-analysis of stability-adjusted longitudinal studies. *Soc Sci Med.* 75(7): 1163–74.

Leka, S. & Jain, A. (2010). *Health Impact of Psychosocial Hazards at Work: An Overview.* World Health Organization, Geneva. https://apps.who.int/iris/handle/10665/44428

Liu, H.-C., Cheng, Y., Ho, J.-J. (2020). Associations of ergonomic and psychosocial work hazards with musculoskeletal disorders of specific body parts: a study of general employees in Taiwan. *Int. J. Ind. Ergon*. 76, 102935.

Macdonald, W.A. 2000. Workload, fatigue and job satisfaction in repetitive blue-collar work. In *Human System Interaction: Education, Research and Application in the 21st Century*, edited de Waard et al, Shaker Publishing BV, Maastricht, The Netherlands.

Macdonald, W. 2003a. Workload and Work Stress. In *Work Stress: Studies of the context, content and outcomes of stress. A Book of Readings.* Chapter 6 (pp.113-138). Baywood: New York.

Macdonald, W. 2003b. Work demands and stress in repetitive 'blue-collar ' work. In Work Stress: Studies of the context, content and outcomes of stress. A Book of Readings. Chapter 7 (pp.139-163). Baywood: New York.
Macdonald W. (2012). Conceptual framework for development of a toolkit for prevention of work-related musculoskeletal disorders. *Work*. 41(Suppl 1):3933–6.
Macdonald, W., & Evans, O. (2006). *Research on the prevention of work-related musculoskeletal disorders Stage 1 - literature review.* Canberra: Australian Safety and Compensation Council. https://www.safeworkaustralia.gov.au/system/files/documents/1702/research_prevention_workrelated_musculoskeletal_disorders_stage_1_literature_review.pdf
Macdonald W, Evans O, Armstrong R. (2008). Research on the prevention of musculoskeletal disorders – Stage 2. A study of a small sample of workplaces in high-risk industries. Department of Employment and Workplace Relations, Project No. CIR 1001439–07.
Macdonald W, Oakman J. (2022). The problem with "ergonomics injuries": What can ergonomists do? *Applied Ergonomics*. 103:103774.
Marras W, Cutlip R, Burt S, Waters T. (2009). National occupational research agenda (NORA) future directions in occupational musculoskeletal disorder health research. *Appl Ergon.* 40(1):15–22.
Metzler YA, Bellingrath S. (2017). Psychosocial hazard analysis in a heterogeneous workforce: determinants of work stress in blue-and white-collar workers of the European steel industry. *Frontiers in Public Health*. 5:210.
Meyers DC, Durlak JA, Wandersman A. (2012). The quality implementation framework: a synthesis of critical steps in the implementation process. *Am J Community Psychol.* 50(3–4):462–80.
National Research Council (US), Institute of Medicine (US) (2001). Panel on Musculoskeletal Disorders and the Workplace. *Musculoskeletal disorders and the workplace: low back and upper extremities*. Washington, DC: National Academies Press.
Neupane, S., Leino-Arjas, P., Nygård, C.-H., et al. (2017). Developmental pathways of multisite musculoskeletal pain: what is the influence of physical and psychosocial working conditions? *Occup. Environ. Med.* 74 (7), 468–475.
Norton, S., Ellis, B., Suarez, B.S., Schwank, S., Fitzpatrick, R., Price, A. & Galloway, J. (2019). Validation of the Musculoskeletal Health Questionnaire in inflammatory arthritis: a psychometric evaluation. *Rheumatology. 58:45-51.*
Oakman, J., 2014. Investigation of MSD toolkit risk and hazard measures in relation to claim rates and other indicators. Institute for Safety, Compensation and Recovery Research. https://research.iscrr.com.au/prevention/workplace-injury-and-illness /work-related-musculoskeletal-disorders.
Oakman, J., Chan, S. (2015). Risk management: where should we target strategies to reduce work-related musculoskeletal disorders? *Saf. Sci.* 73, 99–105.
Oakman, J., Macdonald, W., Wells, Y. (2014). Developing a comprehensive approach to risk management of musculoskeletal disorders in non-nursing health care sector employees. *Appl. Ergon*. 45 (6), 1634–1640.
Pejtersen, J.H., Kristensen, T.S., Borg, V., & Bjorner, J.B. (2014). The second version of theCopenhagen Psychosocial Questionnaire. *Scandinavian Journal of Public Health,* 38(Suppl 3): 8-24.
Peterson, U., Bergstrom, G., Demerouti, E., Gustavsson, P., Asberg, M., & Nygren, A. (2011). Burnout levels and self-rated health prospectively predict future long-term sickness absence. *Journal of Occupational & Environmental Medicine,* 53:7, 788-793.
Punnett, L. & Wegman, D.H. (2004). Work-related musculoskeletal disorders: the epidemiological evidence and the debate. *Journal of Electromyography and Kinesiology.* 14, 13-23.
Roquelaure, Y., Ha, C., Leclerc, A., Touranchet, A., Sauteron, M., Melchior, M., Imbernon, E., & Goldberg, M. (2006). Epidemiological surveillance of upper-extremity musculoskeletal disorders in the working population. *Arthritis & Rheumatism (Arthritis Care & Research).* 55:5, 765-778.
Safe Work Australia (2013). *How to determine what is reasonably practicable to meet a health and safety duty*. guide_reasonably_practicable.pdf (safeworkaustralia.gov.au)
Stanton, N.A., (2016). On the reliability and validity of, and training in, ergonomics methods: a challenge revisited. *Theor. Issues Ergon. Sci.* 17 (4), 345–353.
Sullivan, G.M. (2011). A Primer on the Validity of Assessment Instruments. *J Grad Med Educ.* 3 (2): 119–120.
Taibi, Y., Metzler, Y.A., Bellingrath, S., et al. (2021). A systematic overview on the risk effects of psychosocial work characteristics on musculoskeletal disorders, absenteeism, and workplace accidents. *Appl. Ergon*. 95, 103434.
Takala, E.-P., Pehkonen, I., Forsman, M., et al., 2010. Systematic evaluation of observational methods assessing biomechanical exposures at work. *Scand. J. Work. Environ. Health* 36 (1), 3–24.

Tang, K.H.D., 2020. A review of psychosocial models for the development of musculoskeletal disorders and common psychosocial instruments. *Archives of Current Research International.* 20 (7), 9–19.

Widanarko, B., Legg, S., Devereux, J., et al. (2015). Interaction between physical and psychosocial risk factors on the presence of neck/shoulder symptoms and its consequences. *Ergonomics.* 58 (9), 1507–1518.

World Health Organization (2020). *Occupational health: Stress at the workplace.* https://www.who.int/news-room/questions-and-answers/item/ccupational-health-stress-at-the-workplace#:~:text=Work%2Drelated%20stress%20is%20the,challenge%20their%20ability%20to%20cope

Yang, F.M. and Kao, S.T. (2014) Item Response Theory for Measurement Validity. *Shanghai Archives of Psychiatry*, 26, 171-177.