WORK DISABILITY AMONG INJURED WORKERS WITH CHRONIC UPPER-EXTREMITY DISORDERS: MEASUREMENT AND DETERMINANTS

by

Kenneth Tang

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto

© Copyright by Kenneth Tang (2014)
WORK DISABILITY AMONG INJURED WORKERS WITH CHRONIC UPPER-EXTREMITY DISORDERS: MEASUREMENT AND DETERMINANTS

Kenneth Tang
Doctor of Philosophy (2014)
Institute of Health Policy, Management and Evaluation
Dalla Lana School of Public Health
University of Toronto

ABSTRACT

Work disability represents an important source of the burden for individuals recovering from a work-related upper-extremity disorder (WRUED). Historically, a small subset of injured workers has shown to experience a prolonged recovery course characterized by poor health and work outcomes, which are associated with high workers’ compensation costs. Beyond achieving an initial return-to-work, the sustainability of work (i.e. staying at work) and health-related work limitations (i.e. “on-the-job” problems) are increasingly recognized as relevant concerns in this population. In this thesis, three quantitative studies were conducted to contribute insights to the measurement and determinants of work disability associated with WRUEDs, using survey data gathered from injured workers attending a Workplace Safety & Insurance Board Upper-Limb Specialty Clinic in Ontario. Specifically, these studies investigated 1) the factorial validity of the Work Limitations Questionnaire (WLQ-25), 2) the predictive ability of the Upper-Limb Work Instability Scale (UL-WIS) for subsequent transitioning out of work (i.e. poor sustainability of work), and 3) the relationship between the work environment and health-related work limitations, among
injured workers with a WRUED. Study findings suggest that: 1) due to a specific area of psychometric underperformance, some adaptations are needed for the WLQ-25 before it is suitable for use; 2) the UL-WIS is an independent predictor of transitioning out of work, although some limitations as a standalone prognostic tool were also revealed; and 3) against initial hypothesis, broad differences in the work environment (based on variations across 12 specific workplace characteristics) did not have significant associations with the extent of health-related work limitations experienced by injured workers with a WRUED. Study findings were discussed in terms of how they advance current understanding on the measurement and determinants of work disability, and potential implications in terms of the assessment and management of work disability among injured workers recovering from a WRUED. Future research directions were also proposed to build on current work.
ACKNOWLEDGEMENTS

First and foremost, I wish to express my sincere gratitude to my thesis supervisor, Dr. Dorcas Beaton. I am forever indebted for the tremendous guidance, unwavering support, and endless patience you have provided since day one. Over the years you have given me opportunities far beyond any student can ask, and I know that any career success I might have in the future will be because of the excellent mentorship you have provided. For that and so much more, I can’t say thank you enough.

I also wish to sincerely thank other members of my thesis committee: Drs. Sheilah Hogg-Johnson, Pierre Côté, Patrick Loisel, and Benjamin Amick III, for consistently providing insightful and constructive feedback to my thesis work, and for providing encouragement and support throughout the process. Like Dr. Beaton, you have all been an integral part of my professional development and for that I am extremely grateful.

I wish to express my appreciation to the staff and colleagues at St. Michael’s Hospital, the Institute of Work & Health, the WSIB Shoulder and Elbow clinic at the Holland Orthopaedic & Arthritic Centre, and also, collaborators from OMERACT and the I-CAN-WORK Alliance. There are too many of you to name, but you have all been tremendously helpful and supportive to me. Each of you has inspired me in your own unique ways and I have also learned so much from all of you. Thank you very much.

Special thanks to my thesis examiners Drs. Marie-José Durand and Paul Demers for their time and effort to review my thesis, and for sharing valuable insights and feedback on my work.

Lastly, I wish to acknowledge funding support provided by the Canadian Institutes of Health Research, Institute for Work & Health, WSIB Shoulder & Elbow Specialty clinic at the Holland Orthopaedic & Arthritic Centre, Canadian Arthritis Network, and Institute of Health Policy, Management and Evaluation at the University of Toronto. I also wish to acknowledge research funding provided by the WSIB RAC (grant #05028) and the WSIB
Shoulder and Elbow Specialty clinic. Without such generous support this thesis would not be possible and for that I am truly grateful.
TABLE OF CONTENTS

ABSTRACT .......................................................................................................................... II

ACKNOWLEDGEMENTS ...................................................................................................... IV

TABLE OF CONTENTS ....................................................................................................... VI

LIST OF TABLES ................................................................................................................ IX

LIST OF FIGURES ............................................................................................................. X

LIST OF APPENDICES ..................................................................................................... XI

ABBREVIATIONS .............................................................................................................. XII

CHAPTER 1: GENERAL INTRODUCTION ................................................................. 1

1.1 EPIDEMIOLOGY OF WORK-RELATED UPPER-LIMB DISORDERS ....................... 2

1.1.1 Origins and typical presentation ........................................................................... 2

1.1.2 Classification .......................................................................................................... 3

1.1.3 Burden .................................................................................................................... 4

1.1.3.1 Prevalence and incidence ................................................................................. 4

1.1.3.2 Workers’ compensation costs ......................................................................... 6

1.1.3.3 A small but costly subgroup .......................................................................... 7

1.1.3.4 Cost to injured workers .................................................................................. 8

1.1.3.5 Upper-Extremity Specialty clinics in Ontario .................................................... 8

1.2 WORK DISABILITY FOLLOWING OCCUPATIONAL INJURIES ...................... 9

1.2.1 Overview and definitions ...................................................................................... 9

1.2.2 Productivity loss: an indirect cost of occupational injuries .................................. 11

1.2.3 Work disability associated with WRUEDs ......................................................... 12

1.3 AREAS IN NEED OF RESEARCH ...................................................................... 14

1.3.1 Theme #1: Measurement of work disability ...................................................... 14

1.3.1.1 Emergence of measures of health-related work limitations .......................... 14

1.3.1.2 Limited psychometric evidence in WRUEDs ................................................. 15
1.3.2 Theme #2: Determinants of work disability ................................................................. 16
  1.3.2.1 Prognostic and contextual factors .............................................................................. 17
  1.3.2.2 Multifactorial nature of work disability ...................................................................... 17
  1.3.2.3 Evidence for WRUEDs (heterogeneous conditions) ............................................... 18
  1.3.2.4 Evidence for hand disorders or carpal tunnel syndrome ........................................ 20
  1.3.2.5 Methodological quality of studies ............................................................................ 21
  1.3.2.6 Gaps in current understanding ............................................................................... 22

1.4 AIMS AND OBJECTIVES OF PROPOSED RESEARCH .................................................. 24

CHAPTER 2: CONFIRMATORY FACTOR ANALYSIS OF THE WORK LIMITATIONS QUESTIONNAIRE (WLQ-25) WORKERS’ COMPENSATION CLAIMANTS WITH CHRONIC UPPER-LIMB DISORDERS (OBJECTIVE #1) 25

2.1 ABSTRACT .......................................................................................................................... 26
2.2 INTRODUCTION .................................................................................................................. 27
2.3 METHODS .......................................................................................................................... 30
2.4 RESULTS ............................................................................................................................... 35
2.5 DISCUSSION ........................................................................................................................ 38
2.6 TABLES AND FIGURES ........................................................................................................ 42

CHAPTER 3: PREDICTIVE ABILITY OF THE UPPER-LIMB WORK INSTABILITY SCALE FOR TRANSITIONING OUT OF WORK AMONG INJURED WORKERS WITH CHRONIC UPPER-EXTREMITY DISORDERS (OBJECTIVE #2) ................................................................. 47

3.1 PREFACE ............................................................................................................................. 48
3.2 ABSTRACT ............................................................................................................................ 49
3.3 INTRODUCTION .................................................................................................................... 50
3.4 METHODS ............................................................................................................................ 52
3.5 RESULTS ............................................................................................................................... 58
3.6 DISCUSSION ........................................................................................................................ 61
3.7 TABLES AND FIGURES ........................................................................................................ 64

CHAPTER 4: HEALTH-RELATED WORK LIMITATIONS FOLLOWING WORK-RELATED UPPER-EXTREMITY DISORDERS: DOES THE WORK ENVIRONMENT PLAY A ROLE? (OBJECTIVE #3) ................................................................. 70

4.1 PREFACE ............................................................................................................................. 71
**LIST OF TABLES**

Table 2.1   Demographic, health, and work-related characteristics of study participants \( (n=2262) \)

Table 2.2   Univariate statistics of WLQ-25 items \( (n=2262) \)

Table 2.3   Model fit statistics for nested WLQ-25 and post-hoc WLQ\textsuperscript{PD} factor structures (MI subscale intact vs. expanded)

Table 3.2   Item-level distribution of the UL-WIS

Table 3.3   Predictive ability of UL-WIS on subsequent transitioning out of work assessed by multivariable logistic regression

Table 4.1   Socio-demographic, health, and work-related characteristics of the study sample \( (n=349) \)

Table 4.2   Model fit statistics estimated from latent profile analyses of 12 organizational- or job-level factors (multiple solutions compared)

Table 4.3   Work characteristics stratified by broad work environment profiles \( (n=4) \) derived from latent profile analysis

Table 4.4   Multiple-group path analysis examining the main and moderation effects of work environment profiles
LIST OF FIGURES

Figure 2.1  Standardized parameter estimates from confirmatory factor analyses of nested WLQ-25 models: 4-factor model with MI subscale intact (top) and 5-factor model with MI subscale divided into MD and IP subscales (bottom)

Figure 3.1  Distribution of baseline UL-WIS scores among injured workers attending a Workplace Safety & Insurance Board Upper-Limb Specialty clinic (n=309)

Figure 3.2  Distribution of work status over 12 months for injured workers who were working at time of initial Upper-Limb Specialty clinic attendance (n=356)

Figure 3.3  Positive predictive values and negative predictive values across the range of UL-WIS score cut-offs. The highest positive predictive value (UL=WIS>80.8) and negative predictive value (UL-WIS<41.9) were identified as potential scale cut-scores

Figure 4.1  Hypothesis of “worker-workplace interactions” to be investigated in the current study

Figure 4.2  Four-profile solution derived from latent profile analysis of 12 organizational- or job-level factors
LIST OF APPENDICES

Appendix 1  Measures of Work Disability and Productivity: Rheumatoid Arthritis specific
Work Productivity Survey (WPS-RA), Workplace Activity Limitations Scale (WALS), Work Instability Scale for Rheumatoid Arthritis (RA-WIS), Work Limitations Questionnaire (WLQ), and Work Productivity and Activity Impairment Questionnaire (WPAI)

Appendix 2  Documentation of research ethics approval from the University of Toronto
## ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLS</td>
<td>Bureau of Labor Statistics</td>
</tr>
<tr>
<td>CTS</td>
<td>Carpal tunnel syndrome</td>
</tr>
<tr>
<td>DASH-W</td>
<td>Disabilities of the Arm, Shoulder, and Hand Work Module</td>
</tr>
<tr>
<td>ICD</td>
<td>International Classification of Diseases</td>
</tr>
<tr>
<td>LPA</td>
<td>Latent profile analysis</td>
</tr>
<tr>
<td>MI</td>
<td>Mental-interpersonal (subscale of WLQ)</td>
</tr>
<tr>
<td>OD</td>
<td>Output demands (subscale of WLQ)</td>
</tr>
<tr>
<td>PD</td>
<td>Physical demands (subscale of WLQ)</td>
</tr>
<tr>
<td>P-E fit</td>
<td>Person-environment fit</td>
</tr>
<tr>
<td>RA-WIS</td>
<td>Work Instability Scale for Rheumatoid Arthritis</td>
</tr>
<tr>
<td>RTW</td>
<td>Return(ed)-to-Work</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SPS</td>
<td>Stanford Presenteeism Scale</td>
</tr>
<tr>
<td>TM</td>
<td>Time management (subscale of WLQ)</td>
</tr>
<tr>
<td>UL-WIS</td>
<td>Upper-Limb Work Instability Scale</td>
</tr>
<tr>
<td>WI</td>
<td>Work instability</td>
</tr>
<tr>
<td>WRF</td>
<td>Work Role Functioning</td>
</tr>
<tr>
<td>WSIB</td>
<td>Workplace Safety &amp; Insurance Board</td>
</tr>
<tr>
<td>WLQ</td>
<td>Work Limitations Questionnaire</td>
</tr>
<tr>
<td>WRUED(s)</td>
<td>Work-related Upper-Extremity Disorder(s)</td>
</tr>
</tbody>
</table>
CHAPTER 1: GENERAL INTRODUCTION
“A return to work does not necessarily mark the end of the spells of work disability associated with a chronic condition” (Baldwin and Butler, 2006)

1.1 EPIDEMIOLOGY OF WORK-RELATED UPPER-LIMB DISORDERS

1.1.1 Origins and typical presentation

Upper-extremity disorders encompass a variety of conditions that affect the neck, shoulder, arm, wrist or hand regions, due to impairments to the peripheral neuro-musculoskeletal system. These conditions are typically characterized by pain, discomfort, or tingling in localized or diffuse upper-extremity regions (Lemasters et al., 1998). In addition to impaired sensorimotor function, other symptoms like tissue swelling and muscle fatigue are also common (Lemasters et al., 1998; Rempel et al., 1992). As a result, patients often complain of an inability to perform activities of daily living and work-related activities, as well as sleep disruptions (Lemasters et al., 1998; Rempel et al., 1992; Shaw et al., 2002). Since many jobs involve tasks that tax the upper-extremity, many upper-extremity disorders have an occupational origin (Descatha et al., 2004; Latko et al., 1999; Leclerc et al., 2004; Leclerc et al., 2001; Punnett and Wegman, 2004). Work-related upper-extremity disorders (WRUEDs) can be of acute onset (e.g. result from a discrete injurious event), but more typically tend to be of gradual onset (e.g. repetitive strain injury) (Bureau of Labor Statistics, 2013; Workplace Safety and Insurance Board of Ontario, 2012). Often, WRUEDs are caused or precipitated by a combination of workers applying excessive force, poor hand gripping techniques, adopting awkward postures, repeated strenuous movement, and/or insufficient rest or recovery following exertions. WRUEDs are now recognized as one of the most common types of occupational disorders, and studies have shown that WRUEDs are common in a variety of occupational groups including manual labourers (Latko et al., 1999), computer users, and performance artists like musicians (Pascarelli and Hsu, 2001). Following initial onset, residual effects and recurrence of condition are ongoing concerns which can significantly affect work functioning (Baldwin and Butler, 2006; Keogh et al., 2000; Pransky et al., 2002).
1.1.2 Classification

WRUEDs are heterogeneous in terms of affected body part(s) and clinical pathology (Stiens et al., 1996). As such, valid classification and differentiation of WRUEDs can contribute to more specific understanding and evaluations of occupational disorders. Historically, however, this has proven to be challenging as a universal diagnostic criteria or classification system for WRUEDs remains elusive (Buckle and Devereux, 1999; Katz et al., 2000; Van Eerd et al., 2003; World Health Organization, 1985). A related issue is that highly varied terminologies have been used to describe WRUEDs with 14 different umbrella terms identified in a recent review (Boocock et al., 2009). Available classification systems for WRUEDs are equally diverse provoking ongoing discussions on this topic (Boocock et al., 2009; Buchbinder et al., 1996; Davis, 1998; Harrington et al., 1998; Huisstede et al., 2007; Katz et al., 2000; Palmer et al., 2000; Van Eerd et al., 2003). In a 2003 review, Van Eerd et al. identified a total of 27 classification systems and identified 44 unique WRUEDs conditions. Interestingly, no two of the identified systems have an identical set of conditions and only 14 of the 44 disorders were included in five or more systems. One common suggestion among scholars is the need to distinguish between “specific” and “non-specific” conditions (see below). Beyond these two major categories, some have further argued for an additional category of rare conditions where considerable controversies remain regarding diagnosis and work-relatedness (Boocock et al., 2009; Sluiter et al., 2001; Van Eerd et al., 2003).

Among WRUEDs, “specific” conditions are generally associated with a well-defined set of diagnostic criteria established from evidence-based approaches (Health Council of The Netherlands, 2000; Sluiter et al., 2001; Tyrer, 1999). However, existing classification systems are not in complete agreement about which conditions to qualify into this group or how they should be organized within this group. For example, while Piligian et al. identified seven specific conditions of the distal upper-extremity based on prevalence or “strong evidence for work-relatedness in the literature” (Piligian et al., 2000), Sluiter et al. identified 11 specific conditions based on a literature review (Sluiter et al., 2001). Meanwhile, instead of specific conditions, Helliwell et al. proposed seven diagnostic sub groups of soft-tissue disorders of the arm (Helliwell et al., 2003). In the most recently published classification of WRUEDs, Boocock et al. proposed five broad groupings for 14 different conditions based on a systematic literature review (Boocock et al., 2009). The groupings proposed were: 1) tendon-
related conditions, 2) nerve-related conditions, 3) circulatory/vascular conditions, 4) joint-related conditions, 5) painful syndromes.

In contrast, “non-specific” conditions can be thought of as less-well-defined WRUEDs often without a clear pattern of signs and symptoms, and therefore, a specific diagnosis is elusive. Symptoms may include diffuse aching, weakness, and/or muscle tenderness. From a Delphi consensus, Harrington et al. defined these as conditions “not linked to any anatomical structure, with diagnosis being confirmed by the exclusion of specific signs and symptoms” (Harrington et al., 1998). Although the relevance of such a group is generally acknowledged, different terms have been previously used to capture conditions in this category as a whole. These include, for example, “non-specific diffuse forearm pain” (Harrington et al., 1998), “non-specific upper limb disorder” (Helliwell et al., 2003), and “non-specific upper extremity musculoskeletal disorders” (Sluiter et al., 2001).

The International Classification of Diseases (a standard diagnostic tool for epidemiology, health management and clinical purposes, endorsed by the World Health Organization) is sometimes applied in research as the basis for identifying WRUED claims from a workers’ compensation caseload. Typically, qualifying conditions include mononeuropathies (e.g. carpal tunnel syndrome, CTS [354.0]), enthesopathies (e.g. lateral epicondylitis [726.32]), tendon disorders (e.g. tenosynovitis [727.05]), soft tissue disorders (e.g. myalgia and myositis [729.1]), nerve root and plexus disorders (353), cervical disorders (e.g. cervicalgia [723.1]), osteoarthrosis (715), disorders of the muscle, ligament, and fascia (728), and peripheral vascular diseases (443) (Feuerstein et al., 1998).

1.1.3 Burden

1.1.3.1 Prevalence and incidence

To date, highly-varying estimates of the prevalence of upper-extremity disorders have been reported, suggesting a degree of regional- and population-specificity (e.g. general population vs. specific occupational group) in prevalence. This is borne out by a recent systematic review that synthesized published estimates of the prevalence and incidence of upper-extremity disorders (Huisstede et al., 2006). Among 13 studies deemed to be of sufficient methodological quality (data collected between 1983 and 1998, seven different countries
represented), the point prevalence of upper-extremity disorders ranged between 1.6-53.0% (Huisstede et al., 2006). None of these studies, however, reported incidence rates, and therefore such insights remain limited at present. Another important challenge noted was that unique case definitions of upper-extremity disorders had been applied across the studies, and therefore, available data could be not pooled. Moreover, many case definitions do not clearly differentiate between upper-extremity disorders of occupational origin and those that are not. Reviews of the prevalence of specific types of upper-extremity disorder, such as shoulder pain or CTS, are also available, and likewise, a wide range of estimates has been found. For example, the point prevalence of shoulder pain in the general population among different countries worldwide was estimated to be 7-27% (Luime et al., 2004). The prevalence of CTS was estimated to be 1-61% among various occupational groups (Hagberg et al., 1992).

Recent reports from the Bureau of Labor Statistics (BLS) suggest that WRUEDs remain common among US workers. A division of the US Department of Labor, the BLS publishes prevalence and incidence rates for various occupational injuries and illnesses on a national level. Its latest annual statistical report suggests that, in 2012, WRUEDs (inclusive of shoulder, arm, wrist, and hand injuries, but not neck) accounted for 30% (n=347590 of 1153980) of all nonfatal occupational injuries requiring days away from work (Bureau of Labor Statistics, 2013). The incidence rate of WRUEDs was estimated to be 33.8 cases per 10,000 full time workers, which was highest among all occupational injuries including those affecting the back (21.2 cases/10,000 full-time workers) and trunk (27.9 cases/10,000 full-time workers). Among all WRUEDs cases, the median number of days away from work was 10. From this perspective, shoulder injuries were actually the most disabling of all occupational injuries, with a median of 24 days away from work (Bureau of Labor Statistics, 2013).

Another perspective on the prevalence of WRUEDs can be gained from research using workers’ compensation claims data (i.e. administrative data). On the low end, some studies have suggested that WRUEDs account for <5% of workers’ compensation caseloads (Brogmus et al., 1996; Fabrizio, 2002; Feuerstein et al., 1998; Hashemi et al., 1998). However, other reports have suggested a much higher proportion. For example, claims data
between 1999-2004 from the Ohio Bureau of Workers’ Compensation found WRUEDs (combining shoulder, elbow, and hand/wrist cases) accounted for 38.5% of all claims ($n=220173$ of $572508$). Hand/wrist injuries, in particular, represented a fairly high proportion (20.6%) of all claims, second only to lumbar spine cases (Dunning et al., 2010). In Ontario, WRUEDs annually accounted for 21.7-25.2% of all lost-time claims from 2003-2012 (“sprain and strains” most common), according to reports from the Workplace Safety and Insurance Board (WSIB) (Workplace Safety and Insurance Board of Ontario, 2014). Interestingly, WRUED cases have actually shown a slight declining trend over recent years. In 2004, WRUEDs constituted 24.7% (18,678 of 75,537 new cases) of all lost time claims in Ontario, but this proportion has actually dropped to 22.3% (9,244 of 41,508 new cases) in 2013. A general limitation with claims data is that they may underestimate WRUEDs as not all work-related injuries result in a claim. In two studies of Connecticut workers, it was found that workers’ compensation claims were filed for <10% of work-related musculoskeletal disorders identified through employee interviews (Morse et al., 2005; Morse et al., 1998). Similarly, among newspaper workers in Ontario, only a small subset of workers with pain and symptoms reported to the workplace or filed a claim (Polanyi et al., 1997). And of course, it should also be recognized that not all workers are covered by workers’ compensation insurance (Leigh et al., 1997).

#### 1.1.3.2 Workers’ compensation costs

WRUEDs are responsible for substantial lost work time, as well as medical and indemnity costs (Baldwin and Butler, 2006; Cheadle et al., 1994; Courtney and Webster, 1999; Feuerstein et al., 1998; Tate, 1992). Based on estimates from previous U.S. or Canadian studies (using claims data from 1989-1998), Baldwin and Butler suggested that the mean cost per WRUED claim was likely between $5000-8000 (Baldwin and Butler, 2006). More recently published estimates of the workers’ compensation costs for WRUEDs are generally inline with this figure. For example, based on 1999-2004 claims data from the Ohio Bureau of Workers’ Compensation, the average cost in US dollars per claim (median, SD) for injuries to the shoulder, elbow, and hand/wrist were $6668 (639, 17478), $1971 (329, 7544), and $2081 (320, 8068), respectively (Dunning et al., 2010). Moreover, the percentage of claims with >7 lost work days for these specific body regions were 30.6%, 15.6%, and 16.0%, respectively. Considering the total workers’ compensation (medical and indemnity)
costs, shoulder injury represented the 2nd most costly condition at $457 million (i.e. average cost per claim x number of claims), behind lumbar spine injury ($1.4 billion total, average cost per claim = $8750). The particularly high cost of occupational injuries to the shoulder was reinforced in a recent analysis of 1998-2008 claims data from the Injured Workers’ Insurance Fund – a large workers’ compensation insurer in Maryland (Mroz et al., 2014). In this study, it was estimated that the average cost of shoulder claims (in 2010 US$, adjusted for inflation) was $8530 (SD = 25596), which actually exceeded that of back claims $7535 (SD = 35804) (Mroz et al., 2014).

### 1.1.3.3 A small but costly subgroup

It is generally known that a majority of workers' compensation claims are in fact quickly resolved and involve little or no time off work, and therefore, minimal medical and/or indemnity costs. Baldwin and Butler estimated that >80% claims involve only medical costs with no indemnity benefits payments (Baldwin and Butler, 2006). Further indication of this comes from reports showing a median of zero days absent from work among all WRUED cases (Feuerstein et al., 1998; Hashemi et al., 1998). On the flip side, it has been suggested that a small but non-trivial proportion of injured workers is responsible for much of the burden (i.e. disability days, costs), including a small proportion (<5%) with very low probability of returning to work following initial injury (Baldwin and Butler, 2006).

Converging reports support the existence of a small but costly group of injured workers with WRUEDs (Cheadle et al., 1994; Feuerstein et al., 1998; Hashemi et al., 1998; Webster and Snook, 1994). In Webster and Snook, the median cost per claim was in fact only $824 (mean = $8070), with the most costly 25% of claims accounting for 89% of the costs (Webster and Snook, 1994). Cheadle et al. found that, among Washington State workers’ compensation claims for carpal tunnel syndrome (CTS) from 1987-1989, 18% were absent from work for at least six months, 12% were absent at least 12 months, and 7% were absent at least 2 years (Cheadle et al., 1994). In a study of 1994 claims from a single large US workers’ compensation insurer, 7% of the claims were responsible for 60% of the costs and 75% of the total disability days (Hashemi et al., 1998). For lost time claims accepted by the US Department of Labor, Office of Workers’ Compensation Programs from 1993-1994, 18% of CTS claims accounted for 53% of the indemnity costs of all such cases, while 23%
of enthesopathy of the elbow claims accounted for 67% of the indemnity costs of all such cases (Feuerstein et al., 1998).

1.1.3.4 Cost to injured workers

Beyond the costs to workers’ compensation systems, WRUEDs are also costly to the injured workers themselves, although this perspective has not been well-characterized to-date. One important concern that has been raised is the potential financial hardship due to under-compensation from lengthy and recurring spells of work absence. Keogh et al. found that following a WRUED claim, 16% of families could not afford to maintain their car, 32% borrowed money from friends or family, and 36% had been contacted by a collection agency (Keogh et al., 2000). Similar concerns were reported in a study of Connecticut injured workers, which found that persons with WRUEDs are significantly more likely to report that they have moved for financial reasons, lost a home or car, or lost health insurance, compared to a control group without WRUEDs (Morse et al., 1998).

1.1.3.5 Upper-Extremity Specialty clinics in Ontario

The Workplace Safety & Insurance Board (WSIB) of Ontario is the provincial workers’ compensation board and is an organization mandated under the Workplace Safety and Insurance Act (legislated by the provincial government) to promote health and safety at workplaces in Ontario. The WSIB provides no-fault liability insurance for employers, as well as health care and loss of earnings benefits for employees. Recognizing the high burden and complexity of managing injured workers, in 1998, the WSIB initiated the Specialty Clinics Program which established several Upper-Extremity Specialty clinics at academic health sciences centers across the province to provide specialized clinical assessments for claimants with chronic or complex WRUED cases (Workplace Safety and Insurance Board, 2012). As such, these cases reflect the subset of all WRUED claims that are particularly costly and burdensome to various stakeholders, and thus most in need of research attention. At these clinics, consultations are provided by a multidisciplinary team of rehabilitation therapists, orthopedic surgeons, social workers, and case managers to evaluate recovery progress and prognosis, suitability for work, or candidacy for surgery or rehabilitative (e.g. vocational) interventions. A key goal at these clinics is to help injured workers achieve a sustainable RTW, and/or an improved level of work functioning. Currently, ten WSIB Upper-Extremity Specialty clinics
exist in Ontario, which is more than the number of clinics dedicated to any other type of occupational injury – another indication of the high burden of WRUEDs.

### Key Points:

- **Work-related upper-extremity disorders (WRUEDs) are responsible for substantial cost to the workers’ compensation system. An average cost of between $5000-8000 per claim has been estimated based on a collection of US and Canadian studies.**

- **Converging reports suggest that a small proportion of injured workers is responsible for much of the burden (disability days, costs) associated with WRUEDs. A focus on reducing the burden of this subset may be especially important.**

- **Recognizing the high burden and potential complexity of WRUEDs, Upper-Extremity Specialty clinics have been established by the WSIB of Ontario to help manage claimants with more chronic and complex cases of WRUEDs**

### 1.2 WORK DISABILITY FOLLOWING OCCUPATIONAL INJURIES

#### 1.2.1 Overview and definitions

Employment is considered a valued role in society. Beyond contributing to financial independence, work is also important for cultivating an “occupational identity” (Kielhofner, 2002). Moreover, a productive and satisfying work life contributes to psychosocial well-being (Waddell and Burton, 2006). For individuals with health concerns, work may even have added importance as it can help provide feelings of normality, a sense of stability and belonging, individual self-esteem and worth (Saunders and Nedelec, 2014). As such, work disability represents a significant source of the burden associated with occupational injuries. Historically, “work disability” simply referred to work absences due to health reasons.
Chapter 1: General Introduction

However, over the years, a more expansive view of this concept is being adopted, as it is now known that various types of employment-related changes may occur as a result of health. Accordingly, there are different indicators of work disability. Certainly, being unable to work (i.e. not working, amount of lost work time over time) remains one of the most obvious indicators. In order to work, some may require some form of job accommodations (e.g. modified duties or reduced work hours), representing a change in work status. For others, a transition to a different job may be necessary to accommodate health concerns. The common thread among these work-related changes is that they represent an inability to participate at a pre-injury level. From the perspective of the International Classification of Functioning, Disability, and Health (ICF) framework, this is generally regarded as an impact at the participation level, since the individual is unable to fulfill his or her usual societal role due to health reasons. This of course, can be transient, since a person may transition between different work statuses/states over time. Accordingly, the “sustainability” of work is emerging as an issue of interest particularly for workers with chronic or episodic health conditions (e.g. rheumatoid arthritis). For example, recurrent episodes of work disability are not uncommon following occupational back disorders (Abenhaim et al., 1988; Carey et al., 1999; Rossignol et al., 1992; Watson et al., 1998).

For those able to work, “on-the-job” problems due to health reasons represent another form of work disability, yet this issue has only gained attention more recently. In the literature, this is sometimes referred to as “at-work disability”, “presenteeism” or “health-related work limitations”. In this thesis, the term “health-related work limitations” will be primarily used. Broadly, health-related work limitations can be defined as the inability of a worker to meet specific demands within his or her work role due to health reasons (e.g. difficulties performing work tasks). From an ICF perspective, this mainly pertains to impact at the activities level, although this can sometimes blend with impact at the body structure and function level (e.g. having pain while performing work tasks). Historically, health-related work limitations have been somewhat of a hidden problem (Hemp, 2004); however, it is now increasingly recognized that initial RTW following an occupational injury does not necessarily imply successful functioning at work (Burton et al., 2005; Burton et al., 2004; Butler et al., 1995). Beyond impacts on a person’s quality-of-(work)-life, health-related work limitations may also foreshadow future problems, such as the need to go off-work later on
Chapter 1: General Introduction

(i.e. poor sustainability of work). Accordingly, measures of health-related work limitations have been increasingly applied as an outcome (i.e. study endpoint) for observational studies or clinical trials involving musculoskeletal disorders (see (Tang et al., 2013c) for a review). Some measures may also have merit as prognosticators of future work states (e.g. predicting work loss) (Gilworth et al., 2007; Gilworth et al., 2006; Gilworth et al., 2003; Gilworth et al., 2009a).

1.2.2 Productivity loss: an indirect cost of occupational injuries

Productivity loss due to health reasons is generally considered a source of indirect health costs. Since a productive workforce is desirable, indirect health costs have particular relevance from the perspectives of employers and society-at-large (i.e. impact on overall economy). In health economic research, a broad distinction is commonly made between productivity loss/costs attributed to missed work time due to health reasons (“absenteeism”), and decreased productivity while at work due to health reasons (“presenteeism”) (Brouwer et al., 1999; Hemp, 2004). Naturally, occupational injuries can contribute to both sources of productivity loss, with health-related work limitations a driver of the latter. The cost of presenteeism is by no means trivial, and a growing number of reports have shown that for more chronic or episodic health conditions (e.g. arthritis, migraine, depression), the cost of presenteeism can exceed that of absenteeism (Bramley et al., 2002; Goetzel et al., 2004; Hemp, 2004; Li et al., 2006; Ozminkowski et al., 2004) and of medical care (Cockburn et al., 1999). For example, the American Productivity Audit – which involves a random telephone survey of 28,902 US workers – estimated that lost productivity due to health costs employers a staggering $225.8 billion per year ($1685 per employee per year), of which 71% is attributed to presenteeism (Stewart et al., 2003). Similarly, in a study of 8267 US employees at 47 employer locations, the total productivity costs associated with 15 common health conditions were estimated to be $2962 per employee per year, with presenteeism accounting for 69% of those costs ($2049 per year) (Lamb et al., 2006). Bank One (now JP Morgan Chase) estimated presenteeism to be as much as 84% of their total lost productivity costs, with absenteeism and disability comprising the other 16% (Hemp, 2004). Clearly, health-related work limitations represent an aspect of work disability that deserves attention.
1.2.3 Work disability associated with WRUEDs

As previously discussed, there is ample evidence from workers’ compensation claims data to suggest that prolonged work absence is common in WRUEDs, particularly among the most severe cases (Baldwin and Butler, 2006; Cheadle et al., 1994; Feuerstein et al., 1998; Hashemi et al., 1998). Moreover, there is emerging evidence to suggest that, for those able to work (or have RTW) following initial injury, sustainability of work is an issue of concern (Baldwin and Butler, 2006; Bultmann et al., 2007; Keogh et al., 2000; Roquelaure et al., 2004). Baldwin and Butler found that among permanently impaired workers’ compensation claimants in Ontario with various work-related musculoskeletal disorders, 25% of workers with WRUEDs had experienced a second spell of work absence, which was greater than the proportions for either back pain (18%) or fractures (12%) (Baldwin and Butler, 2006). Parallel to this finding, a more recent study of lost-time claimants in Ontario with either back pain or a WRUED found that among those who had achieved RTW at study baseline (approximately 1-month post-injury), 38% reported being off-work at 6-month follow-up (Bultmann et al., 2007).

While health-related work limitations represent another potential concern for injured workers recovering from a WRUED, documentation of them remains relatively scarce at present. Two US studies provided some insights on this topic. In the Connecticut Upper Extremity Surveillance Project, 35% of workers with a WRUED (n=102 of 374) reported having to cut down on their work pace, although only 4% reported having to go on disability (Morse et al., 1998). Among Maryland workers recovering from a WRUED (1-4 years post-onset), over 50% of workers reported that their symptoms continued to interfere with work (Keogh et al., 2000). More recently, efforts have been made to quantify the extent of health-related work limitations among injured workers attending WSIB Upper-Extremity Specialty clinics in Ontario. In a sample of 80 Specialty clinic attendees, moderate health-related work limitations was found based on assessments using four different self-report multi-item measures (Tang et al., 2009). These measures include the Disabilities of the Arm, Shoulder, and Hand Work Module (DASH-W), Work Limitations Questionnaire 16 (WLQ-16), Stanford Presenteeism Scale 6 (SPS-6), and Work Instability Scale for Rheumatoid Arthritis (RA-WIS). In particular, scores from the WLQ-16 subscales suggested that meeting time management demands and physical demands at work were especially problematic in this cohort (TM=47.4/100, PD = 36.3/100: 0=more limitations). In another study that
recruited from the same population, Pichora et al. examined 460 attendees of a WSIB Specialty clinic in Kingston (southeastern Ontario) with WRUEDs (47% with hand/finger injuries), and assessed health-related “work instability”, defined as the extent of mismatch between work ability and demands. Based on RA-WIS scores, they concluded that 25% and 15% of the sample were at “medium” (score >10/23) and “high” risk (score >17/23) for subsequent work loss, respectively (Pichora and Grant, 2010).

Key Points:

- **Work disability is a major contributor of the cost and burden of occupational injuries.** Health-related work limitations and poor sustainability of work are increasingly recognized as part of the concern.

- **Recent evidence suggests that both health-related work-limitations and poor sustainability of work are relevant concerns among injured workers recovering from a WRUED.**
1.3 AREAS IN NEED OF RESEARCH

Since work disability represents an important source of the burden associated with WRUEDs, reducing work disability is an important goal in this population. To this end, research on two broad themes can contribute insights towards this goal. These are: 1) the measurement of work disability, and 2) the determinants of work disability. Background, relevance, and existing knowledge on these themes are discussed here.

1.3.1 Theme #1: Measurement of work disability

1.3.1.1 Emergence of measures of health-related work limitations

Advancements in the measurement of work disability are essential from two different perspectives. First, in any scientific discipline, the ability to measure pertinent variables is of fundamental importance, paving the way for ongoing research investigations. In work disability research, work disability naturally represents a key variable of interest, and therefore, the ability to measure this concept is pivotal to any progress made in the field. Second, measurement of work disability is also necessary for real-world applications, such as disability case management and clinical assessments (i.e. as a patient-oriented outcome). In the literature, common outcomes include indicators (or proxies) that aim to differentiate between work states at a single point in time (e.g. working, working with job accommodations, not working), or capture the extent of work disability from a longitudinal perspective, such as disability duration (e.g. time to initial RTW, total lost time) or compensation duration (e.g. length of episode of wage replacement benefits). With emerging interest in health-related work limitations across different disciplines (e.g. occupational disability, rehabilitation medicine, health economics), instruments that aim to capture this aspect of work disability has seen a significant upsurge in recent years (Amick III et al., 2000b; Beaton et al., 2009; Escorpizo et al., 2007; Loeppke et al., 2003; Lofland et al., 2004; Mattke et al., 2007; Prasad et al., 2004; Tang et al., 2011a). A recent review by Beaton et al. found 24 instruments featuring a component on health-related work limitations (or presenteeism) that have potential applicability in arthritis and musculoskeletal disorders (Beaton et al., 2009). These measures are highlighted by substantial diversity in conceptualization and instrumentation approaches. Some are primarily designed as patient-oriented outcomes and are best suited for descriptive (i.e. assessing the burden) or evaluative
purposes (i.e. assessing changes over time), some are prognostic tools designed to predict future adverse employment-related events (e.g. RA-WIS, UL-WIS), while others are primarily designed to estimate and monetize productivity loss to facilitate health economic evaluations. A number of them may have potential for multiple applications (e.g. WLQ).

1.3.1.2 Limited psychometric evidence in WRUEDs

Currently, it is unclear which measures of health-related work limitations are most suitable for use in chronic/complex WRUEDs, as data on their psychometric properties (e.g. reliability, validity, responsiveness) remain relatively scarce. Psychometric performance can be context-dependent, and therefore, it is important to affirm psychometric properties in specific populations/settings of interest (e.g. health condition, phase of disease/injury), particularly where there may be propensity for some differential item functioning (i.e. measures consisting of items that may have more/less intrinsic relevance to a given population/setting) (Tang, 2011). For WRUEDs, supporting evidence on internal consistency, content validity, and/or construct validity have been demonstrated for the DASH-W (Tang et al., 2009), RA-WIS (Roy et al., 2011; Tang et al., 2009), SPS-6 (Tang et al., 2009), WLQ-16 (Beaton and Kennedy, 2005; Tang et al., 2009), as well as WLQ-25 (Roy et al., 2011). However, in the broader context, this represents only a relatively small proportion of available instruments, as it is likely that a number of others are also appropriate for use in this population. A head-to-head comparison of four measures (DASH-W, SPS-6, RA-WIS, and WLQ-16) in this population found WLQ-16 had the best overall psychometric performance while the RA-WIS was most preferred among injured workers recovering from a WRUED (Tang et al., 2009). As such, these two measures may be particularly promising for use in this population, despite the fact that they are conceptualized from slightly different perspectives. The family of WLQ measures is designed to assess limitations related to specific demands in four aspects of work: time management demands, physical demands, mental interpersonal demands, and output demands (Lerner et al., 2001). In contrast, the RA-WIS quantifies work instability (WI), which is defined by the developers as “a state of mismatch between an individual’s functional abilities and the demands of his or her job”. Importantly, it is believed that a state of high WI can threaten continuing employment if not resolved (Gilworth et al., 2003). To further improve the RA-WIS for use in WRUEDs, modifications to this measure were made so that it met the stringent
psychometric requirements of the Rasch measurement model (i.e. meets assumption for interval-level scaling). Ultimately, this led to a new 17-item Upper-Limb Work Instability Scale (UL-WIS) (Tang et al., 2011b). Overall, empirical testing of measures of health-related work limitations in WRUEDs is still in its relative infancy, as psychometric evidence remains incomplete although several of the tools have shown some initial promise.

**Key Points:**

- *As a key outcome of interest in occupational disability research, the ability to measure work disability well is of fundamental importance.*

- *Psychometric testing of measures of health-related work limitations in WRUEDs remains limited to-date and is a topic requiring further research attention*

1.3.2 Theme #2: Determinants of work disability

An understanding of the determinants of work disability can provide important insights towards effective management of injured workers. On this topic, there are two main components to consider: 1) the identification of prognostic or contextual factors for different aspects of work disability, and 2) an understanding of the process (or mechanism) by which such factors contribute to or influence the extent of work disability. A classic example of this is an “interaction” effect, where the relationship between a factor and an outcome (e.g. work disability) is influenced by the level of another factor. Research in this area is potentially important for several reasons: it can 1) contribute to a fundamental understanding of the phenomenon of work disability itself, 2) inform intervention targets that should be prioritized during management of injured workers, and 3) provide the foundation for developing measurement tools that have applicability in clinical settings, for example, prognostic tools to identify injured workers at high risk for continuing/worsening work disability (i.e. triaging) and/or risk remediation purposes.
1.3.2.1 Prognostic and contextual factors

Prognostic and contextual factors can perhaps be differentiated in the following manner. Generally, a prognostic factor refers to a measurable variable that predicts the chances of an outcome occurring in the future (e.g. course of disease, quality-of-life at follow-up), and can be verified through empirical testing. In other words, a variable would be considered a prognostic factor if it can be shown to contribute to an estimate of the probability of a future event. Contextual factors, on the other hand, are somewhat less concretely defined. These factors represent a key component of the ICF framework and can be viewed as personal or environmental variables that contribute to the basis or understanding of the resulting disease impact, such as the extent of functional limitations or participation restrictions (World Health Organization, 2001). Importantly, the nature of the relationship between specific contextual factors and disease impacts (e.g. associative, predictive, interactions) is not explicitly defined in the ICF. As such, empirical testing is needed, which may require different designs (e.g. cross-sectional or longitudinal studies) depending on hypothesized “mechanism of action”, such as the expected time lag of influence and/or postulated interactions with other factors. Accordingly, it is possible for the same variable to meet the definitions of both a prognostic and contextual factor (for a given outcome), as these two types of factors are not mutually exclusive by definition. It should also be recognized, however, that neither case necessarily implies a cause-and-effect relationship.

1.3.2.2 Multifactorial nature of work disability

To understand work disability, a perspective beyond the biomedical model is needed. Previous epidemiological studies have shown that musculoskeletal pain is able to explain only 5% of work disability (Waddell and Burton, 2006), and that clinical variables often do not represent the strongest determinants of work disability (Turner et al., 2000; Waddell et al., 2003). Consistent with theories of disablement that emphasize the role of the social environment (Lawrence and Jette, 1996; Verbrugge and Jette, 1994), the relevance of psychosocial, workplace, and societal factors in work disability is increasingly recognized. As such, the bio-psychosocial model (Waddell 1992) or person-environment frameworks (Loisel et al., 2001; Sandqvist and Henriksson, 2004; Tang et al., 2011c) represent more useful paradigms to understand work disability. Conceptualized from the perspective of work disability prevention, a framework by Loisel and colleagues suggested four broad systems
make up the “arena” of work disability: personal system, workplace system, health care system, and legislative and insurance system (Loisel et al., 2005; Loisel et al., 2001). In parallel, others have similarly posited the relevance of both person and environmental factors in the RTW process following initial injury (Feuerstein et al., 2003; Krause et al., 2001b; Schultz et al., 2007), and for understanding individual work functioning (Sandqvist and Henriksson, 2004). Broadly speaking, examples of personal factors include sociodemographics, biomedical health factors (i.e. clinical pathophysiology and symptomatology of the underlying health condition, such as injury type or bodily pain), as well as psychosocial health factors. Postulated psychosocial health factors have origins from diverse research disciplines, which include for example, attributes related to predisposing personality (e.g. motivation, vitality, mental health), cognition (e.g. beliefs, perceptions, expectations, locus of control and self-efficacy), and individual adaptation (e.g. coping beliefs, coping styles, and support from peers, family) (Schultz et al., 2004). Certainly, both biomedical and psychosocial health factors can be influenced by ongoing healthcare interventions. In terms of workplace factors, Krause et al. proposed a pragmatic organization that broadly distinguished between “individual task-level physical and psychosocial job characteristics” and “organizational-level employer factors” (Krause et al., 2001b), and then further divided the first category into physical job characteristics (e.g. physical task demands and postural demands), psychosocial job characteristics (e.g. job strain, job control, psychological demands), and workplace social support (e.g. co-worker or supervisor support). Key domains of organizational-level employer factors that have been theorized to include people-oriented culture/climate, proactive in-house RTW program (disability management), active safety leadership (safety climate), and ergonomic job design practices (Amick III et al., 2000a; Habeck et al., 1998; Habeck et al., 1991). Beyond specific job- and organizational-level factors, the jurisdictional healthcare and insurance system and other broader societal-level influences such as the legislative, sociopolitical, and macroeconomic climate can also play a role (Feuerstein et al., 2003; Krause et al., 2001b; Turner et al., 2000).

### 1.3.2.3 Evidence for WRUEDs (heterogeneous conditions)

Identifying prognostic or contextual factors of work disability in specific populations is worthwhile since potential disease-specificity of relationships is possible (Krause et al., 2001b; Krause and Ragland, 1994; Lotters and Burdorff, 2006; Sinclair et al., 1998). To-date
Chapter 1: General Introduction

however, such evidence has largely focused on back pain. Although the prognosis of WRUEDs is a topic of some research attention, surprisingly few studies to-date have actually considered work disability outcomes. Moreover, such studies are characterized by considerable heterogeneity, including source of data, type(s) of WRUEDs considered, phase of injury (e.g. acute vs. chronic conditions), range of factors and outcomes assessed, operationalization of these variables, and overall study design (e.g. cross-sectional vs. longitudinal design). To-date, diverse prognostic or contextual factors for work disability have been revealed across individual studies. However, despite a shared interest on this general topic, direct comparability of results is somewhat challenging given important differences in research objectives, methodology, and context between studies. To the best of my knowledge, efforts to formally synthesize collective findings on this topic have also not yet emerged in the literature.

A number of studies focused on the prognostication of work disability. In the Survey of Ontario Workers with Permanent Impairments, Baldwin and Butler found that, following an initial WRUED, older workers and those working in physically demanding industries were less likely to RTW, whereas those with union membership or longer job tenure (≥5 years with their time-to-injury employer) were more likely to RTW (Baldwin and Butler, 2006). Using patient data from a large Dutch occupational health registry, Lotters and Burdorf found that prolonged sickness absence (disability duration) due to WRUEDs (including neck disorders) was predicted by being female, higher physical workload, having visited a medical specialist, poorer perceptions of general health, and having a chronic health complaint in the previous 12 months (Lotters and Burdorf, 2006). Among patients attending a Michigan clinic with a diagnosis of upper-extremity disorder, the prognostic value of a concept called “workstyle”, defined as the cognitive and behavioral response to perceived increases in work demands (Nicholas et al., 2005), was specifically assessed. In this study, a higher workstyle score (i.e. more positive response) was revealed as the only significant predictor of being able to perform unrestricted work (vs. restricted work) at 6-month follow-up. Non-significant factors in their prediction model included pain, grip strength, diagnosis, and job category (Harrington et al., 2009).
A few cross-sectional studies have been conducted to reveal factors associated with more “favorable” work state or less health-related work limitations. While some of these factors would meet the ICF definition of a contextual factor, most authors have not necessarily labelled them as such. In a study of Maryland compensated workers with a WRUED, it was found that a CTS diagnosis (vs. other types of WRUEDs), having an activities of daily living score within the normal range, and being a union member at time of injury were associated with being employed (vs. not employed) at time of survey (which was conducted 1-4 years post claim) (Keogh et al., 2000). Feuerstein et al. surveyed a group of US federal government employees who were unable to resume their normal work after filing a workers’ compensation claim for a WRUED, and found that compared with those working modified duties, those who were not working were more likely to report a diagnosis of mononeuropathy (vs. enthesopathy), higher pain ratings, greater functional limitations, and more ergonomic stressors (Feuerstein et al., 2003).

1.3.2.4 Evidence for hand disorders or carpal tunnel syndrome

In contrast to the above studies, some research focused only on a specific subset of WRUEDs, most commonly, hand disorders and CTS. In a study of patients from The Netherlands who were operatively treated for a hand disorder and were employed prior to surgery, a delay in RTW was predicted by greater pain, accident being sustained at work (vs. sustained elsewhere), and greater symptoms of post-traumatic stress disorders (Opsteegh et al., 2009). Among workers with CTS claims from Washington state, the strongest predictors of chronic work disability (defined as >180 days of wage replacement compensation for temporary total disability in the first year after claim submission) were found to be older age, greater functional limitations, not being offered job accommodation, greater physical job demands, and lower recovery expectations (Turner et al., 2007). Insights on prognostic factors for work disability were also gained from the Maine Carpal Tunnel Study, which followed community-based patients following carpal tunnel release surgery. In the first set of reports, Katz et al. found being off-work six months following carpal tunnel release surgery was predicted by worse baseline CTS symptom severity, functional status, and mental health, and receiving workers’ compensation benefits (Katz et al., 1997). A re-evaluation at the completion of the study found that being off-work 18 months following surgery was predicted by greater baseline functional limitations and poorer mental health status (Katz et
Chapter 1: General Introduction

A second study, “Work and Carpal Tunnel in Maine”, was initiated with the aim to increase awareness on health-related work limitations among patients with CTS. Several reports were based on this study. First, Amick III et al. found that more successful work role functioning (WRF) six months following carpal tunnel release surgery was most strongly predicted by baseline WRF, improved self-efficacy between baseline and 2-month follow-up, and supportive organizational policies and practices (Amick III et al., 2004). Uniquely, the authors had defined WRF as a three-level ordinal variable that combined information on work status and health-related work limitations (based on the 15-item Work Role Functioning Questionnaire). The levels were: 1) returned to work functioning successfully [able to meet job demands at least 90% of time], 2) returned to work functioning with limitations [able to meet job demands <90% of time], and 3) not returned to work for health reasons. Second, Katz et al. focused on identifying predictors of work absence following carpal tunnel release surgery. Independent predictors of work absence at 6 months included having more than two painful sites, involvement of an attorney, having the same or worse self-efficacy at 2 month follow-up as compared to baseline, and lower income (Katz et al., 2005). Interestingly, different independent predictors of work absence at 12 months emerged and included older age, worse preoperative functional status, and less supportive organizational policies and practices (Katz et al., 2005). Third, in an analysis guided by the Demand-Control model (Karasek et al., 1998), Gimeno et al. found that active work (high demands combined with high control) predicted a less successful work outcome two months after carpal tunnel release surgery, and high job strain (high demands combined with low control) predicted a less successful work outcome six month after surgery (Gimeno et al., 2005). Cumulatively, these findings indeed point to a multifactorial nature of work disability, and also raised the possibility of considerable time-specificity in predictive factors (Amick III et al., 2004; Gimeno et al., 2005; Katz et al., 2005).

1.3.2.5 Methodological quality of studies

Some considerations of the methodological strengths and limitations of the reviewed studies are worthwhile. One common strength is that the multifactorial nature of work disability is consistently recognized, as all studies had considered factors that encompass socio-demographics, health-related (biomedical and psychosocial) and workplace-related variables. In a few studies, an extensive number of highly specific constructs were considered (Amick
Chapter 1: General Introduction

III et al., 2004; Katz et al., 2005; Turner et al., 2007). A second overall strength is that all studies generally provided descriptions of relevant psychometric evidence for the key measures applied. Two main areas of limitations are also worthy of attention. First, some of the data is perhaps slightly dated. For example, the Survey of Ontario Workers with Permanent Impairments was conducted between 1989 and 1990. In Keogh et al., the Maryland workers’ compensation claims that were filed between 1994 and 1996. And lastly, patients were recruited between 1992 and 1993 for the Maine Carpal Tunnel Study. As such, it is unclear to what extent these findings remain relevant, given advancements in the medical management of musculoskeletal disorders, as well as concurrent changes in the workers’ compensation systems and/or other aspects of the macroeconomic environment over time.

Second, the sample size across the studies are generally quite small, as most studies had \( n \leq 200 \) for their main analysis. Along these lines, over-fitting in multivariable regressions is a concern in some cases. Exceptions in this regard were Baldwin and Butler \( (n=879) \), Keogh et al. \( (n=537) \), and Turner et al. \( (n=899) \), which stood out as having fairly large sample sizes.

1.3.2.6 Gaps in current understanding

Although some insights on the determinants of work disability associated with WRUEDs are available, there remain gaps in the literature and areas that require greater understanding. Two major areas stood out. First, issues of particular concern for injured workers who are able to work have been given limited attention. Sustainability of work is one such issue, and this is important since recurrence of initial injury is common among WRUEDs and an initial RTW does not necessarily mark the end of the recovery process (Baldwin and Butler, 2006; Butler et al., 1995; Dasinger et al., 1999; Krause et al., 1999). To that end, research to identify prognostic or contextual factors linked to the ability or inability to remain in a working role is needed. Another issue, of course, is health-related work limitations. With few exceptions (e.g. (Amick III et al., 2004; Gimeno et al., 2005)), health-related work limitations have not been widely considered or incorporated as a study outcome, despite significant potential relevance for various stakeholders (e.g. quality-of-work-life, productivity loss while at work).

Second, there is a paucity of insights on the interaction between (prognostic or contextual) factors of work disability associated with WRUEDs. Within the arena of work disability, it is proposed that interactions between the four major systems may contribute to the disability
process (Loisel et al., 2001). Similar ideas have also been put forth by others, focusing on worker-workplace interactions. For example, Sandqvist and Henriksson suggested that it is the degree of “fit” between worker and workplace that determines the level of individual work functioning (Sandqvist and Henriksson, 2004). Yet, the nature of such interactions currently remains not well understood, and the formulation of specific and testable hypotheses in this area is also lacking at present. Accordingly, experts in the field have urged scholars to explore interactions in statistical models in work disability research towards further understanding of the basis (or mechanisms) of work disability (Krause et al., 2001b; Turner et al., 2000).

**Key Points:**

- **An understanding of the determinants of work disability can contribute insights towards effective management of injured workers associated with WRUEDs (e.g. identify precise intervention targets, develop prognostic tools)**

- **Important gaps in current understanding include: 1) lack of attention on outcomes that concern those at work, such as health-related work limitations or the sustainability of work, and 2) the nature and relevance of worker-workplace interactions as a contributing mechanism of work disability**
1.4 AIMS AND OBJECTIVES OF PROPOSED RESEARCH

The overall aim of this thesis is to contribute understanding on two main themes: 1) the measurement of work disability, and 2) the determinants of work disability associated with WRUEDs. Three research studies, corresponding to Chapters 2, 3 and 4 of this thesis, were conducted with the following main objectives:

**Objective 1:** To assess the factorial validity of the Work Limitations Questionnaire (WLQ-25) when applied to injured workers with WRUEDs, to inform its suitability for use in this population.

**Objective 2:** To examine the predictive ability of the Upper-Limb Work Instability Scale (UL-WIS) for transitioning out of work among injured workers with WRUEDs who are at work.

**Objective 3:** To profile injured workers with WRUEDs by their job context, and to examine the relationship between work environment and extent of health-related work limitations.
CHAPTER 2: Confirmatory factor analysis of the Work Limitations Questionnaire (WLQ-25) workers’ compensation claimants with chronic upper-limb disorders (Objective #1)

A version of this chapter has been published in *J Occup Rehabil* 23(2):228-38 (Tang et al., 2013a), reproduced with permission.
2.1 ABSTRACT

**Purpose:** To examine the factorial validity of the Work Limitations Questionnaire (WLQ-25) among workers’ compensation claimants with chronic upper-limb disorders. **Methods:** Attendees of the WSIB Shoulder and Elbow Specialty clinic in Toronto, Ontario, Canada, completed a survey that included the WLQ-25 (4 subscales: time-management [TM], physical demands [PD], mental-interpersonal [MI], and output demands [OD]). Confirmatory factor analyses (n=2262) were conducted to evaluate and compare alternative 4- and 5-factor WLQ-25 structures (MI subscale intact vs. separated into mental demands [MD] and interpersonal demands [IP] subscales). Model fit indices, saliency of factor loadings, and convergent/divergent validity of latent factors (r=0.4-0.85 expected) were concurrently assessed. **Results:** The 4-factor WLQ-25 showed acceptable model fit after allowing the residuals of a pair of PD items to correlate (CFI=0.924, TLI=0.915, RMSEA=0.057, SRMR=0.054); however, significantly lower-than-expected correlations between the PD factor and all other factors (r=-0.11 – -0.03) were also observed. Model fit for the 5-factor WLQ-25 was even more optimal (CFI=0.934, TLI=0.925, RMSEA=0.053, SRMR=0.051), with MD and IP factors correlating at r=0.83. **Conclusions:** Evidence of factorial validity was demonstrated by the WLQ-25; however, users should be attentive to an instrumentation issue that could be directly related to the psychometric performance of its PD subscale.
Chapter 2: CFA of WLQ-25

2.2 INTRODUCTION

Upper-extremity injuries are common in the workplace, and a significant cause of work disability (Baldwin and Butler, 2006; Fabrizio, 2002; Silverstein et al., 1998). Annually, work-related upper-limb disorders account for approximately 20% of all workers’ compensation claims in Ontario (Workplace Safety & Insurance Board, 2009), and similar incidence rates are found in other Canadian provinces and also in the US (Baldwin and Butler, 2006; Fabrizio, 2002; Silverstein et al., 1998; WorkSafeBC, 2011). In addition to contributing to work absenteeism, residual and re-occurring pain episodes can occur long after initial return-to-work (Baldwin and Butler, 2006; Pransky et al., 2000), which represents an important concern for injured workers. Instruments that measure health-related “at-work” disability are important for providing a view of the impact of occupational injuries beyond what could be captured with more traditional ranges of work outcomes (e.g. return-to-work status, duration of disability benefits). Assessments of specific work impairments are informative when evaluating the recovery progress of individual workers (i.e. capturing more subtle improvements and deteriorations) and also enable precise investigations on the effectiveness of interventions aimed at addressing on-the-job problems. Moreover, such instruments are needed to provide a basis to facilitate better estimations of the costs associated with occupational injuries. The economic relevance of at-work disability has gained significant research attention in recent years, as health-related at-work productivity loss (presenteeism) has shown to be very costly (Burton et al., 2003; Collins et al., 2005; Goetzel et al., 2004; Schultz and Edington, 2007), contributing up to 63% of the total costs of worker illness (compared to 7% for absenteeism) (Hemp, 2004).

The Work Limitations Questionnaire (WLQ-25) (Lerner et al., 2001) is a promising measure to fulfill these needs, as it was designed to have applicability in a wide range of chronic health conditions and occupational sectors. An important strength of the WLQ-25 is that it differentiates impairments over 4 broad work domains: time management (TM, 5 items), which assesses difficulties with handling a job’s time and scheduling demands; physical demands (PD, 6 items), which examines difficulties when performing job tasks that involve bodily strength, movement, endurance, coordination, and flexibility; mental-interpersonal (MI, 9 items), which examines difficulties with cognitively-demanding tasks and on-the-job social interactions;
and output demands (OD, 5 items), which assesses impact on productivity at work (Amick III et al., 2000b; Lerner et al., 2001). The WLQ-25 also has potential applicability for economic costing purposes (Lerner et al., 2003a; Ozminkowski et al., 2004; Rosekind et al., 2010; Zhang et al., 2010) based on its Index score (weighted sum of the 4 subscale scores), which represents percentage productivity loss relative to healthy controls (Lerner et al., 2003b). Cumulative evidence to support the basic psychometric properties of the WLQ-25 has been synthesized in recent reviews (Abma et al., 2012; Prasad et al., 2004; Tang et al., 2011a). For musculoskeletal conditions, previous studies have shown support for its internal consistency, construct validity, and responsiveness among workers with rheumatoid arthritis (Beaton et al., 2010; Lerner et al., 2001; Walker et al., 2005; Wolfe et al., 2005), osteoarthritis (Beaton et al., 2010; Lerner et al., 2002), and upper-limb disorders (Roy et al., 2011). Supporting psychometric evidence is also available for various non-musculoskeletal conditions such as depression and epilepsy (Erickson et al., 2009; Lerner et al., 2001; Sanderson et al., 2007). Moreover, several recent studies have also applied the WLQ-25 as an outcome measure to assess the efficacy of drug or workplace interventions (Erman et al., 2008; Lerner et al., 2012; Pizzi et al., 2009; Reich et al., 2011).

The WLQ-25’s popularity may be attributed to the fact that it is one of the few available measures that assesses the impact of health on a comprehensive set of theoretically-delineated work domains. However, one relevant psychometric property that has received limited testing to-date is the validity of its factor structure, which is an important aspect of the overall construct validity of a multi-dimensional measure. Demonstration of factorial (or structural) validity is important to ensure that the subscale organization of the WLQ-25 reflects initial domain conceptualization. This also serves to verify that each subscale comprises solely of items that inform a common concept (i.e. unidimensional), thus affirming the validity of summing constituent items into a single score. Although slight variations in factor analytic performance may be expected when a measure is applied to different settings or populations, significant variations across different populations might indicate that the ideal organization of a measure’s constructs may be population-dependent. If so, recognition of such population-specificity would be important, as well as any potential adjustments needed (e.g. re-organization or re-weighting of subscales) to optimize the
validity of results when a measure is applied to specific populations. Additional care would also be needed when comparing results across different studies.

Of particular interest for the WLQ-25 is the nature of the MI domain, namely, the possibility of a subdivision. During original development of the scale, mental demands (MD) and interpersonal demands (IP) were initially conceptualized as distinct dimensions, but were later collapsed into the MI subscale when the scale was introduced (Lerner et al., 2001). Yet, several studies using variants of the WLQ-25 have applied MD and IP as separate subscales to differentiate limitations associated with these work domains (Amick III et al., 2004; Amick III et al., 2000b). This raises the question of whether separating MI into distinct MD and IP subscales might offer an even more optimal structural representation of this measure. To these ends, the objectives of this study are to examine the factorial validity of the WLQ-25 when applied to workers’ compensation claimants with chronic upper-limb disorders, and also, to inform which of the alternative domain organizations (MI intact vs. MI expanded) is more optimal in this population.
2.3 METHODS

Patient recruitment
The complexities of managing work-related upper-limb disorders have led to the development of the Workplace Safety & Insurance Board (WSIB) Upper-Extremity Program, which operates a number of Shoulder & Elbow Specialty clinics at major urban centers in Southern Ontario (Workplace Safety and Insurance Board, 2012). These Specialty clinics provide specialized clinical consultations for claimants experiencing an atypical recovery course (i.e. insufficient progress approximately ~6 months after injury onset) from soft tissue injuries (e.g. repetitive strain injuries, acute or cumulative trauma disorders) or uncomplicated fractures (Workplace Safety and Insurance Board, 2012). Clinical assessments are provided collaboratively by a multidisciplinary team of rehabilitation therapists, orthopaedic surgeons, social workers, case managers to evaluate recovery progress and prognosis, suitability for work, or candidacy for surgery or other interventions. At the Specialty clinic in Toronto, newly referred injured workers who are able to understand written English are asked to complete a survey that constitutes part of their routine clinical care (summary of responses provided to clinicians), and to provide consent for future research use of their de-identified responses (ethics approval required for research protocols). The use of survey data collected from 2006-2010 for the current research objective was approved by the research ethics board at the Sunnybrook Health Sciences Centre (host site of clinic), St. Michael’s Hospital, and University of Toronto. To be included in the current analysis, injured workers must have provided written consent for the use of their survey responses in research, and be working at the time of assessment to be eligible to complete the WLQ-25 within the survey. An initial sample of $n=2342$ met these additional inclusion criteria and among them, $n=2262$ (96.6%) attempted the WLQ-25 (all others provided no responses), which constitutes our final study sample.

Data collection
Two modes of survey administration were available to injured workers: traditional paper format ($n=953$, 42.1%, available just ahead of or during initial appointment as survey packages and consent forms were mailed to all workers referred to the Specialty clinic), or using a touch-screen monitor ($n=1309$, 57.9%, also available during initial appointment if
preferred by the worker). Data in this survey include an array of socio-demographic variables, health measures (e.g. Medical Outcomes Study Short Form-36, version 2, SF-36 (Ware and Sherbourne, 1992); pain subscale from the Shoulder Pain and Disability Index, SPADI (Roach et al., 1991); Quick Disabilities of the Arm, Shoulder and Hand outcome measure, QuickDASH (Beaton et al., 2005)), a body diagram for indicating specific upper-limb regions with pain, indicators of work status (e.g. working: yes/no; job accommodations: normal duties/modified duties/reduced hours), and 2 at-work measures: the Upper-Limb Work Instability Scale (UL-WIS) (Tang et al., 2011b), which has been validated to assess the extent of mismatch between functional abilities and job demands, and the WLQ-25.

**WLQ-25**

Scale instructions differed across the 4 WLQ-25 subscales as intended by the original developers. For the TM, MI, and OD subscales, respondents were instructed to indicate the proportion of time with difficulty due to physical or emotional problems in the past two weeks. For the PD subscale, respondents were instructed to indicate the proportion of time without difficulty in the past two weeks. All items were provided with the same 6 response categories: “all of the time”, “most of the time”, “some of the time”, “a little of the time”, “none of the time”, or “does not apply to my job” (N/A), which is treated as missing as per developer recommendation (likewise for item non-response, N/R) (Lerner et al., 2003b). Our two modes of survey administration offered the same item content, but the layout differed. For the touch-screen format, items were provided one at a time with accompanying subscale instructions displayed alongside at all times. For the paper format, we have condensed the WLQ-25 onto 2 total pages with instructions provided once at the beginning of each subscale to reduce the overall length of the survey package. This differed from the copyright version of the WLQ-25 (available from developers) where scale items span 7 pages in large fonts, with instructions repeated for each item in the PD subscale. In our analyses, item scores for the TM, MI, and OD were reverse-coded to achieve a consistent orientation among all scale items (1=least work limitations, 5=most work limitations).
Chapter 2: CFA of WLQ-25

Analysis

Univariate statistics
Descriptive statistics (mean, %) were used to describe the socio-demographic, health, and work-related characteristics of the sample. For each WLQ-25 item, univariate statistics (mean, SD) and frequency of the two sources of “missing” data (N/A and N/R) were assessed. To investigate whether data missingness due to item non-response could be linked to non-ignorable mechanisms, for each WLQ-25 item, polychoric correlations between N/R (yes/no) and the presumed level of the attribute (inferred from the mean of completed items within the associated subscale) were assessed. A high (positive or negative) correlation ($|r|>0.5$) would suggest that data missingness is strongly related to the level of the attribute (low or high work limitations).

Confirmatory factor analysis (CFA)
Since WLQ-25 domains have been previously conceptualized (Amick III et al., 2000b; Lerner et al., 2001), our aim was to apply CFA approaches to examine the performance of pre-specified measurement models. Given that our data was collected from different modes of administration, separate single-group CFAs for paper and touch-screen groups were initially conducted to evaluate the factor analytic performance of the WLQ-25 and to assess whether results differed between the groups. If no significant differences were found, then the data would be pooled for subsequent analyses. For the objective of determine the optimal factor structure for the WLQ-25, additional CFAs were conducted to compare the performance of nested structures: a 4-factor model representing an intact MI subscale (more constrained, less free parameters) versus a 5-factor model with the MI divided into MD and IP subscales (less constrained, more free parameters).

Model estimator and sample size
Full information maximum likelihood parameter estimation with robust standard errors was considered the most appropriate CFA approach, since assumptions of normality were not met for the WLQ-25 at the item- (Shapiro-Wilk test of normality, $p<0.001$) (Shapiro and Wilk, 1965) or scale-levels (Mardia tests of multivariate normality, $p<0.001$) (Mardia, 1980). A sample size of $\geq 10$ per parameter to be estimated is recommended for CFA (Kline, 2005; Streiner and Norman, 1995), which our current cohort exceeds (total $n=2262$: paper, $n=953$;
touch-screen, \( n = 1309 \)) for all proposed analyses. CFAs were conducted using Mplus version 6.0 (Muthén & Muthén, Los Angeles, CA).

**Criteria for model fit and factorial validity**

CFA model fit was assessed by a series of established fit indices based on current guidelines (Bentler, 1990; Hu and Bentler, 1999; Kline, 2005). For each model we examined the overall chi-square (\( \chi^2 \), difference between expected and observed covariance matrices, non-significance=good fit), Comparative Fit Index (CFI) (Bentler, 1990; Iacobucci, 2010), which assesses goodness-of-fit adjusted for model complexity or parsimony, >0.90=adequate fit, >0.95=good fit) (Hu and Bentler, 1999; Kline, 2005), Tucker-Lewis Index (TLI) (Tucker and Lewis, 1973; Widaman and Thompson, 2012), which is based on comparisons between a null model and an incrementally more complex model (>0.90=adequate fit, >0.95=good fit), root mean square error of approximation (RMSEA) (Browne and Cudeck, 1993), which is an absolute measure of fit not considered to be highly sensitive to sample size or distribution (<0.05=close approximate fit, 0.05-0.08=adequate fit, >0.10=poor fit), and standardized root mean square residual (SRMR), which is defined as the standardized difference between the observed and predicted correlation (<0.08=adequate fit, <0.06=good fit) (Hu and Bentler, 1999). Since the chi-square test is known to be highly sensitive to sample size (i.e. null hypothesis generally rejected in large sample size) (Kline, 2005) and our data did not exhibit a normal distribution, evaluations of model fit considered the collective performance of all fit indices assessed. In addition, factor loadings were also assessed for saliency (standardized estimates [\( \beta \)]>0.4) and statistical significance (\( p < 0.05 \) expected). To demonstrate factorial validity, our *a priori* expectation was that the WLQ-25 measurement structure would demonstrate: 1) overall model fit based on the various fit indices and criteria specified above (“adequate” fit at minimum), and 2) salient loading of all items onto their associated factor. In addition, we also expected correlations within 0.40-0.85 among latent factors in the CFA models, to provide evidence of convergent and divergent validity among the WLQ-25 subscales. Since WLQ-25 domains were conceptualized to assess different dimensions of work limitations, it was expected that the 4 associated subscales would be at least moderately inter-related (\( r > 0.40 \), indicating proper convergent validity), but not excessively correlated (\( r < 0.85 \), indicating proper divergent validity) to affirm the lack of
significant redundancy. As such, this inter-factor correlation range was considered a conceptually meaningful and important *a priori* expectation.

**Model modifications**

For all CFA models, model modification indices were inspected to identify unspecified parameters that, if freely estimated, would contribute to decrements in model chi-square (thus improving model fit). For measurement models, such unspecified parameters would represent either items loading onto a different domain than originally hypothesized, or strong correlations among item residuals (i.e. variance unexplained by the common factor). In our analyses, we would only consider performing a modification if there were substantive, underlying conceptual bases.
2.4 RESULTS

Study participants
The study sample ($n=2262$) has a mean age of $46.4\pm 9.0$, with 52.0% female (Table 1). Specific upper-limb site most commonly affected (i.e. have pain) was the shoulder (85.3%) and upper-arm (52.8%). Bilateral upper-limb pain was indicated by 29.4%. The SPADI pain subscale mean was 30.2 (SD=11.5), and QuickDASH mean was 53.2 (SD=20.1), indicating moderate levels of upper-limb pain and functional limitations among claimants in our cohort on average. In terms of job accommodations at initial clinic attendance, 50.8% were on modified duties, 3.5% were working reduced hours, and 20.6% indicated that they were on modified duties as well as working reduced hours. The UL-WIS mean was 10.0 (SD=4.0), indicating moderate work instability.

Univariate statistics of the WLQ-25
Item means varied among the WLQ-25 subscales (Table 2). In the TM subscale, item means ranged from 2.4 to 2.6, while in the PD subscale a higher range (3.1 to 3.3) was found. Wider ranges of item means were found in both the MI (1.6 to 2.5) and OD (2.0 to 2.9) subscales. We differentiated two sources of “missing” data in the WLQ-25: respondents selecting “does not apply to my job” (N/A) or item non-response (N/R). On average, respondents provided 1.6 N/A (median=0) and had 0.8 N/R (median=0) responses. Accordingly, more N/A than N/R was found for 20 of the WLQ-25 items. The highest proportion of N/A was found among 3 items from the PD subscale: walk/move around work locations (13.7%), lift, carry, move objects (11.3%), and use handheld tools, equipment (12.0%), while 2 other items also exceeded 10% (both from MI subscale): speak in person/on phone (12.1%), and help others to work (12.5%). The proportion of N/R ranged from 1.8 to 3.2% among WLQ-25 items. Across the 25 scale items, we found only low-to-moderate polychoric correlations ($r=-0.36 – 0.16$) between N/R and the associated subscale mean (from the completed items), suggesting that data missingness due to item non-response was not strongly related to the level of the attribute in question.
CFA

Single-group CFAs found comparable factor analytic performance for the WLQ-25 between paper and touch-screen groups. During initial model fitting, borderline adequate model fit was evident in both the paper group (CFI=0.905, TLI=0.895, RMSEA=0.065, SRMR=0.060) and the touch-screen group (CFI=0.904, TLI=0.892, RMSEA=0.062, SRMR=0.062). In both groups, factor loadings were salient ($\beta>0.4$) and statistically significant ($p<0.05$) for 23/25 items, except for walk/move around work locations (paper, $\beta=0.09$; touch-screen, $\beta=0.11$) and sit, stand, stay in 1 position (paper, $\beta=0.36$; touch-screen, $\beta=0.36$), both from the PD subscale. Moreover, in both cases, modification indices revealed that the uncorrelated residual variances between these same 2 items were also the most significant source of model misfit (largest model $X^2$ contribution). Allowing their residual variances to correlate in the revised model (i.e. to specify interconnectedness of these 2 items above and beyond their inter-relationship with other subscale items) was considered a justifiable modification since it represented a conceptually-relevant distinction among PD items – these 2 items emphasized mobility/postural-related demands whereas the other 4 PD items emphasized demands that more directly tax the upper-limb(s). Such a conceptual distinction has psychometric implications since the subset of items focused on demands that tax the upper-limb would be expected to have much greater intrinsic relevance in this population. After this modification, the resultant model fit was much improved (paper group: CFA=0.926, TLI=0.917, RMSEA=0.057, SRMR=0.053; touch-screen group: CFA=0.927, TLI=0.919, RMSEA=0.054, SRMR=0.056). In the revised models, correlations among (latent) factors representing the TM, MI, and OD subscales ranged between 0.68 – 0.70 in the paper group, and 0.73 – 0.79 in the touch-screen group. However, significantly lower-than-expected correlations between these 3 factors and the PD factor (range: paper, -0.07 – -0.10; touch-screen, -0.03 – -0.01) were also found, thus the criterion of convergent validity among WLQ-25 domains (inter-factor $r>0.40$ expected) was not completely met in our testing. Overall, we found little difference in all aspects of the factor analytic performance between paper and touch-screen groups, suggesting that the survey administration method did not have a significant influence on CFA findings.
Comparing between 4- and 5-factor WLQ-25 models

Comparisons of nested structures were conducted using a single-group CFA approach pooling data from both modes of survey administration (Table 3). Results found that the 5-factor model showed a moderate improvement in overall model fit (CFI=0.934, TLI=0.925, RMSEA=0.053, SRMR=0.051) compared to the 4-factor model (CFI=0.924, TLI=0.915, RMSEA=0.057, SRMR=0.054). A statistically significant model $X^2$ difference ($200.534$ [df difference=4], $p<0.001$) was also observed between the nested models. This suggests that the item variances/covariances were more fully explained by the expanded WLQ-25 structure. A correlation of 0.83 between the MD and IP factors was also found. In both models, lower-than-expected inter-factor correlations involving only the PD domain remained evident ($r=-0.11$ $-0.03$) (Figure 1).

In light of the performance of the PD domain and its potential implications (see Discussion), we performed additional post-hoc CFAs without the PD subscale to ascertain findings from the comparison of nested models above. Using a modified 19-item WLQ with the 6 PD items removed (henceforth referred to as the WLQ$^{PD}$), CFA models with the MI intact (3 factors) and expanded (4-factors) were re-specified and compared. A re-examination of the dataset of eligible participants identified $n=2252/2342$ (96.2%) that qualified for post-hoc analyses (i.e. contributed data to the WLQ$^{PD}$). During initial fitting, both the 3- and 4-factor WLQ$^{PD}$ models demonstrated acceptable-to-good overall model fit (Table 3), and salient and statistically significant loadings were evident for all 19 scale items. Proper convergent and divergent validity among all WLQ$^{PD}$ factors was also evident (range of inter-factor correlations: 3-factor model=0.72-0.74; 4-factor model=0.67-0.83). Similar to the previous findings, the performance of model fit indices was superior in the parent model (i.e. 4-factor WLQ$^{PD}$) and a statistically significant $X^2$ difference ($183.234$ [df difference=3], $p<0.001$) between nested models was found, affirming that item variances/covariances were more fully explained when separate MD and IP subscales were specified.
2.5 DISCUSSION

The current domain/subscale structure of the WLQ-25 was originally developed by Lerner et al. through psychometric field testing in patients with rheumatoid arthritis, chronic headache syndrome, or epilepsy (Lerner et al., 2001). In this study, we examined whether factor analytic performance of the WLQ-25 in workers’ compensation claimants with upper-limb disorders would demonstrate results that are consistent with the proposed organization, or reveal an alternative subscale organization that may be more optimal for this population. There were several key findings: 1) adequate-to-good model fit was generally evident (after one key modification), 2) lower-than-expected inter-factor correlations pertaining only to the PD domain; 3) the emergence of a subdivision in the PD domain that differentiated two broad types of physical work demands, and finally, 4) improved factor analytic performance when the MI subscale was specified as separate MD and IP factors, offering support for an expanded conceptualization of health-related work limitations (5 domains) based on the measure's content.

The WLQ-25 was designed to capture an overall (higher-order) concept of health-related work limitations and thus a degree of convergence among its domains was expected (i.e. $r=0.40-0.85$). Lower-than-expected inter-factor correlations specifically involving the PD domain were notable for 2 main reasons: 1) at the group-level, one would not expect work limitations related to PD to have no relationship or an inverse relationship with other forms of work limitations, and 2) expected levels of correlations were otherwise consistently observed among the other 3 WLQ-25 domains. Interestingly, considerable variations in the magnitude of correlations for the PD subscale (vs. other subscales and/or WLQ Index) have been reported previously. In some studies, these correlations have been similarly low as in our current results, whereas in other studies they have shown to be substantially higher. In a recent study that applied the WLQ-25 in a similar population, Roy et al. (Roy et al., 2011) found only a 0.06 correlation between the PD subscale and the WLQ Index, whereas other subscales had much higher correlations ($r=0.78-0.89$) against the WLQ Index. In arthritis, the PD also showed low correlations ($r=0.17-0.30$) with the other 3 subscales and only a moderate association ($r=0.49$) with the WLQ Index (Beaton et al., 2010). Interestingly, a large study ($n=836$) examining workers with rheumatoid arthritis reported consistently high
correlations between the PD and other subscales \((r: TM=0.67, MI=0.70, OD=0.67, WLQ-Index=0.83)\) (Walker et al., 2005); however, we were made aware (Dr. Kaleb Michaud, personal communications) that to prevent respondent misinterpretations, instructions for the PD subscale were purposely altered in the study to achieve consistent orientation with the other 3 subscales (i.e. all WLQ-25 items asked about proportion of time \textit{with} difficulty). Psychometric performance for the shorter WLQ-16 (Beaton and Kennedy, 2005; Tang et al., 2009), where the instruction provided is designed to be consistent across all 4 subscales, were also sought to shed further light on this issue. In a recent study, Beaton \textit{et al.} (Beaton and Kennedy, 2005) \((n=42)\) found consistently high correlations between the PD and the TM \((r=0.83)\), MI \((r=0.82)\), and OD \((r=0.62)\) subscales, respectively. These converging reports raise the possibility that the specific orientation of instructions provided for the PD subscale could have important influences on the scale’s psychometric performance. In cases where the PD offered reversed instruction (i.e. asks about time \textit{without} difficulty), is it possible that some respondents may have missed this “flip”, resulting in lower-than-expected inter-subscale correlations? We believe this specific instrumentation issue is noteworthy for users. Given that correlations between the PD and other subscales approximated zero, this further suggests that perhaps only about half of the sample had missed the change in instruction, since a larger negative correlation might be expected had the majority of respondents missed the flip. Since the WLQ-25 was constructed from a solid theoretical basis, and expected levels of correlation among subscales had been achieved elsewhere (i.e. in WLQ-16 or when instructions provided are consistent among WLQ-25 subscales), it is unlikely that the various instances of suboptimal performance for the PD subscale reflect issues at the conceptual level. We reiterate that in our administration of the WLQ-25 via the \textit{paper} format, our specific layout differed from the copyrighted version of the questionnaire (i.e. where instructions was repeated for each single PD item), and thus the possibility exists that perhaps this subtle difference could have been an important contributing factor. That said, CFA results from the \textit{touch-screen} group, where items were presented one-at-a-time with accompanying instructions at all times, offered no discernible improvements in this regard.

Moreover on the PD subscale, CFA results also suggested a segregation of subscale items representing 2 subdivisions of physical work demands (postural/mobility vs. upper-limb specific) that has high conceptual relevance for this population. Dividing the PD into 2
subscales, however, may not be ideal given only 2 items focused on postural/mobility demands in the WLQ-25. Thus, perhaps a more sensible approach for users is to be attentive of whether there are noteworthy score differences between these item subsets during applications of the WLQ-25 to relevant populations such as worker with upper-limb disorders. We do raise caution that when interpreting the factor loadings of PD items in our study, we have made one important assumption. Given the concurrent instrumentation issue (i.e. possibility of response error in the PD), only under the assumption that – if one had missed the “flip” in instruction, one would have responded to all items in the section under the same (erroneous) premise – would the co-variances of PD items remain interpretable (i.e. if scores for all PD items have been reversed [by mistake], inter-item correlations would remain unaffected). We do believe this to be a probable assumption.

Which of the 4- vs. 5-factor structures represent the more optimal WLQ-25 organization among claimants with chronic upper-limb disorders? In truth, to answer this question requires considerations from various competing perspectives. There is, probably at the forefront, 1) considerations of the conceptual meaningfulness and relevance of the alternative domain structures, 2) psychometric (e.g. factorial analytic) performance, more specifically, whether improved model fit with a more complex structure (with more parameters) outweigh the parsimony offered by a simpler structure assuming it also performs well, and lastly, 3) practicality considerations (e.g. whether there are sufficient items to adequately represent each conceptualized domains within a measure). It should be considered that insights from this study contributed mainly to the psychometric aspect of these considerations. While dividing MI into MD and IP led to improved CFA model fit, factors representing the MD and IP domains were also quite highly correlated \( (r=0.83) \), suggesting limited divergence between these 2 subscales. Although applying the MI as a single subscale is the current recommended approach to score the WLQ-25 (and some support for this is also offered in this study), it should also be considered that the ability to differentiate the impact of health on well-conceptualized dimensions of work is an important strength of this measure, and separation into more specific MD and IP domains could further enhance this strength in certain contexts. Where the mental and interpersonal work demands placed on the worker differs substantially (e.g. cognitively demanding jobs that require minimal co-
worker interactions), the opportunity to differentiate health-related difficulties associated with each could be potentially informative.

Limitations in terms of the generalizability of study findings should be considered. The emergence of subdivisions in the PD could be a population-specific effect. Such a subdivision may have less relevance in populations with greater heterogeneity in terms of painful body sites (e.g. arthritis) or non-musculoskeletal conditions (e.g. depression). On the other hand, the observation of lower-than-expected inter-factor correlations involving the PD subscale may be independent of the specific health population under investigation, as this appears to be directly linked to a specific instrumentation issue. We might expect factor analytic performance to improve upon current results if effective strategies to minimize (or perhaps nullify) the risk of response error could be implemented. Finally, the full-information maximum likelihood method applied in our CFA assumes no significant indications of non-ignorable data missingness mechanisms in the dataset, which appeared to be case based on our evaluation.

Based on our study results, we conclude that the WLQ-25 demonstrated evidence of factorial validity when applied to injured workers with chronic upper-limb disorders. Moreover, significantly lower-than-expected inter-factor correlations involving the PD domain was also an important finding that has implications for future users, since we believe this relates to a (correctable) instrumentation issue. To this end, we suggest users to consider approaches that would minimize the potential for response error in the PD subscale. Some options may include: 1) applying a questionnaire layout that closely resembles the developer’s copyright version, 2) emphasize and repeat instruction for every PD item (e.g. using large, bold fonts), or 3) consider interviewer-administration of questionnaire to ensure respondents understood the instructions. Alternatively, opportunities to completely mitigate this issue may also be worthwhile to explore. Users may also consider 4) a re-orientation of the instruction for the PD subscale in order to be consistent with the other subscales, or 5) applying alternative versions of this instrument (e.g. WLQ-16 (Beaton and Kennedy, 2005) or Work Role Functioning (Amick III et al., 2000b)).
## 2.6 TABLES AND FIGURES

**Table 2.1** Demographic, health, and work-related characteristics of study participants $(n=2262)$

<table>
<thead>
<tr>
<th>Attribute</th>
<th>$n$ available</th>
<th>Mean ± SD or %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic characteristics:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>2180</td>
<td>46.4 ± 9.0</td>
</tr>
<tr>
<td>Sex</td>
<td>2178</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>52.0%</td>
</tr>
<tr>
<td>School</td>
<td>2152</td>
<td></td>
</tr>
<tr>
<td>Completed high school or less</td>
<td></td>
<td>49.2%</td>
</tr>
<tr>
<td>Some college/university/technical school</td>
<td></td>
<td>14.2%</td>
</tr>
<tr>
<td>Completed college/university/technical school</td>
<td></td>
<td>36.6%</td>
</tr>
<tr>
<td>Marital status</td>
<td>2169</td>
<td></td>
</tr>
<tr>
<td>Married or living with significant other</td>
<td></td>
<td>73.3%</td>
</tr>
<tr>
<td>Divorced or separated</td>
<td></td>
<td>14.1%</td>
</tr>
<tr>
<td>Widowed</td>
<td></td>
<td>1.7%</td>
</tr>
<tr>
<td>Single (never married)</td>
<td></td>
<td>11.0%</td>
</tr>
<tr>
<td><strong>Health-related variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper-limb region affected (with pain)</td>
<td>2214</td>
<td></td>
</tr>
<tr>
<td>Shoulder</td>
<td></td>
<td>85.3%</td>
</tr>
<tr>
<td>Upper-Arm</td>
<td></td>
<td>52.8%</td>
</tr>
<tr>
<td>Elbow</td>
<td></td>
<td>47.9%</td>
</tr>
<tr>
<td>Forearm</td>
<td></td>
<td>32.1%</td>
</tr>
<tr>
<td>Wrist</td>
<td></td>
<td>28.0%</td>
</tr>
<tr>
<td>Hand</td>
<td></td>
<td>27.6%</td>
</tr>
<tr>
<td>Neck Involvement</td>
<td></td>
<td>43.7%</td>
</tr>
<tr>
<td>Bilateral upper-limb involvement</td>
<td></td>
<td>29.4%</td>
</tr>
<tr>
<td>Pain subscale from the SPADI: (0-50, 50=most pain)</td>
<td>2244</td>
<td>30.2 ± 11.5</td>
</tr>
<tr>
<td>QuickDASH: (0-100, 100=most upper-limb symptoms/functional limitations)</td>
<td>2188</td>
<td>53.2 ± 20.1</td>
</tr>
<tr>
<td>SF-1 (1=excellent, 5=poor)</td>
<td>2201</td>
<td>3.0 ± 1.0</td>
</tr>
<tr>
<td>SF-36 physical components summary score (PCS)</td>
<td>2117</td>
<td>38.7 ± 6.9</td>
</tr>
<tr>
<td>SF-36 mental components summary score (MCS)</td>
<td>2117</td>
<td>42.2 ± 13.3</td>
</tr>
<tr>
<td><strong>Work-related variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work status (job accommodations)</td>
<td>2238</td>
<td></td>
</tr>
<tr>
<td>On normal duties</td>
<td></td>
<td>25.2%</td>
</tr>
<tr>
<td>On light/modified duties</td>
<td></td>
<td>50.8%</td>
</tr>
<tr>
<td>Working reduced hours</td>
<td></td>
<td>3.5%</td>
</tr>
<tr>
<td>On light/modified duties and working reduced hours</td>
<td></td>
<td>20.6%</td>
</tr>
<tr>
<td># of hours worked per week</td>
<td>2221</td>
<td>35.9 ± 10.5</td>
</tr>
<tr>
<td>Perceived exertion at work (0-10 Borg scale, 10=most exertion)</td>
<td>2201</td>
<td>3.3 ± 2.2</td>
</tr>
<tr>
<td>UL-WIS: (0-17, 17=most work instability)</td>
<td>1991</td>
<td>10.0 ± 4.0</td>
</tr>
</tbody>
</table>
SPADI, Shoulder Pain and Disability Index; SF-36, Medical Outcomes Studies Short-Form 36 (version 2); SF-1, “In general, would you say your health is”; QuickDASH, Quick Disabilities of the Arm, Shoulder and Hand outcome measure; UL-WIS, Upper-Limb Work Instability Scale.

* assessed from a body diagram in the survey, multiples regions may apply (not mutually exclusive)
Table 2.2 Univariate statistics of WLQ-25 items ($n=2262$)

<table>
<thead>
<tr>
<th>Item</th>
<th>Label</th>
<th>Mean</th>
<th>SD</th>
<th>% Missing</th>
<th>N/R</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Management (TM)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Work required hours</td>
<td>$T_{req}$</td>
<td>2.5</td>
<td>1.4</td>
<td>2.1</td>
<td>4.8</td>
<td></td>
</tr>
<tr>
<td>Get going beginning of work day</td>
<td>$T_{getgo}$</td>
<td>2.6</td>
<td>1.3</td>
<td>2.4</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>Start on work soon after arriving</td>
<td>$T_{start}$</td>
<td>2.4</td>
<td>1.4</td>
<td>1.8</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>Work without breaks or rests</td>
<td>$T_{breaks}$</td>
<td>2.6</td>
<td>1.3</td>
<td>2.4</td>
<td>6.6</td>
<td></td>
</tr>
<tr>
<td>Stick to routine/schedule</td>
<td>$T_{stick}$</td>
<td>2.6</td>
<td>1.4</td>
<td>2.8</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>Physical Demands (PD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walk/move around work locations</td>
<td>$P_{walk}$</td>
<td>3.1</td>
<td>1.7</td>
<td>3.1</td>
<td>13.7</td>
<td></td>
</tr>
<tr>
<td>Lift, carry, move objects</td>
<td>$P_{lift}$</td>
<td>3.3</td>
<td>1.5</td>
<td>2.7</td>
<td>11.3</td>
<td></td>
</tr>
<tr>
<td>Sit, stand, stay in 1 positions</td>
<td>$P_{sit}$</td>
<td>3.1</td>
<td>1.5</td>
<td>2.8</td>
<td>8.5</td>
<td></td>
</tr>
<tr>
<td>Repeat motions</td>
<td>$P_{repeat}$</td>
<td>3.1</td>
<td>1.4</td>
<td>3.2</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>Bend, twist, or reach</td>
<td>$P_{bend}$</td>
<td>3.2</td>
<td>1.3</td>
<td>2.9</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>Use handheld tools, equipment</td>
<td>$P_{tools}$</td>
<td>3.1</td>
<td>1.3</td>
<td>3.2</td>
<td>12.0</td>
<td></td>
</tr>
<tr>
<td>Mental-Interpersonal (MI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keep mind on work</td>
<td>$M_{mind}$</td>
<td>2.5</td>
<td>1.2</td>
<td>2.2</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>Think clearly</td>
<td>$M_{think}$</td>
<td>2.3</td>
<td>1.3</td>
<td>2.0</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>Work carefully</td>
<td>$M_{care}$</td>
<td>2.2</td>
<td>1.3</td>
<td>3.0</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>Concentrate on work</td>
<td>$M_{concen}$</td>
<td>2.4</td>
<td>1.3</td>
<td>2.4</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Lose train of thought</td>
<td>$M_{losing}$</td>
<td>2.3</td>
<td>1.2</td>
<td>2.7</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>Easily read/use eyes</td>
<td>$M_{read}$</td>
<td>1.8</td>
<td>1.3</td>
<td>2.8</td>
<td>6.9</td>
<td></td>
</tr>
<tr>
<td>Speak in person/on phone</td>
<td>$I_{speak}$</td>
<td>1.6</td>
<td>1.0</td>
<td>2.3</td>
<td>12.1</td>
<td></td>
</tr>
<tr>
<td>Control temper</td>
<td>$I_{temper}$</td>
<td>1.8</td>
<td>1.1</td>
<td>2.6</td>
<td>5.8</td>
<td></td>
</tr>
<tr>
<td>Help others to work</td>
<td>$I_{help}$</td>
<td>2.3</td>
<td>1.3</td>
<td>2.6</td>
<td>12.5</td>
<td></td>
</tr>
<tr>
<td>Output Demands (OD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handle workload</td>
<td>$O_{handle}$</td>
<td>2.7</td>
<td>1.3</td>
<td>2.8</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>Work fast enough</td>
<td>$O_{fast}$</td>
<td>2.9</td>
<td>1.4</td>
<td>3.0</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>Finish work on time</td>
<td>$O_{finish}$</td>
<td>2.4</td>
<td>1.4</td>
<td>3.2$^c$</td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>Work without mistake</td>
<td>$O_{mistak}$</td>
<td>2.0</td>
<td>1.2</td>
<td>2.8$^c$</td>
<td>5.7</td>
<td></td>
</tr>
<tr>
<td>Do all you’re capable of</td>
<td>$O_{capab}$</td>
<td>2.9</td>
<td>1.4</td>
<td>2.1$^c$</td>
<td>2.3</td>
<td></td>
</tr>
</tbody>
</table>

scores from TM, MI, OD subscales were reverse-coded to achieve consistent response orientation with PD subscale (for all items: 1 = less work limitations, 5 = more work limitations)

$^a$ N/R = missing due to item non-response

$^b$ N/A= missing due to “does not apply to my job” option selected by respondents (treated as missing value as per developer recommendation)

$^c$ for these 3 OD items, data for $n=72$ were lost due to an unexpected technical problem with the touch-screen data entry system, not considered as N/R
### Table 2.3 Model fit statistics for nested WLQ-25 and post-hoc WLQ<sup>PD</sup>-factor structures (MI subscale intact vs. expanded)

<table>
<thead>
<tr>
<th>Model</th>
<th>$X^2$</th>
<th>df</th>
<th>$P$</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>Model difference</th>
<th>Model difference $X^2$</th>
<th>df</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WLQ-25: no modifications</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-factor model (MI intact)</td>
<td>2812</td>
<td>269</td>
<td>&lt;.0001</td>
<td>0.901</td>
<td>0.890</td>
<td>0.065 (0.062-0.067)</td>
<td>0.060</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>5-factor model (MI expanded)</td>
<td>2544</td>
<td>265</td>
<td>&lt;.0001</td>
<td>0.911</td>
<td>0.900</td>
<td>0.062 (0.059-0.064)</td>
<td>0.058</td>
<td>207</td>
<td>4</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td><strong>WLQ-25: allowed correlation of residual variance between $P_{walk}$ and $P_{sit}$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-factor model (MI intact)</td>
<td>2227</td>
<td>268</td>
<td>&lt;.0001</td>
<td>0.924</td>
<td>0.915</td>
<td>0.057 (0.055-0.059)</td>
<td>0.054</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>5-factor model (MI expanded)</td>
<td>1956</td>
<td>264</td>
<td>&lt;.0001</td>
<td>0.934</td>
<td>0.925</td>
<td>0.053 (0.051-0.055)</td>
<td>0.051</td>
<td>201</td>
<td>4</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td><strong>Post-hoc WLQ&lt;sup&gt;PD&lt;/sup&gt;-models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-factor model (MI intact)</td>
<td>1641</td>
<td>149</td>
<td>&lt;.0001</td>
<td>0.925</td>
<td>0.914</td>
<td>0.067 (0.064-0.070)</td>
<td>0.055</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>4-factor model (MI expanded)</td>
<td>1405</td>
<td>146</td>
<td>&lt;.0001</td>
<td>0.937</td>
<td>0.926</td>
<td>0.062 (0.059-0.065)</td>
<td>0.051</td>
<td>183</td>
<td>4</td>
<td>&lt;.001</td>
<td></td>
</tr>
</tbody>
</table>

Criteria for model fit: chi-square: non-significant $p$ value; CFI, Comparative Fit Index: >0.90=adequate, ≥0.95=good; TLI, Tucker-Lewis Index: >0.90=adequate, ≥0.95=good; RMSEA, root mean square error of approximation: <0.08=adequate, <0.05=good; SRMR, standardized root mean square residual: ≤0.08 adequate, <0.06 good.

WLQ<sup>PD</sup>-model: 19 items (6 items from PD subscales removed from WLQ-25).

Chi-square difference tests were adjusted for non-normality scaling correction factor.
Figure 2.1 Standardized parameter estimates from confirmatory factor analyses of nested WLQ-25 models: 4-factor model with MI subscale intact (top) and 5-factor model with MI subscale divided into MD and IP subscales (bottom)
CHAPTER 3: Predictive ability of the Upper-Limb Work Instability Scale for transitioning out of work among injured workers with chronic upper-extremity disorders (Objective #2)
3.1 PREFACE

A number of measures of health-related work limitations currently exist in the literature, and it is well-recognized that they capture on-the-job problems from different perspectives (Beaton et al., 2009; Escorpizo et al., 2007; Tang et al., 2011a). However, despite such differences, it is likely that a number of them are suitable for use in WRUEDs, but for potentially unique purposes. This can be illustrated when comparing between the WLQ-25 and UL-WIS – two measures that are promising for use in this population. The WLQ-25 is designed to assess limitations while performing specific demands in four aspects of work: time management demands, physical demands, mental interpersonal demands, and output demands (Lerner et al., 2001). Beyond its potential as an outcome measure, it also has potential for health economic research, since its scores can be translated into a productivity-oriented metric compatible for costing (i.e. WLQ Index) (Lerner et al., 2003b). In contrast, the UL-WIS (Tang et al., 2011b) is designed to capture a concept referred to as work instability (WI), which is defined as “a state of mismatch between an individual’s functional abilities and the demands of his or her job”. If not resolved, it is believed that a state of high work instability may threaten continuing employment (Gilworth et al., 2003). As such, the UL-WIS may have potential utility as a prognostic measure for outcomes concerning work sustainability.

To help affirm whether various measures are suitable for use in WRUEDs, research evaluations of their measurement properties are important. The lack of available evidence on the factorial validity of the WLQ-25 was identified, and therefore, this was investigated in Chapter 2 to contribute to the existing body of evidence on this tool. Similarly, it was also recognized that there are gaps in evidence for the UL-WIS, which is a relatively new measure. Most notably, the postulated link between work instability and adverse employment outcomes has yet to be empirically verified in WRUEDs to-date. This was considered a worthwhile question to explore, as insights are needed on the relevance of WI, and also the potential utility of the UL-WIS as a prognostic tool in this population. This was the main objective for Chapter 3 (Objective #2), which, complementing the work done in Chapter 2, is intended to help advance the measurement of work disability in WRUEDs (thesis theme #1).
3.2 ABSTRACT

Objectives: To test the predictive ability of the Upper-Limb Work Instability Scale (UL-WIS) for transitioning out of work among injured workers with work-related upper-extremity disorders (WRUEDs), and to identify UL-WIS cut-scores. Methods: Injured workers working at initial attendance of a Workplace Safety and Insurance Board Upper-Extremity Specialty clinic were identified (n=356) from an existing dataset for a 12-month prospective cohort study (follow-up at 3, 6, and 12 months). Data were collected from self-report surveys assessing socio-demographics, health-related, and work-related variables, including work status at each study follow-up. Multivariable logistic regressions were conducted to investigate whether baseline UL-WIS (range: 0-100) is an independent predictor of a subsequent transition out of work within the study period. Nine variables were initially considered as model covariates, including age, gender, upper-limb pain intensity, mental health status, job strain, upper-limb physical demands, co-worker support, supervisor support, and organizational policies and practices. Positive predictive values (PPVs) and negative predictive values (NPVs) across the scale range were examined as a basis to identify meaningful UL-WIS cut-scores. Results: Among injured workers in our sample, 51.1% had transitioned out of work at some point over the study period. Baseline UL-WIS was shown to be a significant independent predictor of a subsequent transition out of work (adjusted odds ratio=1.03 [1.01, 1.06], p=0.003). Across the UL-WIS score range, the cut-score of >80.8 was shown to have the highest PPV (0.80 [95%CI: 0.52, 0.95]), while the cut-score of <41.9 was shown to have the highest NPV (0.81 [95%CI: 0.61, 0.94]). Conclusions: Transitioning out of work was common for injured workers with chronic/complex WRUEDs. The UL-WIS demonstrated independent predictive ability of this outcome; however, as a standalone tool, some limitations in predictive accuracy were also evident. Three bands of UL-WIS representing low, moderate, and high risk of a subsequent transition out of work were proposed based on the identified cut-scores.
Chapter 3: Predictive ability of the UL-WIS

3.3 INTRODUCTION

Recent reports suggest that work-related upper-extremity disorders (WRUEDs) have remained prevalent in recent years (Bureau of Labor Statistics, 2013; Workplace Safety and Insurance Board of Ontario, 2012). Given the potential chronicity of WRUEDs, it is suggested that an initial return-to-work may not necessarily represent the end of the recovery process for injured workers (Baldwin and Butler, 2006; Pransky et al., 2000). Accordingly, the sustainability of work has emerged as an important issue in this population, as injured workers who are able to work may still be at risk for transitioning out of work due to recurrence and/or residual effects of initial injury (Baldwin and Butler, 2006; Keogh et al., 2000; Mayer et al., 1999; Pransky et al., 2000). Among those able to work, one question faced by clinicians and workplace parties is which injured workers are more likely to transition out of work? Measurement tools with the ability to predict this outcome have not been well-established in the literature. If available, the ability to identify those at the greatest risk for poor work sustainability could help facilitate timelier (i.e. earlier) applications of health/workplace interventions (e.g. job accommodations) to help mitigate such risk. In turn, this could also help ensure that scarce healthcare resources and disability management efforts are directed to those who have the greatest need, and thus contributing to more optimized allocation of these resources and efforts.

In recent years, the extent and significance of “on-the-job” problems among working individuals while managing existing musculoskeletal disorders have been increasingly recognized (Burton et al., 2005; Burton et al., 2004; Hemp, 2004). A connection between precariousness of a person’s ability to work and the risk of transitioning out of work seems plausible, although this relationship requires empirical testing and further understanding. One available measurement tool that has specific applicability for WRUEDs is the 17-item Upper-Limb Work Instability Scale (UL-WIS) (Tang et al., 2011b). Initially derived from the rheumatoid arthritis version of the scale (i.e. the 23-item RA-WIS), the family of WIS tools is designed to quantify “work instability” (WI), which the original developers have defined by the developers as “a state of mismatch between a worker’s functional capabilities in relation to job demands due to a health disorder” (Gilworth et al., 2003). As such, high WI represents some precariousness in an individual’s existing work role, which may increase the
likelihood of adverse work outcomes in the future (e.g. a subsequent transition out of work) (Gilworth et al., 2003). In fact, a recent study of workers with arthritis has shown that the RA-WIS is predictive of an adverse employment event within one year (e.g. going off-work, reducing work hours, needing job modifications) (Tang et al., 2010). The predictive ability of the UL-WIS among WRUEDs, however, remains to be verified.

The main objectives of this study are: 1) to investigate whether the level of WI, as measured by the UL-WIS, is predictive of a subsequent transition out of work for injured workers with WRUEDs who are initially at work, and if so, 2) to determine meaningful cut-scores for this purpose. Based on its conceptualization and intended purpose, we hypothesize that the UL-WIS will demonstrate an independent ability to predict this work outcome.
3.4 METHODS

Study setting
Participants for this study were injured workers attending an Upper-Extremity Specialty clinic operated by the Workplace Safety and Insurance Board (WSIB) of Ontario (www.wsib.on.ca). The Workplace Safety & Insurance Board is a single-source workers’ compensation insurer in the region, funded by employers in Ontario and legislated by the provincial government. These clinics provide specialized clinical consultations for claimants experiencing an atypical recovery course (i.e. insufficient progress after ~6 months, more chronic/complex cases) following soft tissue musculoskeletal injuries (e.g. repetitive strain injuries, acute or cumulative trauma disorders of muscle/tendon/ligaments) or uncomplicated fractures to the upper-limb. Attendees of WSIB Specialty clinics are characterized by varied symptomatology (e.g. localized or radiating pain, paresthesia), functional limitations, and extent of work disability due to WRUEDs (i.e. ranging from able to work with minimal limitations to completely unable to work). At these clinics, clinical consultations are provided by a multidisciplinary team of rehabilitation therapists, orthopaedic surgeons, social workers, and case managers to evaluate recovery progress and prognosis, suitability for work, or candidacy for surgery or rehabilitative interventions.

Study design and data collection
A secondary data analysis was performed, using data from a 12-month prospective cohort study that followed 614 attendees of one of two WSIB Upper-Extremity Specialty clinics (Toronto or London sites). This data was originally collected between 2004-2005. During this study, survey questionnaires were fielded to injured workers at initial clinic attendance (baseline), and 3, 6, and 12-month follow-up. These surveys comprised questions on work status, socio-demographics, as well as health- and job-related variables based on a literature review of prognostic factors of work disability in musculoskeletal health. Participants who had missed a follow-up survey were allowed to continue in the study (i.e. participate in the next follow-up survey).
Inclusion criteria
Since our aim was to test the predictive ability of a tool that is specifically applicable for working individuals, to be eligible for this study the injured worker must: 1) have been working at the time of initial clinic attendance (i.e. baseline), 2) have understood written English, and 3) have provided written consent for the use of their de-identified survey data for research purposes. All participants were explicitly assured that neither the WSIB nor their employer would have access to their survey responses. In our dataset, 356 of 614 (58%) injured workers were working at baseline, and therefore eligible for the current analysis. Approval for this study was obtained from research ethics boards at the Sunnybrook Health Sciences Centre, St. Michael's Hospital, University of Toronto, and University of Western Ontario.

Variables

Study outcomes
A transition out of work over the course of the study was the primary outcome of interest. At each of the three follow-up time-points, an injured worker’s work status was determined by the survey question “Are you currently working? (yes or no)”. To operationalize “transitioning out of work”, a binary outcome (#1) was constructed to differentiate workers who reported working at all three follow-up time-points (“non-event”), from those who reported not working at one or more of the follow-up time-points (“event”). If there were missing data on work status, and this precluded the worker from being definitively classified into either of the outcome categories, the observation would be excluded from the relevant analyses. To provide a sensitivity analysis, we constructed two secondary binary outcomes based on the timing of workers’ initial transition out of work. For outcome #2 (“early” transition out of work), an “event” was defined as not working at the 3-month follow-up. For outcome #3 (i.e. “late” transition out of work), an “event” was defined as not working only at either 6- and/or 12-month follow-up. For both these secondary outcomes, the same “non-event” group as per outcome #1 was applied.

Main predictor of interest: UL-WIS
Work instability (WI) was assessed at study baseline by the 17-item UL-WIS (Tang et al., 2011b). The UL-WIS consists of items that assess perceptions of symptom control at work,
task performance at work, stamina at work, issues of time management, cognitive distresses associated with work, and perceptions of sustainability of current work role (collectively referred as work instability by original scale developers) (Gilworth et al., 2003). A dichotomous response option (yes [1] / no [0]) is provided for each item, and the scale is scored by summing all 17 items for raw scores from 0-17. These were then transformed into Rasch scores (range: 0-100) based on previous insights (Tang et al., 2011b). In this study, we allowed only a maximum of 1 missing item (i.e. <10% of total scale items), otherwise the scale was not scored. Evidence of internal consistency and construct validity of the UL-WIS have been previously demonstrated in WRUEDs (Tang et al., 2011b).

Other variables (descriptors and/or model covariates)
A series of socio-demographic, health-related and job-related variables were also assessed at study baseline. Socio-demographic variables included age (continuous), gender (male or female), marital status (married/common law vs. others), and level of education (completed postsecondary level vs. others). Health-related variables included number of painful sites (assessed from a pain diagram with 15 upper-body regions), upper-limb pain intensity (5-item pain subscale of the Shoulder Pain and Disability Index, [SPADI]) (MacDermid et al., 2006; Roach et al., 1991; Roy et al., 2009), mental health status (5-item mental health subscale from version 2 of the Medical Outcomes Studies Short-Form 36 [SF-36]) (Ware Jr., 2000), and a list of co-morbid conditions based on a modified Sangha scale (Sangha et al., 2003). Job-related variables included several measures from the Job Content Questionnaire (Karasek et al., 1998): psychological job demands (5 items), skill discretion (3 items), and decision authority (6 items), as well as upper-limb physical job demands (based on a 4-item checklist) (Eriksen et al., 1999), co-worker support (4 items from the Psychosocial Aspect of Work Questionnaire (Symonds et al., 1996), and a measure of organizational policies and practices (OPPs, 11-item scale) (Tang et al., 2011d). To further describe the cohort, job title provided by the injured workers in the study survey was also appraised to examine the distribution of occupational sectors (9 total categories) based on the National Occupational Classification from Human Resources and Skills Development Canada.
Analysis

For all analyses, data gathered from the two Upper-Extremity Specialty clinic sites were pooled since they share common meso-/macro-level factors of occupational disability (e.g. healthcare and workers’ compensation system, labor market and socio-political climate), and it was not of primary interest to investigate site-specific differences in the relationship between UL-WIS and subsequent work transitions.

Descriptive analysis

To describe the study participants, univariate statistics ($n$, mean, %) were applied to examine the distribution of baseline socio-demographic, health-related, and work-related variables. The distribution of work status at the four study time-points was described, and also plotted on a cumulative bar graph.

Logistic regressions

A univariable logistic regression was initially performed with baseline UL-WIS as the sole predictor of our transition out of work study outcome. The concordance ($\delta$) statistic of this model was specifically observed to inform the predictive ability of the UL-WIS as a standalone tool. Then, multivariable logistic regressions were performed to assess whether this is an independent predictive relationship (i.e. free from confounding bias). To this end, additional personal (worker-related) and environmental (workplace-related) factors were considered for inclusion in multivariable regressions, since they have also been postulated to be of relevance in occupational disability (Feuerstein et al., 2003; Krause et al., 2001b; Sandqvist and Henriksson, 2004; Schultz et al., 2007), and therefore, have potential to confound the relationship between UL-WIS and transitioning out of work. In terms of personal factors, key postulated domains include socio-demographics (e.g. age, gender), biomedical factors (e.g. clinical pathophysiology and symptomatology of the underlying health condition, such as injury type or bodily pain), and psychosocial factors (e.g. mental health, motivation, vitality). In terms of workplace factors, key postulated domains include physical job characteristics (e.g. physical task demands and postural demands), psychosocial job characteristics (e.g. job strain, job control, psychological demands), and workplace social support (e.g. co-worker or supervisor support) (Krause et al., 2001b). Beyond job-level domains, the relevance of organizational-level factors has also been theorized and major
domains in this area include people-oriented culture, proactive disability management, active safety leadership (safety climate), and ergonomic job design practices (Amick III et al., 2000a; Habeck et al., 1998; Habeck et al., 1991). With considerations given to current guidelines on events-per-variable ratio (i.e. not exceeding 10:1) (Harrell et al., 1996; Peduzzi et al., 1996) and the available variables from our dataset, a total of nine variables covering these domains were selected, including four personal factors (age, gender, upper-limb pain intensity, mental health status) and five environmental factors (upper-limb physical job demands, co-worker support, social support, OPPs, and a measure of job strain: ratio of psychological job demands and job control [sum of skill discretion and decision authority]).

To identify confounders of the relationship between UL-WIS and transitioning out of work, a “change-in-estimate” screening step was initially performed. By this approach, each of the nine factors was individually added to a base model with only UL-WIS as the predictor. If the inclusion of a factor led to a >10% change in the UL-WIS estimate, then this factor would be considered a confounder and included in the multivariable model (Greenland and Neutra, 1980; Kleinbaum et al., 1982). No elimination steps were performed subsequently. As the secondary outcomes (#2 and #3) were designed only for the purpose of sensitivity analysis, the same covariates were kept in these models. For all regression analyses, complete case analysis (listwise deletion) was used where observations with any missing covariates or response were excluded. Effects were expressed as unstandardized parameter estimates ($b$) with the associated standard error (SE), Wald Z value, and adjusted odds ratio (OR) and 95% confidence interval (CI). Proper model fit for each model was verified based on Hosmer-Lemeshow goodness-of-fit statistics ($p>0.05 = acceptable fit$).

**Missing data analysis**

To gain insights on the likelihood of selection bias, a comparison of baseline UL-WIS and outcome proportions was performed between the subset of observations included (i.e. complete cases) and excluded (i.e. incomplete cases) from our regressions.

**Determination of UL-WIS cut-scores**

If the UL-WIS demonstrated independent predictive ability (i.e. $p<0.05$ in multivariable regressions), then cut-scores were sought to differentiate between low, moderate, and high risk of a subsequent transition out of work, similar to previous efforts for the original RA-
WIS (Gilworth et al., 2003). To this end, we assessed the positive predictive values (PPV) and negative predictive values (NPV) (Altman and Bland, 1994) across the full UL-WIS scale range (i.e. for each score increment). In the current context, the PPV would represent the proportion of workers with a “positive” test (i.e. UL-WIS above a given cut-score) that had transitioned out of work over the course of this study. The NPV, on the other hand, would represent the proportion of injured workers with a “negative” test (i.e. UL-WIS below a given cut-score) that had not made such a transition. Guided by these values, our aim was to determine two UL-WIS cut-scores: a threshold (moderate/high risk) where scores greater than this value would be associated with a fairly high likelihood of a subsequent work transition (i.e. high PPV), and a second threshold (low/moderate risk) where scores below this value would be associated with a fairly high likelihood of not experiencing a subsequent work transition (i.e. high NPV). Two main considerations were made in determining such cut-scores: 1) the magnitude of the PPV/NPV (i.e. higher values are more desirable), and 2) overall trend/pattern of these values across the scale range, for example, considerations for location of clear “step ups” in PPV/NPV or whether another cut-score might be more sensible despite potentially lower values. All analyses were conducted with R version 3.0.1.
3.5 RESULTS

Baseline characteristics of injured workers

Injured workers participating in this study (n=356) had a mean age of 44.9 (SD=9.2, range: 19-68), and 55.7% were female (Table 3.1). On average, injured workers reported 5.0 (SD=3.8) painful sites (out of 15 available regions on pain diagram), with 41.0% indicating involvement of bilateral upper-extremities. The distribution of painful sites was as follows (not mutually exclusive): neck (44.1%), shoulder (75.0%), upper-arm (49.7%), elbow (52.0%), forearm (36.8%), wrist (39.0%), hand (39.6%), upper-back (23.9%), and lower-back (16.9%).

At the sample-level, the SPADI pain subscale mean was 30.7 (SD=11.4) and SF-36 mental health subscale mean was 64.9 (SD=21.4). Injured workers also reported having an average of 0.8 (SD=0.9) out of 11 co-morbidities assessed. The mean UL-WIS, completed by 309 workers, was 58.2 (SD=18.0), which was slightly over the midrange of the scale (Figure 3.1). The item-level distributions of the UL-WIS are presented in Table 3.2. Based on the National Occupational Classification, our injured workers were most commonly employed in two sectors: trade, transport and equipment operators (25.5%) or processing, manufacturing and utilities (33.2%).

Extent of transitioning out of work

Work status by follow-up is displayed in Figure 3.2. At 3-month follow-up, n=222 (62.4%) were working, n=78 (21.9%) were not working, and n=56 (15.7%) did not provide a response. At 6-month follow-up, n=211 (59.3%) were working, n=95 (26.7%) were not working, and n=50 (14.0%) did not provide a response. At 12-month follow-up, n=191 (53.7%) were working, n=87 (24.4%) were not working, and n=78 (21.9%) did not provide a response.

Synthesizing available work status data from all three follow-up time-points, n=143 (51.1%) had transitioned out of work at some point over the course of the study (i.e. classified as “event”), while n=137 (48.9%) had worked at all follow-up time-points (i.e. classified as “non-event”). A definitive classification could not be made for n=76 due to incomplete
information on work status (subsequently excluded from analyses). Among those who had transitioned out of work, \( n=78 \) (54.5%) were not working at the 3-month follow-up (“early” transition out of work – “event” for outcome \#2), and \( n=65 \) (45.5%) had first reported to be not working at either 6- or 12-month follow-up (“late transition out of work”, “event” for outcome \#3).

**Predicting transitioning out of work**

Univariable logistic regression revealed baseline UL-WIS to be a statistically significant predictor of a subsequent transition out of work (crude OR=1.04 [95%CI: 1.02, 1.06], \( p<0.001 \)). The \( R^2 \) statistic for this model was 0.69. Screening of covariates by the “change-in-estimate” approach identified four confounders: age, upper-limb pain intensity, mental health status, and OPPs. When these confounders were included in our final multivariable logistic regression, the UL-WIS remained a significant independent predictor (adjusted OR=1.03 [95%CI: 1.01, 1.06], \( p=0.003 \)) (Table 3.3). This estimate was in fact quite similar to the crude estimate, and translates to an adjusted OR of 1.41 [95%CI: 1.26, 1.58] for a 10-point increase in UL-WIS score. Sensitivity analyses revealed similar findings for different timing of transitions out of work. When predicting “early” transitions out of work (outcome \#2), the adjusted OR for the UL-WIS was 1.04 [95%CI: 1.01, 1.07], \( p=0.004 \). When predicting “late” transitions out of work (outcome \#3), the adjusted OR for the UL-WIS was 1.03 [95%CI: 1.01, 1.05], \( p=0.03 \). Adequate model fit was verified for all models based on the Hosmer-Lemeshow goodness-of-fit statistic (\( p>0.05 \)).

**Missing data analysis**

A comparison of observations included \((n=233)\) and excluded \((n=123, \text{ due to missing covariates or response})\) from multivariable logistic regressions revealed similar baseline UL-WIS as well as outcome proportions. The UL-WIS means were 58.8 and 56.4 for the included and excluded groups, respectively, while the proportions of injured workers who had transitioned out of work were 50.2% and 55.3% for the included and excluded groups, respectively.
UL-WIS cut-scores

A plot of the PPVs and NPVs over the full range of UL-WIS scores is illustrated in Figure 3.3. The highest PPV was observed at the cut-score of >80.8 (PPV=0.80 [95%CI: 0.52, 0.95]), whereas the highest NPV was found at the cut-score of <41.9 (NPV=0.81 [95%CI: 0.61, 0.94]). We ultimately felt that both these represented sensible UL-WIS cut-scores, following an inspection of the overall trends in PPVs and NPVs across the scale range. By this, three bands of the UL-WIS emerged as follows: low risk (<41.9), moderate risk (41.9-80.8), and high risk (>80.8) for a subsequent transition out of work among injured workers recovering from WRUEDs.
3.6 DISCUSSION

The sustainability of work is an important issue among injured workers with WRUEDs, given the potential for persisting symptoms and functional impairments following initial injury (Baldwin and Butler, 2006; Keogh et al., 2000). The main aim of this study was to evaluate the predictive ability of the UL-WIS, by investigating the relationship between WI and subsequent transitioning out of work among injured workers with WRUEDs. The main study findings were that 1) transitioning out of work was common among injured workers who were initially working, 2) WI, as measured by the UL-WIS, was shown to be an independent predictor of such transitions, and 3) two UL-WIS cut-scores of 41.9 and 80.8 were identified, dividing the tool into three bands (i.e. low, moderate, and high risk).

Although the concept of WI (mismatch between work ability and demands) is relatively unknown in the occupational disability literature, in some sense this notion is not entirely new. For example, WI resonates with the idea of person-environment (P-E) fit from organizational psychology (Caplan et al., 1975; Harrison, 1985). The person-environment fit theory posits that a discrepancy between the person (i.e. worker) and environment (i.e. workplace) constitute an important basis of job stress (Caplan et al., 1975; Harrison, 1985). What’s unique about WI, perhaps, is that this concept has been operationalized into a family of self-report measures through rigorous research involving both qualitative and quantitative measurement work (Gilworth et al., 2007; Gilworth et al., 2006; Gilworth et al., 2003; Gilworth et al., 2009a; Gilworth et al., 2009b). When applied in research, the concept of WI has shown to resonate well with workers with various musculoskeletal health conditions. Among 250 workers with arthritis, the content comprehensiveness of the RA-WIS was endorsed by 85.6% of the respondents (Tang et al., 2013b). Moreover, a head-to-head comparison of instruments found that injured workers with WRUEDs preferred the RA-WIS over three other at-work disability measures for capturing their on-the-job experience (Tang et al., 2009). The potential versatility of the family of WIS measures should also be recognized. Although we have focused on examining its prognostic value in this study, several recent clinical trials and cohort studies have applied these tools as outcome measures (i.e. study endpoints) (Bernaards et al., 2007; Emery et al., 2011; Macedo et al., 2009).
Chapter 3: Predictive ability of the UL-WIS

The robustness of UL-WIS’s predictive ability was supported in several ways in this study. First, our regressions were reasonably well-adjusted in terms of the range of variables considered as potential confounders, which were selected on the basis of postulated relevance in occupational disability. Second, our sensitivity analyses revealed similar predictive ORs for “early” and “late” transitions out of work. And third, our missing data analyses suggested a low likelihood of significant selection bias, as observations included and excluded from our regressions were shown to be relatively similar in key variables. Furthermore, this finding paralleled results from a previous study where WI similarly emerged as a significant predictor of future adverse employment outcomes among workers with arthritis (Tang et al., 2010). Yet, some limitations of the UL-WIS as standalone predictive tool were also evident. The adjusted OR of 1.03 (1.41 for a 10-point increase) was perhaps not overly impressive, and the observed χ statistic of 0.69 in the univariate regression model would be considered “poor” by the standards of clinical diagnostic tools (Hanley and McNeil, 1982). That said, to offer an additional perspective, a 17-factor prognostic model of work disability in carpal tunnel syndrome reported an area under the receiver operating characteristic curve of 0.76 (Turner et al., 2007). From this vantage point, the predictive ability of the UL-WIS as a standalone tool seems respectable. Nonetheless, towards even better predictive accuracy, future research efforts toward developing a predictive “index” (i.e. inclusive of multiple factors beyond the UL-WIS) may be worthwhile.

Our efforts to determine meaningful cut-scores represent a first step towards “benchmarking” the UL-WIS to enhance its interpretability during future use. To this end, we honed in on the parts of the tool where there are specific strengths in predictive ability (i.e. high PPV or NPV). Beyond the purpose of triaging injured workers (i.e. identifying those at high risks for transitioning out of work), such cut-scores are also potentially useful from the perspective of goal-setting during patient care (e.g. aim to achieve a certain score threshold by next clinic follow-up), or for research purposes (i.e. stratification of UL-WIS scores during analysis). Certainly, limitations associated with these two cut-scores should be recognized during future applications. The estimated levels of PPV and NPV (i.e. ~0.80) for the high and low UL-WIS cut-scores, respectively, still imply some probability of incorrect predictions (~20%).
Several limitations of the study should be considered. It is important to reiterate that our study cohort – attendees of WSIB Specialty clinics – represents only the most complex/chronic cases of WRUEDs. Previous reports have shown that in a majority of all cases of WRUEDs are, in fact, relatively fast resolving (Baldwin and Butler, 2006; Feuerstein et al., 1998; Hashemi et al., 1998). As such, the observed propensity for exiting work is unlikely to generalize to less severe/complex cases. Limitations in the operationalization of our study outcomes should also be recognized. First, for our primary study outcome, all patterns of work transitioning out of work were treated equally (i.e. all coded similarly as “event”). Second, although our secondary outcomes aimed to differentiate between “early” and “late” transitions out of work, the precision of timing was limited given the frequency of survey data collection. As such, our dataset would also not have captured any work transitions taking place between two measurement points. Third and finally, different types of work duties (e.g. regular, modified, or reduced hours) were not differentiated in our outcome. While we made efforts to assess this in our survey, many injured workers had in fact indicated that more than one work status category applied to them, perhaps reflecting some complexity in their existing work status. For future considerations, we propose two main areas for improvements in terms of design: 1) collection of more detailed data on work transitions (e.g. more precise timing of on/off episodes), and 2) repeated measurement of predictors/covariates over time, which would afford opportunity to apply more sophisticated analytic strategies (e.g. time-to-event analysis with time-dependent variables) to elucidate factors linked to sustainability of work.

Based on current study results, we conclude that the UL-WIS, which quantifies WI, demonstrated an ability to predict a subsequent transition out of work among injured workers recovering from a WRUED. However, as a standalone tool, its overall predictive accuracy might be considered somewhat modest. Focusing on areas of the tools with specific predictive strengths, UL-WIS cut-scores of <41.9 (NPV=0.81) and >80.8 (PPV=0.80) were identified, which divides the tool into three bands - low, moderate, and high risks. Recognizing WI to be a prognostic factor of poor sustainability of work, increased awareness to the extent of WI experienced by injured workers may be important, as well as continued research efforts to identify interventions that can lessen WI.
### 3.7 TABLES AND FIGURES

**Table 3.1** Baseline characteristics of injured workers \( (n=356) \)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Instrumentation</th>
<th>Scale range</th>
<th>( n ) available</th>
<th>mean</th>
<th>SD</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socio-demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Single-item</td>
<td>numeric</td>
<td>353</td>
<td>44.9</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Single-item</td>
<td>% female</td>
<td>352</td>
<td>.</td>
<td>.</td>
<td>55.7</td>
</tr>
<tr>
<td>Marital status</td>
<td>Single-item</td>
<td>% married or common law</td>
<td>351</td>
<td>.</td>
<td>.</td>
<td>70.9</td>
</tr>
<tr>
<td>Education</td>
<td>Single-item</td>
<td>% completed postsecondary level</td>
<td>351</td>
<td>.</td>
<td>.</td>
<td>33.3</td>
</tr>
<tr>
<td><strong>Health-related variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of painful sites</td>
<td>Body diagram (15 upper-body sites assessed)</td>
<td>0-15 painful sites</td>
<td>347</td>
<td>5.0</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>Upper-limb pain intensity</td>
<td>SPADI pain subscale (5 items)</td>
<td>0-50 ((50 = high pain intensity))</td>
<td>349</td>
<td>30.7</td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td>Mental health status</td>
<td>SF-36 mental health subscale (5 items)</td>
<td>0-100 ((100 = best mental health status))</td>
<td>346</td>
<td>64.9</td>
<td>21.4</td>
<td></td>
</tr>
<tr>
<td>Number of co-morbidities</td>
<td>Modified Sangha scale</td>
<td>out of 11 health conditions</td>
<td>356</td>
<td>0.8</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Work instability</td>
<td>UL-WIS (17 items)</td>
<td>0-17 ((17 = high work instability))</td>
<td>309</td>
<td>58.2</td>
<td>18.0</td>
<td></td>
</tr>
<tr>
<td><strong>Work-related variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill discretion</td>
<td>JCQ subscale (6 items)</td>
<td>1-4 ((4 = high skill discretion))</td>
<td>336</td>
<td>2.5</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Decision authority</td>
<td>JCQ subscale (3 items)</td>
<td>1-4 ((4 = high decision authority))</td>
<td>336</td>
<td>2.5</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Psychological job demands</td>
<td>JCQ subscale (5 items)</td>
<td>1-4 ((4 = high demands))</td>
<td>317</td>
<td>2.7</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Supervisor support</td>
<td>JCQ subscale (5 items)</td>
<td>1-4 ((4 = high support))</td>
<td>303</td>
<td>2.8</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Organizational policies and practices (OPPs)</td>
<td>OPP-11 (11-items)</td>
<td>1-5 ((5 = most favorable OPPs))</td>
<td>346</td>
<td>3.1</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Upper-limb physical job demands</td>
<td>4-item checklist</td>
<td>0-4 ((4 = high demands))</td>
<td>349</td>
<td>2.4</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Co-worker support</td>
<td>PAWQ subscale (4 items)</td>
<td>1-5 ((5 = high support))</td>
<td>340</td>
<td>3.7</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

SPADI, Shoulder Pain and Disability Index; SF-36, Medical Outcomes Studies Short-Form 36; UL-WIS, Upper-Limb Work Instability Scale; JCQ, Job Content Questionnaire; OPP-11, 11-item Organizational Policies and Practices scale; PAWQ, Psychosocial Aspect of Work Questionnaire.
Figure 3.1 Distribution of baseline UL-WIS scores among injured workers attending a Workplace Safety & Insurance Board Upper-Limb Specialty clinic (n=309)
Table 3.2 Item-level distribution of the UL-WIS

<table>
<thead>
<tr>
<th>Item #</th>
<th>Label</th>
<th>Yes</th>
<th>No</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>reduce hours</td>
<td>69</td>
<td>228</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>worried</td>
<td>203</td>
<td>105</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>pain or stiffness</td>
<td>252</td>
<td>56</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>stamina</td>
<td>207</td>
<td>102</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>holiday</td>
<td>53</td>
<td>252</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>push myself</td>
<td>214</td>
<td>94</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>say no</td>
<td>257</td>
<td>49</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>watch how much</td>
<td>295</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>opening doors</td>
<td>127</td>
<td>180</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>extra time</td>
<td>227</td>
<td>81</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>give up work</td>
<td>102</td>
<td>206</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>get on</td>
<td>255</td>
<td>50</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>feeling tired</td>
<td>176</td>
<td>126</td>
<td>7</td>
</tr>
<tr>
<td>14</td>
<td>restricted</td>
<td>148</td>
<td>156</td>
<td>5</td>
</tr>
<tr>
<td>15</td>
<td>getting up earlier</td>
<td>152</td>
<td>156</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>very stiff</td>
<td>89</td>
<td>218</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>stress</td>
<td>173</td>
<td>133</td>
<td>3</td>
</tr>
</tbody>
</table>

UL-WIS was scored in n=309 (i.e. completed ≥16/17 items)
Figure 3.2 Distribution of work status over 12 months for injured workers who were working at time of initial Upper-Limb Specialty clinic attendance ($n=356$)
Table 3.3 Predictive ability of UL-WIS on subsequent transitioning out of work assessed by multivariable logistic regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>$b$</th>
<th>SE</th>
<th>Wald Z</th>
<th>$P$</th>
<th>Adjusted OR$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work instability</td>
<td>0.03</td>
<td>0.01</td>
<td>3.00</td>
<td>0.003</td>
<td>1.03 (1.01-1.06)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.05</td>
<td>0.02</td>
<td>-2.92</td>
<td>0.004</td>
<td>0.95 (0.92-0.98)</td>
</tr>
<tr>
<td>Upper-limb pain intensity</td>
<td>0.03</td>
<td>0.01</td>
<td>1.82</td>
<td>0.07</td>
<td>1.03 (1.00-1.06)</td>
</tr>
<tr>
<td>Mental health status</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.62</td>
<td>0.54</td>
<td>1.00 (0.98-1.01)</td>
</tr>
<tr>
<td>Organizational policies and practices (more favorable)</td>
<td>-0.46</td>
<td>0.20</td>
<td>-2.28</td>
<td>0.02</td>
<td>0.63 (0.42-0.93)</td>
</tr>
</tbody>
</table>

$n=233$, after listwise deletion of observations with missing covariates or response; omnibus likelihood ratio = 45.06 (df=5), $P<0.0001$; Hosmer-Lemeshow Goodness of Fit: $p>0.05$

$^1$ per unit increase in score of the variable

note: the adjusted OR for 10 unit increase in score for the UL-WIS is 1.41 [1.26, 1.58]
Figure 3.3 Positive predictive values and negative predictive values across the range of UL-WIS cut-scores. Points of highest positive predictive value (UL-WIS > 80.8) and negative predictive value (UL-WIS < 41.9) were identified.
CHAPTER 4: Health-related work limitations following work-related upper-extremity disorders: does the work environment play a role? (Objective #3)
4.1 PREFACE

In Chapter 3 (Objective #2), the relationship between WI and subsequent transitioning out of work among WRUEDs was examined. While this was informative from a measurement perspective (i.e. predictive validity of a tool), it also provided insights on prognostic factors for work sustainability, which is a topic that relates to the second thesis theme (i.e. determinants of work disability). On this theme, it was argued at the outset that more research applying outcomes that concern the injured worker is needed, as well as studies that aim to elucidate the nature of worker-workplace (person-environment) interactions. Putting these elements together, it was clear that an investigation to better understand the basis of health-related work limitations associated with WRUEDs would be worthwhile, as it addresses both the aforementioned gaps. To this end, Chapter 4 (Objective #3) describes research that examines the role of the work environment as a potential contextual factor of health-related work limitations in WRUEDs, in particular, interactions between the work environment and key worker health factors like pain intensity and mental health status and their relationship to work limitations. Two fairly innovative techniques, namely, latent profile analysis (LPA) and multiple-group path analysis were applied in the study. To place this work in additional context, Objective #3 was intended as the first of several studies on this same topic. This initial study specifically took a broader view of the work environment, which will be complemented by future research (outside of thesis) taking a more specific view (i.e. specific workplace characteristics) towards a better understanding of the nature of worker-workplace interactions.
4.2 ABSTRACT

**Objective:** To profile injured workers recovering from work-related upper-extremity disorders (WRUEDs) based on 12 different work characteristics, and to examine the relationship between the work environment and the extent of health-related work limitations experienced. **Methods:** Secondary analysis of cross-sectional survey data was conducted, focusing on injured workers who were working at time of initial attendance \( (n=349) \) at an Upper-Extremity Specialty clinic operated by the Workplace Safety and Insurance Board of Ontario. These surveys assessed socio-demographics, health-related factors (e.g. upper-limb pain, mental health status), work characteristics, and two measures of health-related work limitations: the Disabilities of the Arm, Shoulder and Hand Work Module (DASH-W, range: 0-100), and Upper-Limb Work Instability Scale (UL-WIS, range: 0-100). Two sequential analyses were performed. First, a latent profile analysis was conducted to identify broad work environment profiles based on variations across 12 organizational- and job-level characteristics. Then, applying these profiles, multiple-group path analyses were conducted to examine potential main and moderating associations on health-related work limitations based on a hypothesis of “worker-workplace interactions”. Main and moderating associations were formally tested by applying equality constraints to model intercepts and slopes, respectively, and then assessing resultant changes to model fit using chi-square difference tests. **Results:** Considerable variations across the 12 organizational- and job-level factors were found among injured workers. Following latent profile analysis, four distinct work environment profiles emerged after considerations of theoretical plausibility, statistical performance, and interpretability. Subsequent multiple-group path analysis found no significant main nor moderating associations \( (p>0.05) \) of the broad work environment profiles on health-related work limitations. **Conclusions:** Injured workers with WRUEDs were successfully profiled based on a set of work characteristics with postulated relevance in occupational disability. However, against our hypothesis, membership in a specific work environment profile was not significantly related to the extent of health-related work limitations experienced. Contributing factors to the current findings were discussed, as well as potential research avenues to further elucidate the nature and relevance of “worker-workplace interactions”.

Chapter 4: Role of the work environment

4.3 INTRODUCTION

Work disability is a known concern following work-related upper-limb disorders (WRUEDs) (Baldwin and Butler, 2006; Feuerstein et al., 1998; Keogh et al., 2000). Beyond return-to-work issues, health-related work limitations (or “on-the-job” problems) in this population have been increasingly recognized, affecting those who are able to work following initial injury (Keogh et al., 2000; Morse et al., 1998; Pichora and Grant, 2010; Tang et al., 2009). This is a relevant concern from several perspectives. First, health-related work limitations represent an impact on an individual’s quality-of-(work)-life. Second, this signals a potential discrepancy between a worker’s abilities and existing job demands, which may indicate an unsustainable work role. Third, health-related work limitations have meaningful economic implications. Recent reports suggest that the cost of presenteeism (lost productivity while at-work) can actually exceed that of absenteeism attributed to employee health (Hemp, 2004; Li et al., 2006).

To-date, research efforts to elucidate the basis of health-related work limitations associated with WRUEDs have remained scarce. Theoretically, beyond health factors (e.g. physical and psychological well-being), the work environment is also commonly posited to have influence on the extent of work disability experienced (Feuerstein, 1991; Sandqvist and Henriksson, 2004; Schultz et al., 2007). Broadly, the work environment may be thought of as the collective influence of multiple work characteristics (e.g. organizational- and job-level factors), operating within the broader healthcare and insurance system and other societal-level influences (e.g. legislative, macroeconomic, socio-political climate) (Feuerstein et al., 2003; Krause et al., 2001b; Turner et al., 2000). Specific work characteristics of relevance are diverse, and may include factors related to physical job content (e.g. heavy or repetitive task demands), psychosocial job content (e.g. job autonomy, work flexibility, or co-worker/supervisor relations), and organizational policies and practices (e.g. on safety practices, proactive disability management, or organizational/managerial climate) (Amick III et al., 2000a; Habeck et al., 1998; Habeck et al., 1991; Hunt et al., 1993). Moreover, interactions between worker and workplace factors have also been postulated as a key basis of occupational disability. For example, Sandqvist and Henriksson suggested that the level of work functioning is a direct reflection of how well personal and environmental factors
“intersect” (Sandqvist and Henriksson, 2004). Yet, the nature of such “worker-workplace interactions” remains not well understood and has been rarely tested to-date.

Our overall aim is to elucidate the role of the work environment on extent of health-related work limitations among injured workers recovering from a WRUED. The specific objectives are two-fold: 1) to profile injured workers’ overall work environment based on variations across a set of 12 work characteristics, and 2) applying these initial results, to evaluate main and moderating associations of broad work environment profiles on health-related work limitations, guided by a hypothesis of worker-workplace interactions as per Sandqvist and Henriksson’s theorizations (Figure 1). To our knowledge, research efforts that aim to disentangle between main and moderating associations of environmental factors on occupational disability have been limited to-date; however, this is important towards a more comprehensive understanding on this topic. Based on the postulated relevance of environmental factors in work functioning (Sandqvist and Henriksson, 2004), we hypothesize that both main and moderating associations for broad work environment profiles will be evident.
4.4 METHODS

Study setting
Participants for this study were injured workers (i.e. workers’ compensation claimants) attending a Upper-Extremity Specialty clinic operated by the Workplace Safety & Insurance Board of Ontario. This represents a population where health-related work limitations are a known concern (Beaton and Kennedy, 2005; Pichora and Grant, 2010; Tang et al., 2009). The Workplace Safety & Insurance Board is a single-source workers’ compensation insurer in the region, funded by employers in Ontario and legislated by the provincial government (www.wsib.on.ca). The Specialty clinics were established to provide specialized clinical consultations for claimants experiencing an atypical recovery course (i.e. insufficient progress after ~6 months, more chronic/complex cases) following soft tissue musculoskeletal injuries (e.g. repetitive strain injuries, acute or cumulative trauma disorders of muscle, tendon, and/or ligaments) or uncomplicated fractures to the upper-limb. At these clinics, consultations are provided by a multidisciplinary team of rehabilitation therapists, orthopaedic surgeons, social workers, case managers to evaluate recovery progress and prognosis, suitability for work, or candidacy for surgery or rehabilitative (e.g. vocational) interventions. Injured workers attending these clinics come from diverse occupational sectors and present with highly varied levels of symptoms (e.g. localized/radiating pain and paresthesia), health status, and work disability (Beaton and Kennedy, 2005; Pichora and Grant, 2010; Tang et al., 2009). Such variability makes this an ideal population for investigating the inter-relationships between worker health, work limitations, and the work environment.

Study population and design
This study involved secondary analysis of cross-sectional data collected from consenting injured workers recruited from one of two Specialty clinics (Toronto and London sites). The data was initially collected between 2004-2005, and involved fielding surveys to injured workers at time of initial clinic attendance. The survey included an array of socio-demographic (e.g. age, gender), health-related (e.g. pain intensity, pain diagram) and work-related variables (e.g. organizational- and job-level factors). Since this study focused on “on-the-job” problems, to be eligible the injured worker must: 1) have been working at initial
Clinic attendance, 2) have understood written English, and 3) have provided written consent for the use of their de-identified survey data for research purposes. A total of 614 initially completed our survey, and of them, \( n=356 \) (58.0\%) met the eligibility criteria for the current study (all those excluded were not working). We further excluded seven participants who did not contribute any data on work characteristics, which led to a final study sample of \( n=349 \).

Study participants were explicitly assured that neither the WSIB nor their employer would have access to their survey responses. The current research protocol was approved by the research ethics board at the Sunnybrook Health Sciences Centre (host site of clinic), St. Michael's Hospital, University of Toronto, and University of Western Ontario.

**Key variables and measures**

**Work characteristics**

To characterize injured workers’ work environment, 12 organizational- and job-level variables assessed in the survey were selected for analysis. These were chosen on the basis of theoretical relevance (i.e. captures key postulated domains of the work environment), available psychometric evidence in musculoskeletal populations, and pragmatic considerations (e.g. no excessive missing values). Organizational-level factors (\( n=4 \)) were represented by the four subscales of the 11-item Organizational Policies and Practices measure (OPP-11), which features domains on safety practices (SP, active safety leadership, safety training, and safety diligence: 3 items), ergonomic practices (EP, design of physical work environment and promoting use of work tools: 1 item), disability management (DM, administrative handling of work injuries, proactive return-to-work programs, and job accommodation initiatives: 5 items), and people-oriented culture (POC, management promoting positive and supportive workplace environment: 2 items) (Amick III et al., 2000a; Habeck et al., 1998; Habeck et al., 1991; Hunt et al., 1993). The reliability and validity of employee assessments of the OPP-11 has been previously demonstrated among injured workers with WRUEdS (Tang et al., 2011d).

The selected job-level factors (\( n=8 \)) comprised variables that inform both the physical and psychosocial aspects of the injured workers’ job. Psychosocial job factors featured key concepts from the Demand-Control-Support model (Johnson and Hall, 1988; Karasek, 1979; Karasek and Theorell, 1990) and were assessed using various Job Content
Chapter 4: Role of the work environment

Questionnaire (JCQ) subscales (Karasek et al., 1998): skill discretion (SD, 6 items), decision authority (DA, 3 items), psychological demands (PsyD, 5 items), psychological pace (PP, 3 items), and supervisor support (SS, 5 items). In addition, co-worker support (CS, 4 items) and mental stress of job (MS, 4 items) were assessed using subscales from the Psychosocial Aspects of Work Questionnaire (PAWQ) (Symonds et al., 1996). To evaluate physical job demands (PhyD), we employed a 4-item checklist that comprised constructs thought to be of high relevance in this population (i.e. involved work tasks that directly tax the upper-limb) (Eriksen et al., 1999). Specifically, these items asked respondents whether their job involved: work with hands over the shoulder level, work in same position over a long time, repetitive stereotypic movements, and/or heavy lifting (the number of checked items is tallied into a score out of four, reflecting higher exposure to such activities). The scoring range of measures for each of the 12 organizational- and job-level factors are described in (Table 4.1)

Worker health factors

To provide representation from both biomedical and psychosocial perspectives of health, upper-limb pain intensity (ULP) and mental health status (MHS) were chosen as specific health factors of interest in our current analysis. In the literature, symptom severity and psychological distress/depressive symptoms have often shown to have robust relationships with work disability outcomes in musculoskeletal populations (Allen et al., 2005; Currie and Wang, 2004; Druss et al., 2000; Feuerstein et al., 2003; Hogg-Johnson and Cole, 2003; Lotters and Burdorf, 2006; Lotters et al., 2006; McIntosh et al., 2000; Schade et al., 1999; van der Giezen et al., 2000; van der Weide et al., 1999; Waddell et al., 2003). As such, it was hypothesized that both will demonstrate a significant association with health-related work limitations. ULP was assessed by the 5-item pain subscale of the Shoulder Pain and Disability Index (SPADI) (Roach et al., 1991), and MHS was assessed by the 5-item mental health subscale from version 2 of the Medical Outcomes Studies Short-Form 36 (SF-36) (Ware Jr., 2000). Evidence supporting the reliability and validity for WRUEDs have been previously demonstrated for both the SPADI (Bot et al., 2004; MacDermid et al., 2006; Roy et al., 2009) and SF-36 (Beaton et al., 1996).
Work limitations

Two measures were specifically chosen to offer contrasting perspectives of health-related work limitations (applied in separate analyses): the 4-item Disabilities of the Arm, Shoulder and Hand Work Module (DASH-W) (Tang et al., 2009) and the 17-item Upper-Limb Work Instability Scale (UL-WIS) (Tang et al., 2011b). Both have evidence of reliability and validity in WRUEDs (Tang et al., 2011b; Tang et al., 2009). Comparing between the two, the DASH-W is a short tool focused on assessing difficulties associated with meeting the physical and output demands of work, while the UL-WIS assesses problems over a broader range of constructs that include functional limitations at work, psychological distress at work, as well as perceptions of sustainability of current work role (collectively referred to as work instability by original scale developers) (Gilworth et al., 2003). For the DASH-W, a 5-point Likert response scale (no difficulty [1] to unable [5]) is provided for each item, and the measure is scored by taking the mean from all items (no missing items allowed), subtracting this value by one, and then multiplying it by 25 for a final score (range: 0-100, 100=most limitations). For the UL-WIS, a dichotomous response option (yes [1] / no [0]) is offered for each item and the scale is initially scored by summing all 17 items. The summed scores were then transformed to Rasch scores based on previous insights (range: 0-100, 100=most work instability) (Tang et al., 2011b). We allowed a maximum of 1 missing item for the UL-WIS (i.e. <10% missing allowed), otherwise the scale was not scored.

Analysis

Descriptive statistics

To describe the study sample, univariate statistics (e.g. n, mean, %) were assessed for key socio-demographic, health-related and work-related variables from our survey. In addition, based on job titles provided, injured workers were classified into one of nine broad occupational sectors based on the National Occupational Classification developed by Human Resources and Skills Development Canada (Human Resources and Skills Development Canada, 2011). Descriptive analyses were conducted with R version 3.0.1.

Latent profile analysis

To differentiate injured workers based on their overall work environment, a latent profile analysis (LPA) (Lazarsfeld and Henry, 1968; McCutcheon, 1987) was conducted to identify
distinct work environment “profiles” based on the 12 selected work characteristics (4 organizational- and 8 job-level factors). This, in turn, also served to classify injured workers into meaningful subgroups (who share similar overall work environment). LPA can be viewed as a special case of cluster analysis (finite mixture estimation) and has been advocated by statisticians as a useful method (McLachlan and Peel, 2000; Titterington et al., 1985). In Mplus, parameters are estimated by maximum-likelihood estimation via the Expectation-Maximization algorithm, where the profile parameter $k$ (i.e. number of distinct profiles) is seen as unknown variable (Muthén, 2001). Our overall approach was to evaluate and compare the sensibility and performance of a range of potential solutions (i.e. consisting of 2-5 profiles). As per current expert recommendation, determining the ideal number of profiles was done based on theoretical plausibility, statistical, as well as interpretative considerations (e.g. proportion of workers classified to each profile) (Muthén, 2004; Muthén, 2001). Since there remains no clear consensus on the ideal statistical criterion, we considered performance on several relevant statistics, including information criteria such as Bayesian Information Criteria (BIC)\(^1\) (Schwartz, 1978) and adjusted BIC\(^2\) (Sclove, 1987), as well as likelihood-based techniques such as the Lo-Mendell-Rubin (LMR) test (Lo et al., 2001) and the Bootstrapped Likelihood Ratio Test (BLRT) (McLachlan and Peel, 2000). Both the LMR and BLRT similarly provide a $p$ value that can be used to determine if the inclusion of an additional profile results in a statistically significant improvement in model fit (i.e. $p<0.05$).

**Multiple-group path analysis**

Following LPA, multiple-group path analyses were conducted to examine main and moderating associations of the work environment profiles on work limitations, in accordance to our proposed hypothesis (Figure 4.1). By this, health-related work limitations was specified as the endogenous variable (i.e. outcome), which was represented by either the DASH-W or UL-WIS (tested separately). ULP and MHS were concurrently specified as exogenous variables (i.e. predictors), given their postulated relationship with work limitations. To enhance interpretability of results, both exogenous variables were centered, such that the model intercepts would represent the estimated level of work limitation (i.e. $1$ BIC $= -2 \log L + p \ln n$, where $\log L$ is the loglikelihood and $p$ is the number of free parameters in the model

\[ n^* = \frac{(n+2)}{24} \]

\(^1\) The adjusted BIC replaces the sample size $n$ in the BIC equation above with $n^*$: $n^* = (n+2)/24$
DASH-W or UL-WIS score) for a worker with both ULP and MHS at sample mean. Lastly, work environment was operationalized as a categorical variable based on distinct profiles/subgroups derived from initial LPA. To affirm that all exogenous and endogenous variables were psychometrically appropriate for use in our path analyses, internal consistency of the associated measures was verified using Cronbach’s alpha (Cronbach, 1951) or Kuder-Richardson Formula-20 (Kuder and Richardson, 1937).

A systematic series of models was tested, and the process was repeated with either the DASH-W or UL-WIS as the endogenous variable. Initially, an unconstrained model was fitted where model parameters (i.e. slopes and intercept) were free to vary across subgroups. Then, a test of moderating association was performed, to see whether the relationship between worker health factors and work limitations varied across subgroups (representing distinct work environment profiles). To this end, we refitted the model where an equality constraint was placed on the corresponding slope parameter across subgroups (i.e. ULP/MHS → work limitations forced to be equal between subgroups). Finally, to evaluate the main association of work environment profiles, we fitted an additional model with an equality constraint applied to the model intercept. Formally, both main and moderating associations were quantified by the magnitude of increase in model chi-square after an equality constraint had been introduced to the relevant parameter(s). A p value based on this chi-square distribution was evaluated to test the null hypothesis of no difference between models where the parameter of interest was unconstrained or constrained. A small p value (e.g. <0.05) would indicate significant degradation of fit for the constrained model, providing evidence that the specific parameter differed between subgroups.

Latent profile analyses and multiple-group path analyses were conducted with Mplus version 6.0 (Muthén and Muthén).
4.5 RESULTS

Characteristics of injured workers
A descriptive summary of socio-demographics, health and work-related variables is provided in Table 4.1. In our sample of injured workers, the mean (±SD) age was 44.9±9.2 (range: 19-68) and 56.2% were female. Based on a pain diagram that assessed 15 upper-body regions, injured workers reported an average of 5.0 (SD=3.8) painful sites, with 41.5% indicating bilateral involvement. The distribution of painful sites was as follows (not mutually exclusive): neck (43.6%), shoulder (75.1%), upper-arm (49.6%), elbow (52.1%), forearm (37.0%), wrist (39.0%), hand (39.8%), upper-back (23.8%), and lower-back (16.9%).

At the sample-level, the SPADI pain subscale mean was 30.8±11.4 (out of 50) and the SF-36 mental health subscale mean was 65.1±21.1 (out of 100). In terms of health-related work limitations, the DASH-W (61.5±26.1) and UL-WIS (58.0±18.0) means were both above the midrange of the respective scales, affirming some concern in this area among injured workers. These two measures were only moderately correlated (Pearson r=0.56), affirming some conceptual distinctiveness between their intended concepts.

Our injured workers were most commonly employed in two particular sectors: trades, transport and equipment operators (24.4%) and processing, manufacturing and utility (33.3%). In general, the sample mean for 12 selected work characteristics was near the midrange of the respective scales, with only a few notable exceptions. OPPs on ergonomic practices (mean=2.2±1.1) were well-below the previously suggested cut-point of ≥3.0 considered to represent a “favorable” level (Tang et al., 2011d). This was also low compared to the other three OPPs domains (mean: 3.1-3.3). Levels of psychological demands (mean=2.7±0.5) and psychological pace (mean=2.8±0.6) also appeared somewhat high, in relation to scale midpoint.

Latent profile analysis
To profile injured workers by the 12 organizational- and job-level factors, several LPA solutions were compared (i.e. 2-5 profiles); however, an unequivocal, single best solution did not emerge based on the statistical criteria considered (Table 4.2). Viewed on its own, the LMR test favored the 2-profile solution, since the 3-profile solution did not represent an improvement based on this statistic. Yet, the BIC, adjusted BIC, and entropy all appeared to
favor the larger profile solutions (i.e. 4- or 5-profiles). Comparing between these larger profiles, the 4-profile solution seemed more optimal overall for several reasons: 1) the 5-profile solution was only marginally more favorable on the BIC, adjusted BIC, and entropy, 2) the 5-profile featured some drawbacks in terms of interpretability, as two of the subgroups consisted of ≤10% of the sample, and 3) the LMR test suggested that the 5-profile solution did not represent a significant improvement on the 4-profile solution (i.e. \( p=0.14 \)). The BLRT criterion, unfortunately, was not informative for class enumeration in our analysis (\( p<0.001 \) for all solutions). We felt that the 2- and 4-profile solutions represented the most viable solutions. Between the two, the 4-profile model could be viewed as a general elaboration of the more parsimonious 2-profile model, but with two additional subgroups that featured more “extreme” work characteristics in general (i.e. Profile #1 [10.0%] & Profile #4 [15.8%]) (Figure 4.2). Although these subgroups were relatively small in proportion, we thought that these variations have theoretical plausibility, as several job stress theories in the literature have posited (at least) four broad job types. For example, the Karasek quadrants from the Demand-Control (DC) model suggest four different broad job types (Karasek and Theorell, 1990), and Job Demands-Resources (JD-R) model have hypothesized likewise (also a 2x2 matrix) (Bakker and Demerouti, 2007)). Taking into account statistical, theoretical, and interpretative considerations, we ultimately considered the 4-profile model to be the best solution. Univariate statistics (mean, SD) of the 12 work characteristics stratified by profiles are presented in Table 4.3.

As can be seen in Figure 4.2, considering only the four organizational-level factors, there was in fact relatively little variation between the four profiles, as a profile that was high on one OPPs domain tended to also be high on other OPPs domains (and vice versa). In contrast, variation across the eight job-level factors between the profiles appeared much more complex. Comparatively, Profile #1 was characterized by high OPPs, control (SD, DA), and support (SS, CS), low physical demands (PhyD), and interestingly, a mix of both low (PsyD) and high psychological demands (PP, MS). For the most part, these were what might be considered fairly favorable characteristics. Profile #2 featured moderately high OPPs, control, and support, physical demands, and moderately high-to-high psychological demands. Profile #3 was characterized by moderately low OPPs, control, support, and physical demands, and moderately low-to-low psychological demands, particularly PP and
MS. Lastly, Profile #4 featured low OPPs, control, and support, and high physical and psychological demands. Collectively, these appeared to be largely unfavorable characteristics.

**Multiple-group path analysis**

Adequate internal consistency was affirmed for all exogenous and endogenous variables in our path analyses. Cronbach’s alphas for the SPADI pain subscale, SF-36 mental health subscale, and the DASH-W were 0.86, 0.86, and 0.95, respectively. KR-20 for the UL-WIS was 0.84.

**DASH-W as the endogenous variable**

In the unconstrained model with work limitations represented by the DASH-W, Profile #1 had the lowest model intercept (53.56) compared with the other three subgroups (60.70-65.70) (Table 4.4). Statistically significant associations ($p<0.05$) between both worker health factors (ULP and MHS) and work limitations (i.e. slope parameters) were generally observed, with the exception of Profile #4, where the association between MHS and DASH-W was not statistically significant. Constraining the corresponding slope parameters between subgroups (to test for moderating associations) did not result in significant degradations in model fit. Specifically, when ULP→DASH-W was constrained, the model chi-square increase was 2.86 (df=3, $p=0.41$), and when MHS→DASH-W was constrained, the model chi-square increase was 2.87 (df=3, $p=0.41$). With these constraints maintained, the model intercept between subgroups was then further constrained to provide a test of main association. This led to a chi-square increase of 6.76 (df=3, $p=0.08$). Collectively, these results suggest no significant main or moderating associations for the work environment profiles.

**UL-WIS as the endogenous variable**

Results were generally similar when work limitations were represented by the UL-WIS (Table 4.4). In the unconstrained model, the model intercept for Profile #1 (52.09) was lower than the other three subgroups (57.28-60.11), and statistically significant associations ($p<0.05$) between worker health factors and work limitations (i.e. slope parameters) were also generally observed. The lone exception was the relationship between ULP and UL-WIS in Profile #4, which was not statistically significant. Constraining the corresponding slope
parameters between subgroups did not result in significant degradations in model fit. Specifically, when ULP→UL-WIS was constrained, the chi-square increase was 5.84 (df=3, $p=0.12$), and when MHS→UL-WIS was constrained, the chi-square increase was 5.54 (df=3, $p=0.14$). With these constraints maintained, the model intercept was then constrained to provide a test of main association. This led to a chi-square increase of 5.34 (df=3, $p=0.15$). Again, collectively, these results suggest no significant main or moderating associations for the work environment profiles.
4.6 DISCUSSION

The study objectives were two-fold, 1) to profile injured workers with a WRUED based on 12 different organizational- and job- level characteristics, and 2) to elucidate the relationship between work environment and extent of health-related work limitations experienced by injured workers. The key study findings were as follows. First, based on LPA results, we found that injured workers can be differentiated according to four unique work environment profiles. Second, against our initial hypothesis, significant main nor moderating associations between work environment profiles and health-related work limitations were not evident, according to findings from our multiple-group path analyses.

Given increased recognition of the relevance of both organizational- and job-level factors in occupational disability (Krause et al., 2001b), a “profiling” strategy was applied to help provide a comprehensive view of injured workers’ work environment. We believe our efforts represent a somewhat novel application, since the use of LPA or other clustering techniques to broadly discern the work environment is rare to-date. Yet, it should also be recognized that “profiling” workers based on multiple work characteristics has theoretical basis from current job stress models. For example, Karasek’s Demand-Control model proposed four broad job types differentiated based on combination of high/low psychological job demands and job control (i.e. the Karasek quadrants) (Karasek and Theorell, 1990). Likewise, the Job Demands-Resources model similarly posited that jobs can be characterized by combinations of high/low job demands and job resources (Bakker and Demerouti, 2007). Unique for our current work is that we have emphasized a broader set of work characteristics with postulated relevance for occupational disability, which includes both organizational- and job- level factors.

Overall, we thought that two aspects of our LPA results stood out. First, all four profiles featured certain work characteristics that were challenging, at least in relation to other profiles. Case in point, although most work characteristics appeared quite favorable in Profile #1, its levels of psychological pace and mental stress of job might be considered suboptimal compared to Profile #3. This is perhaps a reflection of the “real-world” reality where all workplaces are likely to have unique strengths and limitations. A second interesting feature
was that it appears certain work characteristics can be differentiated into more categories (or “levels”) than others. On one hand, only two discernible “levels” were revealed for psychological pace and mental stress of job (i.e. Profile #1, 2, and 4 [high] vs. Profile #3 [low]), which is inline with existing theorizations from the Demand-Control model or the Job Demands-Resources model (e.g. 2x2 matrices based on combinations of low/high characteristics). But on the other hand, it was also apparent that four different “levels” can be clearly discerned for some characteristics, like skill discretion and supervisor support. This suggests that a simple low/high distinction may be insufficient (i.e. too coarse), and that potentially greater variation for such characteristics deserves to be considered.

Some reflection on the potential advantages and disadvantages of profiling injured workers using an LPA technique is perhaps worthwhile. First and foremost, we thought that a main strength was that it helped foster a more holistic view of the overall work environment. This was considered meaningful since workers do not experience a specific work characteristic in isolation, but rather, they experience an “environment” characterized by the concurrent influence of multiple factors. In some sense, this is not dissimilar to the more traditional “job type/category” variable that can also discern workers from a broader work environment context. What is perhaps most unique about LPA or other clustering approaches is that these are data-driven techniques, where individual researchers are afforded the choice of specific variables to include, in order to suit different research contexts and needs. Moreover, if needed, insights on cut-scores for individually measured variables can also be gained with these approaches by examining observed score thresholds that separate the subgroups. This can be informative from the perspective of score interpretability. Certainly, some challenges associated with the application of LPA techniques remain. Perhaps the most important issue is that a definitive statistical criterion toward profile/class enumeration remains of some debate (Muthén, 2004; Muthén, 2001). Related to this issue, there is also limited current understanding on how methodological factors (e.g. available sample size) might impact the performance of various statistical criteria, and ultimately, how this might influence interpretation of LPA results. As can be seen in our analysis, an obvious best solution may not emerge as considerations from different perspectives (i.e. statistical, theoretical, and interpretative) can sometimes conflict. Finally, it should be emphasized that LPA techniques are designed to highlight comparative
differences (between profiles), and therefore, it is important to exercise caution when interpreting results in absolute terms (i.e. “low/high” and “favorable/unfavorable” are relative against other profiles).

We thought three main reasons may have contributed to the lack of significant relationships between the work environment profiles and the extent of health-related work limitations. First, this is a unique population in that it featured only more complex/chronic WRUED cases (e.g. almost 70% scored ≥25 out of 50 on the SPADI pain subscale). It may be the case where health problems are more severe, the overall work environment becomes less influential of a factor (i.e. no main association). In a similar vein, in such a population, perhaps only drastic changes in work environment are likely to alter the relationship between worker health factors (ULP/MHS) and the extent of work limitations (i.e. no moderating association). Second, the distribution and nature of the derived work environment profiles may have been a factor. The two largest profiles (#2 & #3) accounted for almost 75% of the sample, but notably, these did not feature particularly large differences across many of the work characteristics assessed. In this regard, the most extreme differences were found between Profiles #1 & #4, which constituted only 25% of the study sample. Moreover, these latter subgroups were also characterized by fairly small sample sizes, which contributed to somewhat larger standard errors in the parameter estimates. Third, it is possible that broad differences in the work environment are simply not sufficiently precise for the purpose of revealing the nature of worker-workplace interactions. In any case, the current findings do not rule out the possibility of such interactions at the level of a specific characteristic that contributed to the broad profiles. Future efforts to explore this possibility may be worthwhile to further advance understanding on the role of the work environment, as shed additional light on the nature of worker-workplace interactions.

The key strengths of this study include: 1) a focus on WRUEDs, which despite high prevalence remains a relatively under-studied population (e.g. compared to low back pain), 2) an examination of health-related work limitations, which are an aspect of occupational disability that has received relatively less attention compared to more traditional outcomes like return-to-work or disability duration, 3) an exploration of LPA as a novel method to facilitate a holistic view of the work environment, and finally, 4) a rare investigation of both
main and moderating associations to more thoroughly elucidate the role of the work environment. Several study limitations should also be considered. Generalizability of study results may be limited since our sample represents only the more complex/chronic WRUED cases. In terms of the work environment profiles, the distribution of occupational sectors among injured workers may not be representative of the broader labor market. Clinical diagnosis of WRUEDs was not available from our surveys, and therefore, we were unable to describe our sample from this perspective. Two methodological issues pertaining to the path analyses should be noted. First, we did not incorporate measurement models into our analyses, as the available sample size was a limiting factor; notwithstanding, reliable and well-validated instruments were used applied in the study. Second, it is important to recognize that our a priori moderation framework was specifically developed for the purpose of this study. The fact that only ULP and MHS were specified as exogenous variables was not intended to suggest that these are the only putative determinants of health-related work limitations among WRUEDs.

Conclusions
Using latent profile analysis, we differentiated injured workers with WRUEDs into one of four unique work environment profiles based on a set of 12 organizational and job-level characteristics. Against our initial hypothesis, however, work environment profiles did not demonstrate main nor moderating associations on the extent of health-related work limitations. This suggests broad differences in the work environment do not represent a significant contextual factor of work limitations. To gain further insights on the nature of worker-workplace interactions, future research may consider developing and testing hypotheses focused on elucidating the role of more specific work characteristics.
Chapter 4: Role of the work environment

4.7 TABLES AND FIGURES

![Diagram]

Figure 4.1 Hypothesis of “worker-workplace interactions” to be investigated in the current study
### Table 4.1 Socio-demographic, health, and work-related characteristics of the study sample ($n=349$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Instrumentation</th>
<th>Scale Range</th>
<th>$n$</th>
<th>Mean</th>
<th>SD</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Socio-demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>continuous</td>
<td>sample range = 19-68</td>
<td>346</td>
<td>44.9</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>single-item</td>
<td>level: female</td>
<td>345</td>
<td>·</td>
<td>·</td>
<td>56.2</td>
</tr>
<tr>
<td>Marital status</td>
<td>single-item</td>
<td>level: married or common law</td>
<td>344</td>
<td>·</td>
<td>·</td>
<td>70.3</td>
</tr>
<tr>
<td>Education</td>
<td>single-item</td>
<td>level: completed post-secondary level</td>
<td>344</td>
<td>·</td>
<td>·</td>
<td>33.1</td>
</tr>
<tr>
<td><strong>Health and disability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of painful sites</td>
<td>pain diagram</td>
<td>15 upper-body regions assessed</td>
<td>340</td>
<td>5.0</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>Upper-limb pain intensity</td>
<td>SPADI pain subscale (5 items)</td>
<td>0-50 (50=high pain)</td>
<td>348</td>
<td>30.8</td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td>Number of comorbidities</td>
<td>modified Sangha checklist (11 conditions assessed)</td>
<td>0-11 (# of comorbidities)</td>
<td>349</td>
<td>0.8</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Mental health status</td>
<td>SF-36 mental health subscale (5 items)</td>
<td>0-100 (100=high mental health)</td>
<td>339</td>
<td>65.1</td>
<td>21.1</td>
<td></td>
</tr>
<tr>
<td>At-work disability</td>
<td>DASH-W (4 items)</td>
<td>0-100 (100=high at-work disability)</td>
<td>334</td>
<td>61.5</td>
<td>26.1</td>
<td></td>
</tr>
<tr>
<td>Work instability</td>
<td>UL-WIS (17 items)</td>
<td>0-100 (100=high work instability)</td>
<td>302</td>
<td>58.0</td>
<td>18.0</td>
<td></td>
</tr>
<tr>
<td><strong>Organizational-level factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safety practices</td>
<td>OPP-11 subscale (3 items)</td>
<td>1-5 (5=most favorable)</td>
<td>340</td>
<td>3.3</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Ergonomic practices</td>
<td>OPP-11 subscale (1 item)</td>
<td>1-5 (5=most favorable)</td>
<td>340</td>
<td>2.2</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Disability management</td>
<td>OPP-11 subscale (5 items)</td>
<td>1-5 (5=most favorable)</td>
<td>320</td>
<td>3.1</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>People-oriented culture</td>
<td>OPP-11 subscale (2 items)</td>
<td>1-5 (5=most favorable)</td>
<td>341</td>
<td>3.1</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td><strong>Job-level factors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skill discretion</td>
<td>JCQ subscale (6 items)</td>
<td>1-4 (4=high skill discretion)</td>
<td>336</td>
<td>2.5</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Decision authority</td>
<td>JCQ subscale (3 items)</td>
<td>1-4 (4=high decision authority)</td>
<td>336</td>
<td>2.5</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Supervisor support</td>
<td>JCQ subscale (5 items)</td>
<td>1-4 (4=high support)</td>
<td>303</td>
<td>2.8</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>Co-worker support</td>
<td>PAWQ subscale (4 items)</td>
<td>1-5 (5=high support)</td>
<td>340</td>
<td>3.7</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Physical demands</td>
<td>checklist (4 items)</td>
<td>0-4 (4=high demands)</td>
<td>349</td>
<td>2.4</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Psychological demands</td>
<td>JCQ subscale (5 items)</td>
<td>1-4 (4=high demands)</td>
<td>317</td>
<td>2.7</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Psychological pace</td>
<td>JCQ subscale (3 items)</td>
<td>1-4 (4=high pace)</td>
<td>337</td>
<td>2.8</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Mental stress of job</td>
<td>PAWQ subscale (4 items)</td>
<td>1-5 (5=high mental stress)</td>
<td>340</td>
<td>3.4</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

SPADI, Shoulder Pain and Disability Index; SF-36, Medical Outcomes Studies Short-Form 36 (version 2); JCQ, Job Content Questionnaire; PAWQ, Psychosocial Aspect of Work Questionnaire; OPP-11, 11-item Organizational Policies and Practices Scale; DASH-W, Disabilities of the Arm, Shoulder and Hand Work Module; UL-WIS, Upper-Limb Work Instability Scale.
Table 4.2 Model fit statistics estimated from latent profile analyses of 12 organizational- or job-level factors (multiple solutions compared)

<table>
<thead>
<tr>
<th>Model fit statistics</th>
<th>Number of Profiles ($k$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Loglikelihood</td>
<td>-4522.23</td>
</tr>
<tr>
<td># of parameters</td>
<td>37</td>
</tr>
<tr>
<td>2*Loglikelihood (vs. model with 1 profile)</td>
<td>224.08</td>
</tr>
<tr>
<td>BIC</td>
<td>9261.09</td>
</tr>
<tr>
<td>Adjusted BIC</td>
<td>9143.72</td>
</tr>
<tr>
<td>LMR test p</td>
<td>0.002</td>
</tr>
<tr>
<td>BLRT test p</td>
<td>0.000</td>
</tr>
<tr>
<td>Entropy</td>
<td>0.83</td>
</tr>
</tbody>
</table>

**Estimated Proportions**

<table>
<thead>
<tr>
<th></th>
<th>Profile #1</th>
<th>Profile #2</th>
<th>Profile #3</th>
<th>Profile #4</th>
<th>Profile #5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.52</td>
<td>0.34</td>
<td>0.10</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.49</td>
<td>0.30</td>
<td>0.41</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.36</td>
<td>0.33</td>
<td>0.36</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.16</td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BIC, Bayesian Information Criteria; LMR, Lo-Mendell-Rubin test; BLRT, Bootstrapped Likelihood Ratio Test.

**Brief notes on interpretation of model fit statistics:**
BIC (smaller = better fit): $-2 \log L + p \ln n$, where $\log L$ is the loglikelihood and $p$ is the number of free parameters in the model.
Adjusted BIC (smaller = better fit): replaces the sample size $n$ in the BIC equation above with $n^*$: $n^* = (n+2)/24$.
VLMR test: low $p$ = better fit than model with $k$-1 profile.
BLRT test: low $p$ = better fit than model $k$-1 profile.
Entropy: larger = better fit.
Chapter 4: Role of the work environment

**Figure 4.2** Four-profile solution derived from latent profile analysis of 12 organizational- or job-level factors

Organizational-level factors: SP, safety practices; EP, ergonomic practices; DM, disability management; POC, people-oriented culture. Job-level factors: SD, skill discretion; DA, decision authority; SS, supervisor support; CS, co-worker support; PhyD, physical demands; PsyD, psychological demands; PP, psychological pace; MS, mental stress of job.

Grey shaded area on the right comprised four work characteristic factors where higher scores could be viewed as generally less favorable, which is in contrast with the orientation of the other eight factors.

Standardized means of work characteristics are presented, raw means can be found in Table 4.3.
Table 4.3 Work characteristics stratified by broad work environment profiles (n=4) derived from latent profile analysis

| Profile #1 | SP (1-5) | 4.6 (0.4) | EP (1-5) | 4.0 (0.8) | DM (1-5) | 4.2 (0.6) | POC (1-5) | 4.6 (0.5) | SD (1-4) | 3.2 (0.4) | DA (1-4) | 3.4 (0.5) | SS (1-4) | 3.5 (0.8) | CS (1-5) | 4.5 (1.3) | PhyD (0-4) | 1.7 (0.4) | PsyD (1-4) | 2.4 (0.5) | PP (1-4) | 3.2 (0.5) | MS (1-5) | 4.0 (0.7)
| Profile #2 | 3.5 (0.8) | 2.4 (1.0) | 3.3 (0.7) | 3.5 (0.7) | 3.2 (0.4) | 2.8 (0.5) | 2.7 (0.6) | 2.8 (0.7) | SD (1-4) | 2.7 (0.4) | DA (1-4) | 2.8 (0.5) | SS (1-4) | 3.0 (0.6) | CS (1-5) | 4.1 (1.2) | PhyD (0-4) | 2.5 (0.4) | PsyD (1-4) | 2.7 (0.4) | PP (1-4) | 3.1 (0.6) | MS (1-5) | 4.0 (0.4)
| Profile #3 | 3.2 (0.8) | 1.8 (0.9) | 3.0 (0.7) | 2.8 (0.8) | 3.2 (0.4) | 2.0 (0.5) | 1.9 (0.6) | 2.2 (0.6) | SD (1-4) | 2.4 (0.5) | DA (1-4) | 2.1 (0.6) | SS (1-4) | 2.1 (0.8) | CS (1-5) | 3.1 (1.1) | PhyD (0-4) | 2.7 (0.5) | PsyD (1-4) | 3.1 (0.5) | PP (1-4) | 3.1 (0.5) | MS (1-5) | 3.8 (0.7)
| Profile #4 | 2.3 (0.9) | 1.5 (0.8) | 2.5 (0.8) | 2.0 (0.6) | 2.4 (0.5) | 2.1 (0.6) | 2.1 (0.6) | 2.1 (0.8) | SD (1-4) | 2.4 (0.5) | DA (1-4) | 2.1 (0.6) | SS (1-4) | 3.1 (0.8) | CS (1-5) | 3.1 (1.1) | PhyD (0-4) | 3.1 (0.5) | PsyD (1-4) | 3.1 (0.5) | PP (1-4) | 3.1 (0.5) | MS (1-5) | 3.8 (0.7)

Organizational-level factors: SP, safety practices; EP, ergonomic practices; DM, disability management; POC, people-oriented culture.

Job-level factors: SD, skill discretion; DA, decision authority; SS, supervisor support; CS, co-worker support; PhyD, physical demands; PsyD, psychological demands; PP, psychological pace; MS, mental stress of job.
Table 4.4 Multiple-group path analysis examining the main and moderation associations of work environment profiles

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unconstrained model</th>
<th>Constrained model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Profile #1 (n=35, 10.0%)</td>
<td>Profile #2 (n=144, 41.3%)</td>
</tr>
<tr>
<td></td>
<td>b (SE)</td>
<td>b (SE)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL-WIS as endogenous variable</td>
<td>ULP→UL-WIS</td>
<td>0.82 (0.26) **</td>
</tr>
<tr>
<td></td>
<td>MHS→UL-WIS</td>
<td>-0.43 (0.15) **</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>52.09 ***</td>
</tr>
</tbody>
</table>

ns = p>0.05, * = p<0.05, ** = p<0.01, *** = p<0.001.

unstandardized model parameters are presented

unconstrained model: model parameters free to vary between profiles; constrained model: equality constraint applied to corresponding parameter between profiles, resulting model chi-square difference was assessed (smaller p indicates the parameter of interest differs across groups)

† test of equality of model intercepts were conducted with the slope parameters already constrained since the latter did not differ significantly between subgroups

note also that our unconstrained models are saturated and have a chi-square of zero (i.e. has zero degrees of freedom, all possible coefficients are estimated)
5.1 SUMMARY OF KEY FINDINGS

The thesis comprised three studies that examined topics on the measurement and determinants of work disability associated with work-related upper-extremity disorders (WRUEDs). All three studies focused on a common population – workers’ compensation claimants attending a WSIB Upper-Extremity Specialty clinic in Ontario. Generally, attendees of these Specialty clinics represent more chronic/complex cases of WRUEDs characterized by an atypical recovery course following initial injury. Importantly, such cases are potentially highly burdensome to various stakeholders, and therefore, represent a population deserving of research attention. Since the ability to clearly differentiate between specific forms of WRUEDs remains elusive (Katz et al., 2000), WRUEDs were broadly considered in these studies, inclusive of both “specific” and “non-specific” conditions with varying extent of anatomical involvement in the upper-extremities (e.g. local or diffuse pain). The key findings from the three studies are summarized here.

Objective #1 (Chapter 2) assessed the factorial validity of the Work Limitations Questionnaire (WLQ-25) among a sample of 2262 injured workers. As one of the more popular and well-tested measures intended for various health conditions, it was anticipated that the WLQ-25 would demonstrate strong psychometric performance. Although several of the factorial validity criteria were met, specific areas of underperformance were also revealed. Most notably, lower-than-expected inter-factor correlations involving the physical demands subscale, raising concerns about the varied orientation of instructions provided to respondents across the different WLQ-25 subscales, and the possibility of systemic measurement error. In addition, from a statistical perspective (i.e. model fit), a 5-factor structure (MI expanded into two smaller subscales) performed significantly better than the traditional 4-factor structure (MI intact), providing some support for an alternative and perhaps an even more optimal subscale structure for the WLQ-25.

Objective #2 (Chapter 3): examined the predictive ability of the Upper-Limb Work Instability Scale (UL-WIS) for subsequent transitioning out of work among injured workers who were initially at work upon Specialty clinic attendance. The main study finding was that the UL-WIS, which quantifies WI – defined as a state of “mismatch between work ability
and demands” (Gilworth et al., 2003) – was shown to have independent predictive ability for subsequent transitioning out of work. This finding is significant not only because it helped validate the concept of WI (i.e. purported to represent risk for adverse employment outcome), but it also shed light on the connection between two forms of work disability (i.e. “presenteeism” as a precursor of “absenteeism”, so to speak). As such, a potential implication is that the extent of WI may warrant attention during the management of injured workers. In this study, transitioning out of work occurred quite frequently (~50% of sample exited work at some point over the one year study period), which affirmed concerns for the sustainability of work in this population as previously suggested (Baldwin and Butler, 2006; Keogh et al., 2000; Mayer et al., 1999).

Objective #3 (Chapter 4) investigated the relationship between the work environment and health-related work limitations among injured workers. This study initially explored the use of latent profile analysis to identify distinct work environment profiles based on a set of 12 organizational- and job-level characteristics. Then, applying initial LPA results, multiple-group path analyses were conducted to test whether membership in a specific work environment profile is related to the extent of health-related work limitations. LPA revealed four distinct work environment profiles; however, against initial hypothesis, significant main or moderating associations of work environment profiles on health-related work limitations did not emerge, suggesting that broad difference in the work environment was not an influential factor.

5.2 CONTRIBUTIONS TO MAIN THEMES
This thesis was designed to contribute understanding to two main themes: the 1) measurement and 2) determinants of work disability associated with WRUEDs.

New insights on the measurement of work disability (theme #1)
The ability to measure various aspects of work disability is of fundamental importance for progress in research. Moreover, different manifestations of work disability also represent meaningful outcomes in the management of injured workers in clinical settings. With increased recognition to the relevance of health-related work limitations and a growing
interest to measure this concept, research to affirm the psychometric properties of measurement instruments is specific populations and/or settings is essential. For end-users, such evidence would help instill confidence that these instruments are indeed trustworthy (e.g. reliable and valid) for various applications. Some progress was made in this area, as limited psychometric testing of measures of health-related work limitations in WRUEDs have been conducted to-date. This thesis informed different aspects of validity for two different instruments in this population. Specifically, **Objective #1** informed the factorial validity of the WLQ-25 and **Objective #2** informed the predictive validity of the UL-WIS.

For the WLQ-25, the current work directly builds upon previous work that has informed other properties of this measure in WRUEDs, such as reliability (Beaton and Kennedy, 2005; Tang et al., 2009), construct validity (Beaton and Kennedy, 2005; Roy et al., 2011; Tang et al., 2009), and patient preference (Tang et al., 2009). This is a popular measure, and is generally well-tested compared to its peers (Tang et al., 2011a; Tang et al., 2013c). Almost surprisingly, however, the factorial validity of the WLQ-25 has been given very little research attention since its original conception (Lerner et al., 2001), and thus the current work helps to fill an important gap.

For the UL-WIS, the current work represents a somewhat novel addition to the field, since the family of WIS measures and their intended concept of WI are not yet well-known, and the relevance of WI for injured workers is only beginning to be explored. The observed relationship between UL-WIS and poor sustainability of work is meaningful, since it precisely verified the central argument for the relevance of WI (i.e. leads to adverse employment outcomes) (Gilworth et al., 2003). To our knowledge, the study is also the first test of UL-WIS's predictive validity (or any measurement property) since this tool has been developed for this population (i.e. refined from the original RA-WIS) (Tang et al., 2011b). Moreover, this study complements existing work on the RA-WIS that has similarly demonstrated an ability to predict adverse employment outcomes among workers with arthritis.

Practical implications for the application of the WLQ-25 and UL-WIS in chronic/complex WRUEDs should be considered. Unfortunately, suboptimal performance of the WLQ-25 in
Objective #1 raised important concerns about the scores derived from the PD subscale, and for this reason, data on the WLQ-25 was not considered for use in subsequent studies in the thesis. With this revelation, it also follows that one must exercise care when interpreting existing findings derived from the WLQ-25 where this very issue could have relevance (i.e. different orientation of instructions for different subscales). That said, looking ahead, this should be an easily correctable issue, and as discussed, it is expected that factor analytic performance of the WLQ-25 would improve upon current results if appropriate corrections are made to the measure.

Can the UL-WIS be used for prognostication? Yes, but perhaps with some important caveats. While evidence from Objective #2 lent support for the predictive validity of the UL-WIS, its overall predictive accuracy as a standalone tool was perhaps rather modest. Therefore, if it were to be used, the strengths and limitations in predictive accuracy for different parts of the tool need to be recognized. Specifically, one must be especially cognizant of the fact that the determined cut-scores were selected based on specific strengths (i.e. high PPV for the moderate/high threshold, high NPV for the low/moderate threshold), and therefore, we believe the tool may be best utilized with this recognition. This also implies that the UL-WIS may not be particularly informative for workers scoring between the two cut-points, which can potentially be a sizable proportion of a given sample (e.g. 82.5% of the sample from Objective #2 scored between the cut-points). To summarize, the main implications based on Objectives 1 & 2 are that while the UL-WIS is affirmed to have some ability to predict poor sustainability of work among WRUEDs, some adaptations may be needed for the WLQ-25 (in its current form) before it is suitable for use.

New insights on the determinants of work disability (theme #2)
Reducing work disability among injured workers with WRUEDs is a goal shared by various stakeholders (e.g. workers, clinicians, employers, insurers). Towards this goal, a better understanding of the determinants of work disability is needed, which involves not only an understanding of the pertinent prognostic/contextual factors, but also the mechanisms by which such factors exert their influence (e.g. mediations and moderations). Such insights can potentially contribute in a number of ways, for example, the identification of key intervention targets in the management of injured workers, or the development of prognostic tools to identify subsets of injured workers most at risk for continuing or
worsening work disability. As discussed in Chapter 1, notable gaps of current literature include 1) a lack of attention on outcomes that concern the sustainability of work or health-related work limitations, and 2) a lack of insights on the nature and relevance of worker-workplace interactions, which has been postulated as an important basis of occupational disability (Sandqvist and Henriksson, 2004).

While both Objectives #2 & #3 contributed to thesis theme #2, it is important to recognize that these studies differ in two notable ways. First, different aspects of work disability were considered as study outcomes (i.e. sustainability of work [Objective #2], health-related work limitations [Objective #3]). Second, while Objective #2 was designed to address the issue of prognostication (i.e. a practical application), Objective #3 was somewhat less practical in comparison, as its primary aim was to better understand a “phenomenon” (i.e. health-related work limitations), so to speak. Yet, one might also consider that these two objectives share a subtheme – in their unique ways, they both explored the interaction between the worker and workplace. The main prognostic factor in Objective #2, WI, arguably, has parallels with the idea of P-E fit from earlier job stress theorizations (i.e. high WI reflects poor “fit” between worker and workplace) (Caplan et al., 1975; Harrison, 1985). Current results suggest that WI is deserving of some attention in the management of injured workers with a WRUED (e.g. part of patient assessment, target of intervention), and also that reducing WI may be an important goal towards improved work sustainability. Objective #3, in contrast, touched on the same subtheme from a different perspective. This was done through formal testing of statistical interactions (i.e. moderation hypothesis involving broad work environment profiles). While this approach does not directly speak to the notion of “fit” as in Objective #2, a positive result in this case would have supported a meaningful “synergism” between worker and workplace factors as it pertains to the basis of health-related work limitations associated with WRUEGs. Against our expectation, however, significant moderation associations did not emerge, and as such, it must be concluded that worker-workplace interactions remain unverified in this particular context. As previously discussed, a holistic view of the work environment was specifically adopted in Objective #3, which meant that worker-workplace interactions involving more specific work characteristics (that make up the broad work environment profiles) have not been ruled out in our current testing. To summarize, Objectives #2 & #3 contributed new
insights on the determinants of work disability, particularly, in terms of the relevance of worker-workplace interactions. The cumulative results suggest that while this concept appears to have implications for the sustainability of work (Objective #2), its relevance as a basis of health-related work limitations remains to be ascertained (Objective #3).

5.3 OVERALL STRENGTHS AND LIMITATIONS

1) Diversity and versatility of measures of health-related work limitations

Several strengths of the overall thesis may be considered. Cumulatively, the three studies applied a number of measures that provide diverse perspectives on health-related work limitations (i.e. WLQ-25, DASH-W, UL-WIS). Notably, the UL-WIS was applied as both a prognostic factor (Objective #2) and an “outcome” (Objective #3), recognizing that high WI is a “problem” in and of itself, but is also a precursor for arguably a more troubling form of work disability (i.e. transitioning out of work). This duality provides an indication of the potentially versatility of this class of measures. Although a direct comparison between the performance of different measures was not a current aim, some insights on this topic already exist in the literature (Beaton et al., 2010; Tang et al., 2013b; Tang et al., 2009). The most relevant of these studies featured a direct comparison of the reliability, construct validity, and respondent preference between the DASH-W, RA-WIS, WLQ-16, and the SPS-6 among 80 injured workers with WRUEDs. The study concluded that, comparatively, the WLQ-16 was strongest psychometrically overall while the RA-WIS was most preferred by respondents. Overall, it is anticipated that this body of thesis work will help raise the general awareness of health-related work limitations as a relevant issue among WRUEDs, and some of the available instruments to quantify this construct. Increasingly, psychometricians and clinimetricians are recognizing the need for measurement studies of high methodological quality. It is anticipated that both Objectives #1 & #2 will fair reasonably well when appraised against existing quality criteria in the literature, such as the COConsensus-based Standards for the selection of health Measurement INstruments (COSMIN) (Mokkink et al., 2010).
2) Disease- and phase-specificity
A number of scholars have argued for the importance of considering phase-specificity in occupational disability research, since unique factors may be relevant at different stages of recovery (i.e. change in the direction and/or size of effect over the course of disabling condition) (Dasinger et al., 2000; Krause et al., 2001a; Krause et al., 2001b; Krause and Ragland, 1994; Lotters and Burdorf, 2006; Sinclair et al., 1998). Moreover, a disease-specific view may also be important since unique factors may be involved in different types of occupational injuries, and thus the need for tailored treatment and rehabilitation strategies (Krause et al., 2001b; Lotters et al., 2006). As such, a strength of this thesis work is the opportunity to focus on a relatively homogenous set of injured workers from several perspectives: common location of injury (i.e. upper-extremities), stage of recovery (all chronic/complex WRUEDs, >6 months post-injury), and macro-environmental factors (i.e. all receiving clinical consultations from WSIB Specialty clinics - same workers’ compensation system, socio-political, legal, and economic climate). Arguably, research attention to this subset of injured workers is also particularly worthwhile in light of the fact that the most chronic/complex cases have historically shown to be responsible for a large proportion of the costs and burden to the workers’ compensation system (Baldwin and Butler, 2006; Feuerstein et al., 1998; Hashemi et al., 1998).

3) Application of innovative analytic approaches
Another strength of this thesis is that it features several analytic strategies that appear to be relatively unique to occupational disability research, thus contributing methodological novelty in the field. To recap, Objective #1 applied CFA techniques, including a direct comparison of nested models to compare performance between alternative subscale organizations of the WLQ-25 (4 vs. 5-factor models). In an effort to aid in the interpretability of the UL-WIS, Objective #2 explored predictive cut-scores based on estimated PPVs and NPVs across the scale range, supplementing findings derived from traditional multivariable regressions. And lastly, Objective #3 was highlighted by the use of two relatively novel approaches, namely, latent profile analysis, to identify distinct work environment profiles based on heterogeneity across a set of 12 workplace characteristics, and multiple-group path analysis, to provide a rare, formal testing of interactions between worker and workplace factors.
4) Selection bias due to sampling strategy

General limitations for this thesis work should also be considered. Convenience sampling was used throughout, as study recruitment depended on voluntary participation of injured workers. As such, these were not completely random samples, which introduce some risk for selection bias. If the likelihood of study participation was related to certain worker characteristics, then the resulting sample would not have been fully representative of the target population. In the case where such characteristics are also related to the research outcome of interest, internal validity of the results could also be threatened. To ascertain the representativeness of the convenience sample, comparisons with population data is most informative; however, such data is not often readily available. Nonetheless, we believe the risk of selection bias from sampling is low for three main reasons: 1) within the data collection period, all eligible Specialty clinic attendees were offered the opportunity to participate in the study, 2) survey questionnaires pose low risk and burden to the respondents, and are not known to strongly or systematically discriminate any particular subset of injured workers from participation, and 3) high participation rate (>85%) has been routinely achieved based on previous research experience at the Specialty clinics.

5) Use of self-report survey data

In occupational disability research, the use of self-report (survey) data or administrative (claims) data is common. Both methods are generally feasible in terms of cost and administrative burden. However, surveys offer one key advantage: the ability to collect more complex variables (e.g. perception of health, job characteristics), which is paramount for research focused on elucidating relationships between work and health constructs. For this reason, all studies in this thesis relied solely on self-report data. Yet several fundamental drawbacks with this method of data collection should be recognized. Two of the most common problems of survey questionnaires are the potential for inaccurate reporting (e.g. due to recall bias/error, social desirability bias, misinterpretation of questions) and missing data (e.g. partial completion of questionnaire, follow-up attrition). While inaccurate reporting can obscure the “true” relationship from emerging, significant missing data can raise concerns about selection bias. As discussed, the issue of inaccurate report was particularly relevant for Objective #1 (CFA of WLQ-25), where misinterpretation of PD items was speculated a key reason for the suboptimal factorial analytic performance of the WLQ-25.
Chapter 5: Synthesis

Missing data was also a non-trivial issue in our dataset, particularly missing follow-up surveys. The issue probably had the largest impact for **Objective #2** (Predictive ability of UL-WIS), as only \( n = 280/356 \) provided sufficient data to be definitively classified into one of two levels in our primary study outcome. Considerations for respondent burden dictate surveys are fielded at reasonable time intervals to the same participants. This also limited our ability to acquire more precise information on work transitions (i.e. timing of transitions).

6) **Limits to generalizability of results**

The flip side of focusing on a highly-specific population is the potential limits on generalizability of research results to other populations. As such, it remains to be ascertained whether study findings are generalizable to injured workers from other contexts (e.g. non-WRUEDs, from regions with different workers’ compensation systems), or even non-compensated workers who have sustained similar work-related injuries. Clearly, this cohort is a fairly unique group, and even more so considering this thesis also only focused on injured workers who are at work at time of initial Specialty clinic assessment, which make up only a subset (~50-60%) of all clinic attendees. This of course is by design given the common emphasis of health-related work limitations across the three studies – a concept not directly relevant for those who are off-work.

5.4 **FUTURE DIRECTIONS**

1) **Re-evaluation of the “corrected” WLQ-25**

Results from **Objective #1** suggest that corrections to the original WLQ-25 are needed and that retesting of a corrected version of the measure may be worthwhile. Several types of investigations may be useful, for example, re-evaluating inter-subscale correlations or conducting a CFA of the corrected version in a similar cohort, or perhaps a more sophisticated (but less feasible) “measurement randomized controlled trial” to compare the psychometric performance between the original and corrected versions of the instruments. Since the WLQ-25 is a popular measure, and the issue raised was thought to have broad relevance (i.e. non-population-specific ) and important for potential end-users to recognize, we have recently engaged Dr. Debra Lerner (copyright holder of WLQ-25) about this issue and potential changes to the measure in order to address this limitation. In subsequent work
outside of this thesis, we have applied a corrected version of the WLQ (called “adapted WLQ-25”) for additional testing among workers with arthritis, where the orientation of the subscale instructions is consistent throughout the questionnaire. Preliminary results have shown that the inter-factor correlations are highly consistent with expectations \((r=0.71-0.87)\) (Tang et al., 2013c), and this corrected version of the WLQ appears promising in terms of addressing the key concern raised in Objective #1 of this thesis. These findings were recently presented to Outcome Measures in Rheumatology (OMERACT), a consensus-based international collaboration focused on recommending outcome measures for use in arthritis and other rheumatic conditions.

2) Informing other psychometric properties of measures of health-related work limitations

The number of available measures in this area is growing at an impressive clip in the literature. There were 24 tools identified in a recent review (Beaton et al., 2009), and more are introduced in the literature on a regular basis. With perhaps a few notable exceptions, the amount of testing to affirm their basic psychometric properties has not kept pace, and some reviews have also revealed less-than-favorable ratings on the methodological quality of available psychometric studies on these measures (Abma et al., 2012). In addition, while most attention has been paid to the reliability and validity of these tools, as a whole, there remains very limited evidence in the areas of responsiveness and score interpretability. Insights on both these properties are crucial for instruments to be applied for evaluative purposes (e.g. as outcome measures in trials or cohort studies requiring pre- vs. post-comparisons). In addition, such testing is increasingly mandated by regulatory agencies like Health Canada and the FDA where patient-reported outcomes are to be used to support claims in approved medical product labelling (U. S. Department of Health and Human Services FDA Center for Drug Evaluation Research et al., 2006). In terms of score interpretability, research efforts to establish score benchmarks and/or the “minimally important difference” are important. Such insights enable results from clinical trials or longitudinal cohort studies (e.g. changed scores) to be assessed beyond purely statistical criteria (i.e. level of statistical significance can be influenced by sample size). These are potential future research avenues to advance the measurement of health-related work limitations, and to facilitate their application in occupational disability research.
3) Towards a better prognostic tool: building on the UL-WIS?

Although the UL-WIS demonstrated independent predictive ability for a subsequent transition out of work, as a standalone tool, some limitations in predictive accuracy was also evident. To build off of the current thesis work, several future research avenues may be worthwhile. First, the comparability of UL-WIS’s predictive performance with other existing tools designed for a similar purpose, such as the Absenteeism Screening Questionnaire (Truchon et al., 2012), could be investigated. Second, as additional prognostic factors (for the same outcome) emerge with new research, the potential to build a tool with even better predictive accuracy could be explored by incorporating such factors alongside the UL-WIS (e.g. a multi-factor predictive index). Third, new research exploring prognostic factors/tools for (improvements in) health-related work limitations among injured workers is also needed, as this is also an important outcome concerning those who are able to work. Fourth, recognizing that UL-WIS is a prognostic factor for a subsequent transition out of work, another important line of future research should involve developing and testing interventions that can effectively reduce WI. As WI is not yet a familiar concept, this remains a relatively unexplored topic at present. A general working hypothesis may be that a trans-disciplinary and integrative rehabilitation approach (i.e. that concurrently addresses both worker and workplace issues in a coordinated manner) is key towards improving the “fit” between work abilities and job demands, and therefore, also likely to reduce WI.

4) Unpacking the nature of worker-workplace interactions: the need for further theorization and empirical testing

Although the general concept has been proposed in a number of theoretical frameworks (Loisel et al., 2001; Sandqvist and Henriksson, 2004), uncovering the nature of worker-workplace interactions is challenging for a number of reasons. For example, there remains a lack of hypotheses about the specific worker and workplace factors involved in a given context, insights on how to best operationalize the relevant variables, or the ideal design and analytic methods to reveal/verify the nature of such interactions. As such, there is much room for growth in all these areas. In light of this, it should be reiterated that our approach in Objective #3 represents but one way to explore this phenomenon, as other sensible approaches are also possible, for example, testing for statistical interactions involving more specific work characteristics within a regression model. In terms of future research, a
worthwhile avenue may involve some qualitative work to develop more theorizations and testable hypotheses on this topic. Such work will be valuable in terms of providing a foundation for subsequent quantitative work to further elucidate the nature of worker-worker interactions in occupational disability.

5.5 CONCLUDING REMARKS

Work disability represents an important concern and source of burden for injured workers recovering from a WRUED. Towards the long-term goal of reducing this burden, research focused on fundamental topics such as the measurement and determinants of work disability are important. This thesis contributed understanding on both these topics. Specifically, new insights were gained on the suitability of two measurement tools for use, as well as the nature and relevance of worker-workplace interactions in this population. Future research avenues to build on the current work were proposed.
COMPLETE BIBLIOGRAPHY


Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 1982; 143: 29-36.


Lerner D, Rogers WH, Chang H. Scoring the work limitations questionnaire (WLQ) and the WLQ index for estimating work productivity loss. Technical report; 2003b.

Li X, Gignac MA, Anis AH. The indirect costs of arthritis resulting from unemployment, reduced performance, and occupational changes while at work. Med Care, 2006; 44: 304-10.

Lo YT, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. Biometrika, 2001; 88: 767-78.


Complete Bibliography


Widaman KF, Thompson JS. On specifying the null model for incremental fit indices in structural equation modeling. Psychol Methods, 2012; 8: 16-37.


APPENDIX 1: Measures of Work Disability and Productivity: Rheumatoid Arthritis
specific Work Productivity Survey (WPS-RA), Workplace Activity Limitations Scale (WALS), Work Instability Scale for Rheumatoid Arthritis (RA-WIS), Work Limitations Questionnaire (WLQ), and Work Productivity and Activity Impairment Questionnaire (WPAI)

A version of this paper has been published in Arthritis Care & Research 63(S11):S337-S349 (Tang et al., 2011), reproduced with permission
INTRODUCTION

The impact of arthritis on work is an area of increasing research interest and a growing number of outcome measures to quantify such impact have become available in recent years. Recent reviews from an Outcome Measures in Rheumatology (OMERACT) initiative have broadly identified 24 instruments in this area (1;2), though only 11 had been used in arthritis to-date. Previous studies have shown that these measures only moderately correlate with each other (3-10), thus it is important to recognize that available instruments offer distinct perspectives on health-related work impacts. For example, some measures are aimed at examining degree of difficulties with specific workplace activities, while others are designed to quantify the extent of absenteeism (e.g. number of days off work) and/or presenteeism (being at-work but working at reduced productivity, also referred to as “at-work productivity loss” or “at-work disability”). Also available are instruments focused on assessing related concepts such as work “performance”, “efficiency”, “instability”, or degree of “interference” at work. For the purpose of this review, we have adopted a broader approach and consider a diverse range of available measures that offer varying perspectives and approaches to quantifying the impact of health problems on work (Table 1).

Beyond their diverse conceptual foci, existing instruments also differ in terms of their scope of measurement (e.g. impact on employment work vs. non-paid work and/or leisure activities), disease attribution (e.g. disease-specific vs. generic), length (e.g. number of sections and items), and recall period. Some are designed as “modular” instruments that assess work impacts using a series of global rating scales and discrete items (often organized into multiple sections) that are generally not intended to be summative. Others are classic “psychometric” measures consisting of summative items that inform disease impact on different specific aspects of work (contributes to an overall construct). Specific impacts examined at the item-level may include problems meeting the physical demands of work, challenges associated with time management, difficulties maintaining interpersonal relationships at work, cognitive concerns (e.g. worries about continued employability) and/or issues related to symptom control/exacerbation and fatigue. The diversity of available measures is a strength in this growing field and has accommodated the growing interest to apply these tools for a broad range of purposes in arthritis. For example,
measures of role functioning and productivity at work have been used to examine the epidemiology (e.g. population trends, determinants) of work disability/participation in the arthritis population, to evaluate the effectiveness of clinical or workplace interventions (e.g. clinical trials), and also, to estimate the economic costs of health-related work productivity loss at the societal level.

Five specific measures were selected for a detailed review in this article. These were chosen on the basis of two main criteria: 1) the availability of measurement evidence specific to the arthritis population (primarily osteoarthritis [OA] or inflammatory arthritis [IA]), and 2) evidence of previous research application in arthritis populations beyond psychometric testing. By these selection criteria, both modular and psychometric measures were represented. Measures chosen included the Rheumatoid Arthritis-specific Work Productivity Survey (WPS-RA) (11), Workplace Activity Limitations Scale (WALS) (12), Work Instability Scale for Rheumatoid Arthritis (RA-WIS) (13), Work Limitations Questionnaire (WLQ) (14), and Work Productivity and Activity Impairment Questionnaire (WPAI) (15). It should also be recognized that disease-specific variants of 2 of the 5 selected measures are available for ankylosing spondylitis (i.e. The Work Productivity and Activity Impairment Questionnaire for Ankylosing Spondylitis [WPAI:SpA] (16), and the Ankylosing Spondylitis Work Instability Scale [AS-WIS] (17)), however, these measures will not be presented in this review. Also, we have emphasized psychometric evidence primarily in arthritis populations; notwithstanding that evidence in non-arthritis soft tissue musculoskeletal conditions (e.g. low back pain, upper limb disorders) and other non-musculoskeletal disorders are also available for a number of these measures.
### Table A.1. Summary of measures of role functioning and productivity at work, sorted by year of original publication

<table>
<thead>
<tr>
<th>Primary author of original article</th>
<th>Year of publication</th>
<th>Measure</th>
<th>Type of measure: <em>modular</em> (M) or <em>psychometric</em> (Ps)</th>
<th>Number of items/sections **</th>
<th>Scope of concept: <em>absenteeism</em> (A), <em>presenteeism</em> (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osterhaus</td>
<td>1992</td>
<td>Osterhaus Technique (OST) (18)</td>
<td>M</td>
<td>4</td>
<td>A &amp; P</td>
</tr>
<tr>
<td>Reilly</td>
<td>1993</td>
<td>Work Productivity and Activity Impairment &amp; General Health (WPAI)* (15)</td>
<td>M</td>
<td>6</td>
<td>A &amp; P</td>
</tr>
<tr>
<td>van Roijen</td>
<td>1996</td>
<td>Health and Labor Questionnaire (HLQ) (19)</td>
<td>M</td>
<td>14</td>
<td>A &amp; P</td>
</tr>
<tr>
<td>Endicott</td>
<td>1997</td>
<td>Endicott Work Productivity Scale (EWPS) (20)</td>
<td>Ps</td>
<td>25</td>
<td>P</td>
</tr>
<tr>
<td>Kopec</td>
<td>1998</td>
<td>Occupational Role Questionnaire (ORQ) (21)</td>
<td>Ps</td>
<td>8</td>
<td>P</td>
</tr>
<tr>
<td>Brouwer</td>
<td>1999</td>
<td>Quantity and Quality Method (QQ) from the Productivity and Disease Questionnaire (9)</td>
<td>M</td>
<td>2</td>
<td>P</td>
</tr>
<tr>
<td>Amick</td>
<td>2000</td>
<td>Work Role Functioning – 26 (WRF) (22)</td>
<td>Ps</td>
<td>26</td>
<td>P</td>
</tr>
<tr>
<td>Lerner</td>
<td>2001</td>
<td>Work Limitations Questionnaire – 25 (WLQ-25)* (14)</td>
<td>Ps</td>
<td>25</td>
<td>P</td>
</tr>
<tr>
<td>Altshuler</td>
<td>2002</td>
<td>Life Functioning Questionnaire (LFQ) (23)</td>
<td>M</td>
<td>16</td>
<td>A &amp; P</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
<td>Scale Name</td>
<td>Measure</td>
<td>Form</td>
<td>Scale Length</td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------</td>
<td>------</td>
<td>--------------</td>
</tr>
<tr>
<td>Koopman</td>
<td>2002</td>
<td>Stanford Presenteeism Scale – 6 (SPS-6)</td>
<td>Ps</td>
<td>6</td>
<td>P</td>
</tr>
<tr>
<td>Kumar</td>
<td>2003</td>
<td>Health-Related Productivity Questionnaire Diary (HRPQ-D)</td>
<td>M</td>
<td>9</td>
<td>A &amp; P</td>
</tr>
<tr>
<td>Kessler</td>
<td>2003</td>
<td>World Health Organization Health and Work Performance Questionnaire (HPQ)*</td>
<td>M</td>
<td>13</td>
<td>A &amp; P</td>
</tr>
<tr>
<td>Gilworth</td>
<td>2003</td>
<td>Work Instability Scale for Rheumatoid Arthritis (RA-WIS)*</td>
<td>Ps</td>
<td>23</td>
<td>P</td>
</tr>
<tr>
<td>Goetzel</td>
<td>2003</td>
<td>Work Productivity Short Inventory (Wellness Inventory by Pfizer) (WPSI)</td>
<td>M</td>
<td>4</td>
<td>A &amp; P</td>
</tr>
<tr>
<td>Shikiar</td>
<td>2004</td>
<td>Health and Work Questionnaire (HWQ)</td>
<td>M</td>
<td>5</td>
<td>P</td>
</tr>
<tr>
<td>Turpin</td>
<td>2004</td>
<td>Stanford Presenteeism Scale – 13 (SPS-13)</td>
<td>Ps</td>
<td>13</td>
<td>A &amp; P</td>
</tr>
<tr>
<td>Gignac</td>
<td>2004</td>
<td>Work Activity Limitations Scale (WALS)*</td>
<td>Ps</td>
<td>12</td>
<td>P</td>
</tr>
<tr>
<td>Stewart</td>
<td>2004</td>
<td>Work and Health Interview – The American Productivity Audit (WHI)</td>
<td>Ps</td>
<td>7</td>
<td>A &amp; P</td>
</tr>
<tr>
<td>Burton</td>
<td>2004</td>
<td>Work Limitations Questionnaire – 8 (WLQ-8)</td>
<td>Ps</td>
<td>8</td>
<td>P</td>
</tr>
<tr>
<td>Beaton</td>
<td>2005</td>
<td>Work Limitations Questionnaire – 16 (WLQ-16)</td>
<td>Ps</td>
<td>16</td>
<td>P</td>
</tr>
</tbody>
</table>
Appendix 1

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Title</th>
<th>Measure Type</th>
<th>Items</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Munir</td>
<td>2005</td>
<td>Work Limitations Questionnaire (6 items)</td>
<td>PS 6 P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feuerstein</td>
<td>2005</td>
<td>Workstyle Scale - Long version (WSL)</td>
<td>PS 91 P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feuerstein</td>
<td>2005</td>
<td>Workstyle Scale - Short version (WSS)</td>
<td>PS 32 P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>van Roijen</td>
<td>2007</td>
<td>Short Form – Health Labor Questionnaire (SF-HLQ)</td>
<td>M 11 A &amp; P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osterhaus</td>
<td>2009</td>
<td>RA-specific Work Productivity Survey (WPS-RA)*</td>
<td>M 9 A &amp; P</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* identified as a candidate Outcome Measures in Rheumatology (OMERACT) work productivity outcome measure (2)

** for modular measures organized by sections, the number of sections (not items) are indicated; count excludes socio-demographic items/sections unless directly pertain to employment work (e.g. current work status, employment income)

Operational definitions:

- **Modular measure (M):** consisting of a series of global rating scales (e.g. visual analog scales or numeric rating scales) and discrete items that are generally not intended to be summative; contents may be organized into multiple sections within a questionnaire
- **Psychometric measure (Ps):** consisting of summative items that contributes to an overall concept (can be multi-dimensional, i.e. consisting of subscales)
- **Absenteeism (A):** examines work status and/or extent of time/frequency of being off-work (e.g. sick leave)
- **Presenteeism (P):** examines “on-the-job” impact (e.g. productivity loss associated with reduced work efficiency, or degree of workplace activity limitations)
RHEUMATOID ARTHRITIS SPECIFIC WORK PRODUCTIVITY SURVEY (WPS-RA)

Descriptive

Purpose. Measures the impact of rheumatoid arthritis (RA) on the productivity of employment work, household work, and daily activities. Original version of WPS-RA was published in 2009, primarily intended for use in clinical trials (11).

Content. Assesses the number of days of work absence (absenteeism), number of days with reduced work productivity, and degree of interference on work productivity due to RA (presenteeism).

Number of items. 9 questions (Q), organized into 3 sections.

Section I: employment status, type of work (Q1)

Section II: employment work
- days of work missed (Q2)
- days with productivity reduced by at least half (Q3)
- degree of interference on work productivity (Q4)

Section III: household work and activities
- days of non-paid work missed (Q5)
- days with productivity reduced by at least half (Q6)
- days missed family, social, or leisure activities (Q7)
- days with outside help (Q8)
- degree of interference on (non-paid) work productivity (Q9)

Response options/scale. Unique for each individual question. Q1: employed outside of home (Yes/No), type of work (8 options); Q2-3, 5-8: number of days (count data); Q4&9: global rating scale, 0 – 10 (0 = no interference, 10 = complete interference).

Recall period for items. 1 month.

Endorsements. None. (The WPS-RA is 1 of 6 measures identified as a candidate OMERACT work productivity outcome)

Examples of use. The WPS-RA has been applied as an secondary study endpoint to evaluate effects of certolizumab pegol in RA (37) (11), and also as an outcome to evaluate the effects of certolizumab pegol with methotrexate in RA (38).
Appendix 1

Practical Application

**How to obtain.** Request for permission to use the WPS-RA should be made to Global Health Outcomes Research, UCB Pharma. A copy of the WPS-RA questionnaire can be obtained at [http://www.biomedcentral.com/content-supplementary/ar2702-S1.doc](http://www.biomedcentral.com/content-supplementary/ar2702-S1.doc)

**Method of administration.** Currently intended for interviewer-administration, but there are future plans to develop a self-administered version.

**Scoring.** Each of 9 questions is scored individually (i.e. 8 separate outcome scores [Q1 not considered an outcome]), item scores are not intended to be combined.

**Score interpretation.** Values for Q2-3, 5-8 reflect number of days impacted; for Q4&9, higher scores indicate greater interference of RA on work. Cut-points and normative values are not yet established.

**Respondent burden.** Low.

**Administration burden.** Low, assuming interviewer administration.

**Translations/adaptations.** English.

Psychometric Information

**Method of development.** The WPS-RA was developed through a literature review on work productivity associated with RA or other chronic health conditions (e.g. migraine, depression) (11). Patients/workers were not directly involved in scale development, but item selection emphasized patient-centered considerations (e.g. relevance, burden).

**Acceptability.** In a randomized trial of n=220 with RA, Osterhaus et al (11) reported relatively low missing frequencies of 0.5% (Q6), 1.4% (Q8), 1.8% (Q9), and 0% for all other scale items. Q2-3, 5-8 provides count data, may have propensity for floor effect. Readability of WPS-RA items is high.

**Reliability.** Test-retest or inter-rater evidence not yet available, internal consistency testing is not applicable.

**Validity.** Modest evidence to-date. Relationship of the WPS-RA with other work-specific measures has yet to be assessed, but evidence of known-group differences against general health indicators is available. In Osterhaus et al (11), WPS-RA item scores for Q2-4 (employment work) were shown to differ (P<0.05) between known-groups (≥ third quartile vs. ≤ first quartile) based on scores from the Health Assessment Questionnaire Disability Index (HAQ-DI), Short Form-36 Physical Component Summary Score (SF-36 PCS), and the Short
Form-36 Mental Component Summary Score (SF-36 MCS). Item scores for Q5-9 (household work and daily activities) also shown to differ (mostly \( P < 0.05 \)) between known-groups (≥ third quartile vs. ≤ first quartile) based on scores from the HAQ-DI, SF-36 PCS, and SF-36 MCS.

**Ability to detect change.** No known reports to-date on the sensitivity of the WPS-RA to known changes in work role functioning or productivity, but effect sizes have been assessed in persons with RA in a 24-week certolizumab pegol trial (11). For clinical responders based on the American College of Rheumatology (ACR 20) criteria of improvement (i.e. individuals showing a large change), SRM>0.8 (large effect size) was observed for Q4,6,9, SRM=0.5-0.8 (moderate effect size) was shown for Q3,5, and SRM<0.5 (small effect size) was shown for Q2,7,8. For ACR20 non-responders (i.e. individual showing small change, no change, or deterioration), SRM<0.5 was evident for all WPS-RA items (Q2-9). In the same trial, clinical responders based on improvements in the HAQ-DI (MCID=0.22) (i.e. a large change in disability) had an SRM>0.8 for Q4,6, SRM>0.5 for Q3,5, and SRM<0.5 for Q2,7,8. Among HAQ-DI non-responders (i.e. small change, no change, or deterioration), SRM<0.5 was observed for all items (Q2-9). Some care is needed when interpreting SRMs reported for “non-responders” as these are derived from assessing a pool of individuals who have undergone varying degrees of change in the trial.

**Critical appraisal of overall value to rheumatology community**

**Strengths.** Provides broad coverage of impacts on several work domains: questionnaire considers work impacts from the perspectives of both absenteeism and presenteeism (at-work productivity), as well as impacts on both employment work and non-paid activities (e.g. household work, social and leisure activities).

**Caveats, cautions.** Patients/workers were not directly involved in conceptualization or scale development; count data (number of days) is gathered for 6 of the 9 questionnaire items (proper statistical treatment required); evidence to support application beyond RA not yet available; large number of “outcomes” (i.e. 8 separate scores derived, not intended to be summated/combined); Q2-4 pertains specifically to employment work, thus will not be relevant for persons who are temporary unemployed at the time of data collection (i.e. lead to missing values); number of “days with productivity reduced by at least half” (Q3,6) could be challenging to appraise, and may overlook days where there is also considerable impact but productivity is reduced by less than half; overall, there is limited psychometric...
testing against work-specific indicators (to support validity and responsiveness) to-date, but it should be considered that the WPS-RA is a relatively new measure with much promise.

**Clinical usability.** Good potential, although more studies to establish clinical parameters (e.g. minimal clinically important difference, patient acceptable symptomatic state) are needed. The WPS-RA could be potentially useful for providing a more complete view of disease impact (i.e. on both paid and unpaid work) to facilitate decision making on clinical management and issues around job modifications and work/life balance. It is also a highly feasible measure given low respondent and administrative burden.

**Research usability.** Promising. There is emerging evidence of its application and psychometric performance from clinical trials in RA. Low administrative and respondent burden.
WORKPLACE ACTIVITY LIMITATIONS SCALE (WALS)

Descriptive

Purpose. Measures limitations experienced while performing workplace activities, including difficulties associated with upper-limb functioning, lower-limb functioning, concentration at work, and the pace and scheduling of work. The WALS is intended for arthritis populations. Original publication in 2004 (12).

Content. Assesses difficulties with mobility, prolonged sitting and standing, lifting, working with hands, crouching, bending or kneeling, reaching, scheduling, work hours, pace of work, concentration, and meeting current job demands. Respondents are asked to respond to questions assuming that assistance from others or gadgets/equipments are not available.

Number of items. 11-item and 12-item versions of the WALS are available. An item asking about “difficulties concentrating on work” is not included in the 11-item version. This item was added to the 12-item version based on patient feedback.

Response options/scale. 4-point Likert scaling: from no difficulty (score=0) to not able to do (score=3). Not applicable to my job and difficulty unrelated to arthritis response options are also available (both scored 0).

Recall period for items. “In general” or “typically”.

Endorsements. None. (The WALS is one of 6 measures identified as a candidate OMERACT work productivity outcome)

Examples of use. Has been used in samples with IA (e.g. RA, psoriatic arthritis), OA, and lupus. Examined as a factor associated with arthritis-related work changes, work transitions and job accommodations (12,39-41), behavioral coping efforts (42), arthritis-work spillover (43), chronic job stress and strain (44), and disclosure of chronic disease in the workplace (45).

Practical application

How to obtain. Information on the WALS can be obtained free of charge, from Monique Gignac, PhD, Toronto Western Research Institute at the University Health Network, Toronto, ON, Canada. Email: gignac@uhnres.utoronto.ca

Method of administration. Self or interviewer administration

Scoring. Can be expressed as mean of all scale items (range: 0-3) or as a summed total score
Appendix 1

(summed score range: 0-33 [for 11-item version] or 0-36 [for 12-item version]); mean values can be imputed for up to two missing items.

**Score interpretation.** Higher scores indicate greater workplace activity limitations.

Preliminary data on cutoff values has been examined (41).

**Respondent burden.** Low.

**Administration burden.** Minimal.

**Translations/adaptations.** English.

**Psychometric information**

**Method of development.** Items were based on a review of the literature and were modeled after the Health Assessment Questionnaire and modified to be specific to workplace tasks and activities. Patients/workers were not directly involved in the development of the scale.

**Acceptability.** In a sample of 250 with either RA or OA, Beaton et al. (3) reported WALS summed scores to be available in \( n = 234 \) (6% completed <10 of 11 scale items); 0.0% at floor score (WALS=0); 3.4% at ceiling score (WALS=33). Dhanhani (40) reported <1% missing values for employed and not employed groups with lupus. Readability of WALS items is high.

**Reliability.** For the 11-item WALS, Gignac et al (12;42;43) reported a Cronbach’s alpha range from 0.78-0.81 (\( n = 349-491 \) with OA or IA) over 4 time-points, each 18 months apart. In a sample of 250 with either RA or OA, Beaton et al. (3) reported a Cronbach’s alpha of 0.87. For the 12-item version, a Cronbach’s alpha of 0.81 has been reported in a sample of \( n = 292 \) with either OA or IA (Gignac et al. 2007). Dhanhani (40) reported a Cronbach’s alpha of 0.86 for an employed sample with lupus and 0.80 for those not working.

**Validity.** In a sample of 250 with either RA or OA, Beaton et al. (3) reported a correlation range of \( r = 0.43-0.66 \) against a series of work-oriented constructs (self-reported global items); the WALS also showed moderate-to-high correlations against other work-specific measures: including the SPS-6 (\( r = 0.66 \)), EWPS (\( r = 0.55 \)), RA-WIS (\( r = 0.77 \)), and WLQ Index (\( r = 0.61 \)). Moreover, the WALS also showed moderate correlations against the Health Assessment Questionnaire (HAQ) Disability Index (\( r = 0.66 \)), self-rated arthritis severity (\( r = 0.62 \)), and self-rated pain intensity (\( r = 0.67 \)), which might be considered somewhat higher than expected, given these were not work-specific indicators. None of the comparators tested in this study might be considered a “gold standard” reference indicator, but overall, there is
moderate support for its construct validity.

**Ability to detect change.** In a sample of 250 with either RA or OA, Beaton et al. (3), moderate responsiveness of the 11-item WALS to 1-year improvements (SRM=−0.79) and 1-year deteriorations in work ability (SRM=0.50) were found (ranked 1st out of 5 at-work measures compared in the study), but smaller effect sizes for 1-year improvements (SRM=−0.37) and 1-year deterioration in work productivity (SRM=0.18) were observed (ranked 2nd out of 5 at-work measures compared). In this analysis, single item global indices of change were used to provide “reference” indicators of change, and individuals showing varying magnitudes of change were pooled in the analysis (i.e. no stratification of individuals based on magnitude of change). Additional studies to examine the responsiveness of the WALS to more defined changes (i.e. “smaller” vs. “larger”) in workplace activity limitations could be informative.

**Critical appraisal of overall value to rheumatology community**

**Strengths.** This is a feasible measure that captures many of the key perspectives of workplace activity limitations, and has shown very good psychometric performance overall. Included in the scale is an assessment of potential difficulties related to mobility (around workplace, to and from work), which is often omitted in other work outcomes. It also offers a difficulty unrelated to arthritis response option (scored as 0) which is also quite unique. To date, available psychometric evidence has been gathered from samples with OA, IA, and lupus, suggesting potential utility across different arthritis conditions.

**Caveats, cautions.** The WALS has yet to be examined to-date as a “true” study endpoint or applied in clinical trials. There is also limited current information on interpretation of scores or cut-points. Users should be aware of the two different versions of the measure, which have shown comparable levels of internal consistency (comparability of other psychometric properties not yet known). Recall time-frame is not a specified time, but is “in general” or “typically”.

**Clinical usability.** Good potential. The WALS is a feasible tool that can inform impact of disease at the “activity” level (specific to the workplace). When used in concert with other health indicators, it has the potential to help guide clinical decisions related to management strategies (e.g. the need for additional therapeutic or workplace interventions), and vocational recommendations and/or decisions (e.g. sick leave); establishing clinical
parameters (e.g. minimal clinically important difference, patient acceptable symptomatic state) in future studies will be useful to this end.

**Research usability.** Good; low respondent burden; minimal administrative burden.
WORK INSTABILITY SCALE FOR RHEUMATOID ARTHRITIS
(RA-WIS)

Descriptive

Purpose. Measures the extent of work instability (WI), which is defined as “a state in which
the consequences of a mismatch between an individual’s functional abilities and the demands
of his or her job can threaten continuing employment if not resolved” (13), originally
developed specifically for RA.

Content. This is a psychometric (summative) measure with items covering a broad range of
specific work-related issues (e.g. symptom control, time management, task difficulties at
work, cognitive distresses due to concerns about future employability) that may signify a
functional ability/job demands mismatch.

Number of items. 23.

Response options/scale. Dichotomous response options: yes/no.

Recall period for items. “At the moment”.

Endorsements. None. (The RA-WIS is 1 of 6 measures identified as a candidate
OMERACT work productivity outcome)

Examples of use. RA-WIS was assessed as a secondary outcome in a clinical trial of anti-
TNF adalimumab on RA (46), and as an outcome for comparing occupational therapy
versus usual care in RA (47). Macedo et al (48) examined the relationships between the DAS
28 (Disease activity score using 28 joint counts), HAQ scores, and RA-WIS (as study
outcome). The RA-WIS also showed ability to predict arthritis-related work transitions (e.g.
disability leave of absence, reducing work hours, or job changes) among workers with RA or
OA within 1 year (49).

Practical application

How to obtain. The RA-WIS is copyrighted to the Psychometric Laboratory for Health
Sciences, University of Leeds;
www.leeds.ac.uk/medicine/rehabmed/psychometric/Scales3.htm

Method of administration. Patient administration (self-report).

Scoring. The RA-WIS is scored by summing responses from all 23 scale items (scale range:
0-23), no instructions for missing value available; conversion into interval-level scaling
Appendix 1

(derived from Rasch analysis) is available for OA (50).

**Score interpretation.** Cut-points have been established to differentiate levels of WI: *low* <10; *moderate* = 10-17; and *high* >17 (13). This 3-level categorization has also demonstrated predictive validity for arthritis-related work transitions within 1 year (49).

**Respondent burden.** Low. Easy to read.

**Administration burden.** Minimal.

**Translations/adaptations.** Available in 18 languages; adaptation of the RA-WIS requires explicit permission from Galen Research in Manchester; cross-cultural validity of the RA-WIS has been shown for UK, Dutch, German, French (51).

**Psychometric information**

**Method of development.** Details of original development of the RA-WIS are reported in Gilworth *et al* (13). Qualitative interviews were conducted in individuals with RA to identify key themes (job flexibility, good working relationships, and symptom control) from which items were generated; 76 initial statements were identified as potential items, which were reduced to 36 items based on ability to discriminate against 5 levels of work instability as assessed by vocational experts, and then finally reduced to 23 items based on assessment of fit to the Rasch model (1-parameter item response theory approach).

**Acceptability.** In a sample of *n*=250 with either RA or OA, Beaton *et al* (3) reported that RA-WIS scores were available in *n*=223 (scores were not calculated if >10% of items were missed), of whom 9.4% had the floor score (RA-WIS=0), while 0.4% was at the ceiling score (RA-WIS=23). Overall, readability of RA-WIS items is high.

**Reliability.** Gilworth *et al* (13) reported a test-retest correlation of *r*=0.89 (*n*=51); in a sample with *n*=250 RA or OA (3), Kuder Richardson Formula 20 (KR-20) for the RA-WIS was 0.91 and item-total correlation ranged from 0.34-0.71. KR-20 was 0.93 for *n*=130 with OA within this sample (50).

**Validity.** In a sample of *n*=250 with RA or OA (3), the level of correlation (Spearman *r*) against work-oriented constructs (self-reported global items) was *r*=0.54-0.74, which was considered best among 5 at-work measures compared in this study. The RA-WIS also showed moderate-to-high correlations against other work-specific measures including the WALS (*r*=0.77), SPS-6 (*r*=0.69), EWPS (*r*=0.64), and WLQ Index (*r*=0.61), and correlated moderately with HAQ (*r*=0.66), self-rated arthritis severity (*r*=0.62) and pain intensity
Appendix 1

Among workers with OA ($n=130$) (50), the range of correlation (Spearman $\rho$) against work-oriented constructs (self-reported global items) was $\rho=0.55-0.77$, and moderate-to-high correlations with the HAQ ($\rho=0.70$), arthritis severity ($\rho=0.75$), pain intensity ($\rho=0.79$) were also found. Of note is that the level of correlations between the RA-WIS and health (non work-specific) measures from these studies appeared to be somewhat higher than might be expected (i.e. comparable to the level of correlation against work-specific measures). Tang et al (50) demonstrated that the scoring structure of the RA-WIS shows adequate fit to the expectations of the Rasch model with only minor modifications, and thus its summed score may be considered compatible for transformation into interval-level scaling. Proper fit to the Rasch model requires the pattern of item response to satisfy a number of criteria: 1) approximation to the Guttman structure, 2) demonstrate the lack of differential item functioning, as well as 3) provide evidence of unidimensionality and local independence of items. These criteria were met when RA-WIS was tested in OA.

**Ability to detect change.** In a sample of 250 with either RA or OA (3), high responsiveness of the RA-WIS to 1-year improvements (SRM=-0.64) and 1-year deteriorations in work ability (SRM=0.88) were found (ranked 2nd out of 5 at-work measures compared in the study), but only small-to-negligible effect sizes for 1-year improvements (SRM=-0.29) and deteriorations in work productivity (SRM=0.00) were found (ranked 5th out of 5 at-work measures compared in the study). In a sample of $n=130$ with OA, large effect sizes were observed in the RA-WIS for 1-year deteriorations in intrusiveness of arthritis on work, (SRM=1.05) and also for 1-year improvements in intrusiveness of arthritis on work (SRM=-0.78) (50). In both these studies, single item global ratings of change were used to provide “reference” indicators, and it should be considered that individuals demonstrating various magnitude of change were pooled into the same analysis. Further studies to examine whether the RA-WIS is similarly responsive to more defined changes (“smaller” vs. “larger”) in work instability could be informative.

**Critical appraisal of overall value to rheumatology community**

**Strengths.** “Work instability” (functional ability/job demands mismatch) is a unique concept among available work outcomes; evidence of ability to predict for future work transition outcomes in RA and OA suggest that it could have a potential role in risk prognostication, in addition to potential applications as a study endpoint. Patient preference for the RA-WIS
over 3 other at-work disability measures was shown a study that sampled injured workers with upper-limb musculoskeletal disorders (4), likely due to the relative simplicity of item statements and/or the dichotomous response options. Overall psychometric evidence is very good in both RA and OA. Alternative versions of the scale have been developed for other rheumatic conditions like ankylosing spondylitis (AS-WIS) (17).

Caveats, cautions. Some concerns about the limited response options (i.e. lack of a middle-ground option between “yes” and “no”) has been suggested by workers with RA or OA (unpublished work, data from the $n=250$ cohort used in Beaton et al (2010)). Some items appeared to exhibit some redundancy (50).

Clinical usability. The RA-WIS is feasible and is a measure that has been well received by patients with arthritis or other musculoskeletal conditions; potential for clinical use for risk prognostication of adverse future work outcomes; support for the predictive validity of the proposed RA-WIS cut-points has been shown which could be applied for risk stratification: (low vs. moderate vs. high WI); further psychometric evidence and establishing clinical parameters (e.g. minimal clinically important difference, patient acceptable symptomatic state) will be helpful for clinical interpretation of scores.

Research usability. Excellent; a versatile tool (intended concept could be of interest as either a prognostic factor or study endpoint); minimal administrative and respondent burden, unless score calibrations are used (i.e. conversion of summed scores to interval-level scores).
WORK LIMITATIONS QUESTIONNAIRE (WLQ)

Descriptive

**Purpose.** Measures the on-the-job impact of chronic health conditions and treatment with a focus on assessing limitations while performing specific job demands. Original 25-item version (WLQ-25) was published in 2001 (14), and several shortened versions have been tested or applied in studies in workers with various musculoskeletal disorders: WLQ-16 (32), WLQ-8 (30;31), and a 6-item version (33). Another variant is the Work Role Functioning scale (WRF-26)(22), which has similar purpose and content, but is reversed in conceptual orientation (i.e. assesses level of role functioning, not limitations). The current review will focus mainly on the WLQ-25.

**Content.** The WLQ-25 can be organized into 4 domains: time management (TM), addresses difficulty with handling a job’s time and scheduling demands; physical demands (PD), examines ability to perform job tasks that involve bodily strength, movement, endurance, coordination, and flexibility; mental-interpersonal demands (MI), addresses cognitively-demanding tasks and on-the-job social interactions; and output demands (OD), concerns reduced work productivity (14).

**Number of items.** 25 total scale items, divided into 4 subscales: TM (5 items), PD (6 items), MI (9 items), and OD (5 items).

**Response options/scale.** Three of the 4 subscales (TM, MI, OD) examines proportion of time with difficulty: “none of the time (0%)”, score=0; “a slight bit of the time”, score=1; “some of the time (50%)”, score=2; “most of the time”, score=3; “all of the time (100%)”, score=4; plus a does not apply to my job option (treated as missing, no score). The PD subscale has reverse instructions and examines proportion of time without difficulty (same response options provided).

**Recall period for items.** 2 weeks for the WLQ-25. Varies with other versions.

**Endorsements.** None. (The WLQ-25 is 1 of 6 measures identified as a candidate OMERACT work productivity outcome)

**Examples of use.** Rohekær *et al* (52) applied the WLQ-25 as an outcome to assess work disability in seronegative spondyloarthritis. Allaire *et al* (53) applied the WLQ-25 to examine the impact of RA on work disability among 5419 older workers (age range: 55-64) with RA. Lerner *et al* (54) recently developed a method to impute work productivity impact using other
health variables, following an examination of the relationship between the WLQ-25 and an array of pain, functioning, and general health measures. *Tang et al* (4) provided a head-to-head comparison of the psychometric performance of 4 at-work disability measures (includes WLQ-16) among injured workers with upper-limb disorders. *Zhang et al* (55) examined the comparability of methods to estimate productivity loss based on 4 different instruments (includes conversions based on the WLQ Index). Associations between medical conditions (31), arthritis (56), and health risks (57) with work limitations had been examined using the WLQ-8.

**Practical application**

**How to obtain.** The WLQ is copyrighted: Work Limitations Questionnaire © 1998. The Health Institute; Debra Lerner, Benjamin Amick III, Glaxo Wellcome Inc. The WLQ is provided free of charge for non-commercial applications.

**Method of administration.** Patient administration (self-report).

**Scoring.** Multiple approaches have been recommended to score the WLQ-25. *Subscales:* scored by multiplying the mean of subscale items by 25 (range: 0-100, 100=most limitations; note: response orientation needs to be reversed for the PD subscale), missing response for up to 50% of subscale items is allowable for subscale scoring. *Scale scoring:* scored by multiplying the mean of all scale items by 25 (range: 0-100, 100=most limitations; note: response orientation needs to be reversed for the PD subscale); missing response for up to 50% of subscale items is also allowable for scale scoring (Benjamin C. Amick, personal communication). *Weighted index scoring (WLQ Index):* calculated from subscale scores using a weighted formula based on an analysis of the relationship between WLQ scores and actual employee productivity loss relative to healthy employees (58). Scores from all 4 subscales are needed to calculate the WLQ index, computer scoring necessary. Formulas for calculating the WLQ Index and conversion to productivity loss estimates versus healthy controls (range: 0-25%) are published in a technical report available from the developers (58).

**Score interpretation.** Normative WLQ subscale means (standard error) have been reported in a small sample of healthy workers recruited from Massachusetts, United States \(n=37\): PD=4.5 (1.4); TM=7.2 (3.1); MI=10.6 (2.7); OD=7.2 (2.8). A 10% increase WLQ subscale score has been proposed to equate a productivity decline of 4-5% (59).

**Respondent burden.** Moderate. *Lerner et al* (60) reported an administration time of
approximately 30 minutes for workers with OA, and approximately 15 minutes for healthy controls. Some workers have expressed concern over the flipping of instructions for different sections of the questionnaire (i.e. the TM, MI, OD subscales ask about amount of time with difficulty, while the PD subscale asks about amount of time without difficulty).

**Administration burden.** Moderate. Computer scoring required for handling/imputation of missing values, and for calculating the WLQ Index.

**Translations/adaptations.** Over 30 official language translations.

### Psychometric information

**Method of development.** Content and format of the WLQ-25 were developed from focus groups (workers with chronic conditions), cognitive interviews, and an alternate form comparison (stem/response) (14); 70 job demand-level limitation items and 7 dimensions were originally generated, which were reduced to 25 items through cognitive interviews (14).

**Acceptability.** *Lerner et al* (60) reported <1% missing data in OA, floor effect = 20.4-25.8% (subscales), and ceiling effect = 0.9-2.2% (subscales). In workers with RA, *Walker et al* (61) reported a missing proportion of 3.1 – 21.8% for individual scale items and 3.1 – 5.6% for subscale scores, and as a result, the WLQ Index score was unavailable in 10.1% of the sample due to missing data. In a study that recruited *n=250* with RA or OA, *Beaton et al* (3) reported 5.6% (WLQ Index) and 19.9-35.5% (subscales) of the sample had the floor score, while 0% (WLQ Index) and 1.2-3.3% (subscales) of the sample had the ceiling score.

**Reliability.** Among workers with OA, *Lerner et al* (60) reported the following item-to-total correlation coefficients and Cronbach’s alpha for the 4 WLQ-25 subscales: PD: 0.72-0.82, alpha=0.93; TM: 0.79-0.92, alpha=0.95; MI: 0.81-0.92, alpha=0.97; OD: 0.82-0.89, alpha=0.96. In RA, *Walker et al* (61) reported a Cronbach’s alpha range of 0.83-0.88. In a pooled sample of workers with RA or OA, *Beaton et al* (3) reported the following item-total correlations and Cronbach’s alphas for the 4 WLQ-25 subscales: PD: 0.38-0.63, alpha=0.77; TM: 0.60-0.76, alpha=0.86, MI: 0.50-0.90, alpha=0.94, OD: 0.61-0.81, alpha=0.88.

**Validity.** In OA (60), each of the 4 WLQ-25 subscale scores showed a logical and statistically significant association (*P*<0.001 in ANOVA *F*-test) with self-reported arthritis severity (4 levels: poor to very good); specific WLQ-25 subscales also showed linear associations (subscales demonstrating *P*<.05 in ANOVA *F* test in parentheses below) with level of arthritis pain (PD, TM, MI, OD), joint stiffness (PD, TM, MI, OD), functional
Appendix 1

limitations due to arthritis (PD, TM), SF-12 PCS (PD), self-reported work productivity (OD) and work absences (OD). In RA (61), the WLQ Index showed low-to-moderate correlations with the SF-36 MCS (r=-0.60), SF-36 PCS (r=-0.49), fatigue VAS (r=0.50), pain VAS (r=0.46); HAQ (r=0.56), HAQ-II (r=0.54), depression score (r=0.46) and anxiety score (r=0.41) from the Arthritis Impact Measurement Scale, days with limited activities (r=0.38), and days unable to work (r=0.29). Wolfe et al (62) reported similar levels of correlation between the WLQ index and HAQ (r=0.57), HAQ-II (r=0.55), modified HAQ (r=0.55), SF-36 PCS (r=0.50), VAS pain (r=0.47), when tested in a population of workers with RA. In a sample of n=250 workers with either RA or OA (3), the WLQ index showed moderate correlations (r=0.49 – 0.67) against a series of work-oriented constructs (self-reported global items), and also against the HAQ (r=0.49), arthritis severity (r=0.42), and self-rated pain intensity (r=0.48). In this study, the WLQ index also showed moderate-to-high correlations against other work-specific measures: including the WALS (r=0.61), SPS-6 (r=0.63), EWPS Scale (r=0.61), and RA-WIS (r=0.67). Overall, relationships between the WLQ-25 and various health and work-related indicators were by and large inline with expectations, providing strong support for its construct validity.

Ability to detect change. In a sample of 250 workers with either RA or OA (3), levels of responsiveness of the WLQ Index to 1-year improvements (SRM=-0.28) and 1-year deteriorations in work ability (SRM=0.20) were modest (ranked 5th out of 5 at-work measures compared in the study). Varying levels of responsiveness to 1-year improvements (SRM=-0.64) compared to 1-year deteriorations in work productivity (SRM=0.08) were also reported (tied for 3rd out of 5 at-work measures compared). In this study, single item global ratings of change were applied as “reference” indicators, and individuals experience varying level of changes were pooled into the same analysis (i.e. no stratification of individuals based on magnitude of change). Further evaluations are needed to examine level of responsiveness of the WLQ-25 against more defined magnitudes of change (“smaller” vs. “larger”) in work limitations.

Critical appraisal of overall value to rheumatology community

Strengths. The WLQ has a solid foundation of development, which has led to well-defined domains that have been consistently applied in various studies to assess the impact of health conditions on specific aspects of work. The breadth of potential work limitations examined
is high, and should have strong relevance for many job types and health conditions. To date, this is one of the most widely used measure in the field, with studies in RA, OA, and several other musculoskeletal conditions; cumulative evidence of its construct validity is excellent. The WLQ Index can be converted to provide an estimate of productivity loss (in %) – a unique property among available measures in the field, which is potentially useful to bridge measurement needs for dual clinical/economic costing purposes.

**Caveats, cautions.** Several variations of the measure exist in the literature (WLQ-25, WLQ-16, WLQ-8, WRF-26) and approaches to score are unique among tools and across existing studies. The different versions also vary in terms of recall period, specific wording of items, and whether specific sections (i.e. PD subscale) have reverse orientation within the full questionnaire. Care must be taken when making comparisons of results across versions. For the WLQ-25, reverse instructions for the PD domain may confuse respondents and could be a source of error. There is also a generous allowance of missing data (up to 50% of missing items may be imputed).

**Clinical usability.** Moderate administrative and respondent burden should be considered. More studies to establish clinically meaningful cut-points (e.g. minimal clinical important difference, patient acceptable symptomatic state) could help improve clinical usability.

**Research usability.** Excellent. Very good-to-excellent psychometric evidence in arthritis as well as in other clinical populations. High potential for use for economic costing purposes given the time orientation of response options, and purported relationship between WLQ index and level of work productivity reported by *Lerner et al* (59).
WORK PRODUCTIVITY AND ACTIVITY IMPAIRMENT QUESTIONNAIRE (WPAI)

Descriptive

**Purpose.** Measures the effect of health and symptom severity on work productivity and non-work activities. Two versions are available: general health (WPAI:GH) or specific health problem (WPAI:SHP), the latter is designed such that it can be modified for any health problem by specifying the disease/condition of interest in the questions (i.e. to derive disease-specific versions of the scale). The original WPAI:GH was published in 1993 (15); approach to scoring the questionnaire has changed since the original publication.

**Content.** Examines the extent of absenteeism, presenteeism and impairment in daily activities attributable to general health (WPAI:GH) or a specific health problem (WPAI:SHP).

**Number of items.** 6 questions (Q), each with unique response options: current employment status (Q1), number of hours missed due to health problem (Q2), number of hours missed due to other reasons (Q3), hours actually worked (Q4), degree to which health affected productivity while working (Q5), degree to which health affected regular (non-work) activities (Q6). Items not intended to be summative.

**Response options/scale.** Q1: Yes/No, Q2-4: number of hours (count data), Q5: global rating scale: 0 – 10 (health problems had no effect on my work – health problems completely prevented me from working), Q6: global rating scale: 0 – 10 (health problems had no effect on my daily activities – health problems completely prevented me from doing my daily activities).

**Recall period for items.** 7 days.

**Endorsements.** None. (The WPAI:GH is 1 of 6 measures identified as a candidate OMERACT work productivity outcome)

**Examples of use.** *Stockl et al* (63) applied the WPAI:GH as a secondary outcome to evaluate the effect of a RA disease treatment management program. In a study by *Zhang et al* (55), productivity costs associated with arthritis estimated with the WPAI:GH was compared against estimates made from 3 other measures (HLQ, HPQ, WLQ-25). *Haibel et al.* (64) applied the WPAI:SpA as an outcome measure to examine the efficacy of infliximab therapy in NSAID-refractory patients with ankylosing spondylitis. Associations among demographic, health-related, and treatment factors with the level of work productivity (WPAI:SpA) have
been recently examined in patients with ankylosing spondylitis (65).

**Practical application**

**How to obtain.** The WPAI is available from [www.reillyassociates.net/Index.html](http://www.reillyassociates.net/Index.html); permission and fees are not required to use the WPAI.

**Method of administration.** Self-administered or interviewer-administered.

**Scoring.** Detailed information is provided in the website; four outcome (OC) scores can be derived from the WPAI: OC1 – Percent work time missed due to health: $Q2 / (Q2 + Q4)$ (percentage of absenteeism); OC2 – Percent impairment while working due to health: $Q5 / 10$ (percentage of presenteeism); OC3 – Percent overall work impairment due to health: $Q2 / (Q2 + Q4) + [(1 - Q2 / (Q2 + Q4)) \times (Q5 / 10)]$; OC4 – Percent activity impairment due to health: $Q6 / 10$.

**Score interpretation.** For all 4 outcomes, greater scores (range: 0-100%) indicate greater impact of health, clinically important cut-points not yet established.

**Respondent burden.** Minimal.

**Administration burden:** Low, only basic calculations needed.

**Translations/adaptations.** Has been translated in more than 80 languages ([www.reillyassociates.net/WPAI_Translations.html](http://www.reillyassociates.net/WPAI_Translations.html)); the WPAI-SHP can theoretically be adapted to any specific disease or health problem; psychometric evidence is available in a wide range of diseases. Recently, a version of WPAI for ankylosing spondylitis has been developed (WPAI:SpA)(16).

**Psychometric information**

**Method of development.** WPAI items were generated from three sources (15): 1) review of work productivity literature; 2) comments from patients with allergic rhinitis on an interviewer-administered version of the WPAI items from a series of clinical studies; and 3) cognitive debriefing of subjects following interviewer-administration and self-administration of WPAI items to determine final wording.

**Acceptability.** In the original study that included a sample of workers with RA or other musculoskeletal disorders (15), up to 21% of sample had missing data on WPAI questions when the measure was self-administered, but minimal missing data when the questions were interviewer-administered. The WPAI:SpA had <10% missing when self-administered (16).
Appendix 1

Reliability. In the original study by Reilly et al (15), only a modest range of Pearson’s correlation coefficient (0.71-0.75) was found in a test-retest comparison of WPAI:GH fielded in workers with non-specific health problems within the same day (at least 4 hours later). Level of agreement (e.g. intraclass correlation coefficients) between test-retest scores was not reported in this study.

Validity. Some support for the construct validity of the WPAI:GH in RA was provided in Zhang et al (66), although the scale appeared to correlate more strongly to health status indictors than work-specific comparators where the opposite might be expected. OC1 showed a correlation of $r=0.56$ with the number of absent workdays in the past three months (question adapted from the PROductivity and DISease Questionnaire (PRODISQ)); OC2 showed a correlation of $r=0.39$ with the number of hours lost due to presenteeism (item from the HLQ); and OC4 showed a correlation of $r=0.39$ with the number of hours getting help on unpaid work activities (item from the HLQ). In this study, 3 of 4 WPAI-GH outcomes (OC2,3&4) also showed moderate-to-high correlations ($r=0.67-0.77$) against a series of health status outcomes (function, pain, patient global estimate on health impact, fatigue, patient global assessment of disease activity). In an adalimumab versus placebo clinical trial in AS (16), WPAI:SpA outcomes were shown to be able to discriminate between “higher” and “lower” scores (split on the basis of median score in sample) in the Bath Ankylosing Spondylitis Disability Activity Index (BASDAI), the Ankylosing Spondylitis Quality of Life (ASQOL), SF-36 PCS, SF-36 MCS, and Health Utility Index-3 at $P<0.05$. Evidence of known-group differences are consistent with apriori expectations in the trial, although it should be recognized that the comparators applied were not work-specific indicators; additional studies to specifically examine the responsiveness of the WPAI to “true” changes in work productivity would be informative.

Ability to detect change. There are no known reports on the responsiveness of the WPAI against comparable indicators of work productivity to-date, but effect sizes for persons with AS in a 24-week adalimumab versus placebo trial have been reported by Reilly et al (16). In this study, clinical responders based on improvements in the BASDAI ($>1.96$ decrease in score, i.e. a large change in disease activity), SRMs for WPAI outcomes were: OC1: -0.25; OC2: -0.86; OC3: -0.89; OC4: -1.29. Among non-responders (i.e. small change, no changes, or deterioration), SRMs were OC1: -0.14; OC2: -0.52; OC3: -0.54; OC4: -0.39. Among clinical responders based on changes in the ASQOL ($>1.8$ decrease in score, i.e. a large
change in quality of life), SRMs were $\text{OC1}:-0.31; \text{OC2}:-0.89; \text{OC3}:-0.94; \text{OC4}:-1.18$. Among ASQOL clinical non-responders (i.e. small change, no change, or deterioration), SRMs were $\text{OC1}:-0.11; \text{OC2}:-0.46; \text{OC3}:-0.38; \text{OC4}:-0.40$. It is important to exercise care when interpreting SRMs reported for “non-responders” as this is derived from a pool of individuals who have undergone varying degrees of change over the course of the trial.

**Critical appraisal of overall value to rheumatology community**

**Strengths.** The WPAI is designed to have generalizability to a broad range of occupations/diseases, and evidence of reliability and validity is available in many musculoskeletal and non-musculoskeletal conditions. Item content is highly consistent across different versions of the measure (WPAI:GH vs. WPAI:SHP), which should facilitate comparison of outcome scores in different studies (only disease attribution of items differ between versions); intuitive method to score; compatibility with economic costing (orientation of response is based on amount of time affected), low respondent burden.

**Caveats, cautions.** Four separate outcome scores are derived from the questionnaire (not intended to be summated/combined). There have been important changes in the approach to score the WPAI questionnaire since its original development.

**Clinical usability.** Good potential; establishing clinical parameters (e.g. minimal clinically important difference, patient acceptable symptomatic state) will be helpful for clinical interpretability; administrative and respondent burdens are low.

**Research usability.** Good. Applicability and good psychometric performance of the WPAI has been shown in several clinical trials with patients with ankylosing spondylitis.
DISCUSSION

The current review has revealed a moderate level of evidence to-date to support the psychometric properties of 5 selected measures of role functioning and productivity at work in arthritis/musculoskeletal populations. It is important to recognize that the WPAI and WLQ-25 were developed as generic (not disease-specific) instruments, and there exists additional evidence in the literature to support their psychometric performance in other populations that have not been reviewed in detail in this article. On the other hand, evidence of psychometric performance for the WALS, RA-WIS, and in particular WPS-RA, have only emerged in the past few years, since these are relatively new measures. We believe there is room for continued growth in this area of research. Specific measurement attributes requiring further research include examinations of test-retest reliability, validation against work-specific constructs and/or productivity data (where relevant), and assessments of responsiveness to more defined magnitudes of change (e.g. sensitivity to “smaller” vs. “larger” known changes). Also, data on clinically-relevant parameters (e.g. MCID) is scarce to-date and will be important to establish to help evaluate treatment efficacy in research trials and longitudinal observational studies. Issues on the interpretability of scores derived from measures of role functioning and productivity at work are also of emerging interest. An important concept to recognize is that the extent of disease impact on work is ultimately a function of both the person and his or her work context (i.e. environmental factors) and the manner in which they interact (2;67). At the individual level, a change in score could reflect a change in the person’s capacity to work, and/or a change in the demands of the job, for example, in the case where a work transition (e.g. job modifications, reduced work hours) has taken place to allow a person with arthritis to function better at work. To provide a more complete understanding of the bases of change in over time, users may consider fielding additional instruments that can offer insights into the work context (e.g. job type, work status, contractual hours, availability of workplace support) to supplement outcome measures designed to quantify the level of work role functioning and productivity. Overall, the current diversity of available measures in this field is impressive. While the availability of a wide range of instruments can provide users with many options, some care is important when selecting an outcome to meet the needs of a particular research study or clinical purpose. The specific work-related concept or measurement perspective being sought, the availability of supporting psychometric evidence, and pragmatic considerations (e.g.
applicability, feasibility) should be concurrently considered. In addition to the summary of evidence provided in the current article, users may also consider additional findings and insights from a number of recent studies (3;4) that have examined the head-to-head psychometric performance of multiple work measures in arthritis/musculoskeletal populations to help inform the selection of outcomes in future research or clinical applications.
REFERENCES


(36) van Roijen L. Short Form - Health and Labour Questionnaire. Institute for Medical Technology Assessment, Erasmus MC, University Medical Centre Rotterdam 2007; HUhttp://www.imta.nl/publications/07103.pdfUH.


(40) Dhanhani A. The Workplace Challenges of Patients with Lupus. Graduate Department of the Institute of Medical Science, University of Toronto 2010; URL: HUhttp://hdl.handle.net/1807/25407UH.

Appendix 1


(58) Lerner D, Rogers WH, Chang H. Scoring the work limitations questionnaire (WLQ) and the WLQ index for estimating work productivity loss. Technical report 2003.


APPENDIX 2: Documentation of research ethics approval from the University of Toronto
Appendix 2

UNIVERSITY OF
TORONTO

OFFICE OF THE VICE PRESIDENT, RESEARCH

PROTOCOL REFERENCE #26385
April 26, 2011

Dr. Robin Richards
Sunnybrook Health Sciences Centre
2075 Bayview Ave., D5-74
Toronto, ON M4N 3M5

Dr. Farshid Tabloie & Mr. Kenneth Tang
Sunnybrook Health Sciences Centre
2075 Bayview Ave., D5-74
Toronto, ON M4N 3M5

Dear Dr. Richards, Dr. Tabloie & Mr. Tang:

Re: Administrative Approval of your research protocol entitled, "Managing the "Tail of the Curve": The Course, Predictive Factors and Work-Related Outcomes of Injured Workers One Year After Attending the WSIB Specialty Clinics for Upper Limb Disorders"

We are writing to advise you that the Office of Research Ethics (ORE) has granted administrative approval to the above-named research study. The level of approval is based on the following role(s) of the University, as you have identified with your submission:

- Graduate Student research – hospital-based only
- Storage or analysis of De-identified Personal Information (data)

This approval does not substitute for ethics approval, which has been obtained from your hospital Research Ethics Board. Please note that you do not need to submit Annual Renewals, Study Completion Reports or Amendments to the ORE unless the involvement of the University changes so that ethics review is required. Please contact the ORE to determine whether a particular change to the University's involvement requires ethics review.

Best wishes for the successful completion of your project.

Yours sincerely,

Daniel Gyewu
Research Ethics Board Manager- Health Sciences

OFFICE OF RESEARCH ETHICS
McMurrich Building, 12 Queen's Park Crescent West, 2nd Floor, Toronto, ON M5S 1A8 Canada
Tel: +1 416 946-3273 • Fax: +1 416 946-5763 • ethics.review@utoronto.ca • http://www.research.utoronto.ca/for-researchers-administrator/ethics/
UNIVERSITY OF TORONTO

OFFICE OF THE VICE PRESIDENT, RESEARCH

PROTOCOL REFERENCE # 27144

December 8, 2011

Dr. Dorcas Beaton
DEPT OF OCCUPATIONAL THERAPY
FACULTY OF MEDICINE

Mr. Kenneth Tang
DEPT OF OCCUPATIONAL THERAPY
FACULTY OF MEDICINE

Dear Dr. Beaton and Mr. Kenneth Tang,

Re: Administrative Approval of your research protocol entitled, "Work limitations among injured workers recovering from upper-limb disorders: measurement issues and validation of a conceptual framework"

We are writing to advise you that the Office of Research Ethics (ORE) has granted administrative approval to the above-named research protocol. The level of approval is based on the following role(s) of the University of Toronto (University), as you have identified with your submission and administered under the terms and conditions of the affiliation agreement between the University and the associated TAHSN hospital:

- Graduate Student research - hospital-based only
- Storage or analysis of De-identified Personal Information (data)

This approval does not substitute for ethics approval, which has been obtained from your hospital Research Ethics Board (REB). Please note that you do not need to submit Annual Renewals, Study Completion Reports or Amendments to the ORE unless the involvement of the University changes so that ethics review is required. Please contact the ORE to determine whether a particular change to the University's involvement requires ethics review.

Best wishes for the successful completion of your research.

Yours sincerely,

Daniel Gyewu
REB Manager

OFFICE OF RESEARCH ETHICS
McMillan Building, 12 Queen's Park Crescent West, 2nd Floor, Toronto, ON M5S 1A9 Canada
Tel: +1 416 946-3773 • Fax: +1 416 946-3763 • ethics.review@utoronto.ca • http://www.research.utoronto.ca/researchers-administrators/ethics/