University of Alberta

# MEASUREMENT PROPERTIES OF A FUNCTIONAL CAPACITY EVALUATION ADMINISTERED ON WORKERS' COMPENSATION CLAIMANTS WITH LOW BACK PAIN

by

DOUGLAS PAUL GROSS

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY

in

REHABILITATION SCIENCE

FACULTY OF REHABILITATION MEDICINE

Edmonton, Alberta

FALL 2003

*Your file Votre référence*
*ISBN: 0-612-87980-1*
*Our file Notre référence*
*ISBN: 0-612-87980-1*

# University of Alberta

**Library Release Form**

**Name of Author:** *Douglas Paul Gross*

**Title of Thesis:** *Measurement Properties of a Functional Capacity Evaluation Administered on Workers' Compensation Claimants with Low Back Pain*

**Degree:** *Doctor of Philosophy*

**Year this Degree Granted:** *2003*



Date: Aug 11, 2003.

# University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **MEASUREMENT PROPERTIES OF A FUNCTIONAL CAPACITY EVALUATION ADMINISTERED ON WORKERS' COMPENSATION CLAIMANTS WITH LOW BACK PAIN** submitted by **Douglas Paul Gross** in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY in Rehabilitation Science**.

Michele Crites Battie

David Cassidy

Sharon Warren

Ambikaipakan Senthilselvan

Tapio Videman

Margareta Nordin

DATE: August 11, 2003

## ABSTRACT

Functional capacity evaluations (FCE) are standardized batteries of physical tests aimed at identifying a subject's performance potential for work. They are based on the theoretical concepts that physical loading in excess of an individual's physical capacity is responsible for injury and pain exacerbation, and that safety may be ensured through matching a worker's functional abilities with his or her required physical job demands. Workers' compensation boards and insurance companies internationally are increasingly using FCEs to inform return to work decisions. Major implications of FCE testing include personal consequences for workers undergoing testing and societal consequences including high administrative costs. The accuracy and validity of FCE determinations are of fundamental importance but have been minimally studied. This thesis investigates the measurement properties and theoretical foundation of the Isernhagen Work Systems' FCE administered on workers' compensation claimants with low back injuries.

Four studies of the measurement properties of the Isernhagen FCE were undertaken on workers' compensation claimants with low back pain. We examined the interrater and test-retest reliability, construct validity, and predictive validity of this tool. Claimants undergoing rehabilitation were enrolled and tested for the reliability study, whereas the validity studies relied on archived database information of the Alberta Workers' Compensation Board. Cross-sectional methods were used for the construct validity study, whereas a historical cohort

design and multivariable analytic techniques were used for the predictive validity studies.

We found the reliability of rater judgments while using this tool to be acceptable for clinical use, however, the validity of FCE determinations and the associated theoretical foundations are questionable. Subject performance and rater determinations during FCE appear to be influenced by subject perceptions of disability and pain intensity. Indicators of FCE performance appear to be weakly associated with timely recovery after controlling for other factors that influence the relationship between FCE performance and outcomes. Additionally, the validity of judgments of safe ability to return to work based on FCE results is suspect as better FCE performance was associated with increased risk of recurrent problems. Users of FCE information will have to decide whether the usefulness of the information obtained outweighs the associated administrative burdens.

## PREFACE

The doctoral research presented examines the measurement properties and theoretical foundations of a functional capacity evaluation (FCE) used internationally and by the Alberta Workers' Compensation Board (WCB-Alberta) to assist in managing the claims of injured workers. Specifically, essential properties were investigated in the group of claimants for which testing is most frequently undertaken, those with work-related low back pain. The objectives of the studies undertaken were developed in consultation with policy makers, clinicians and quality assurance professionals at the WCB-Alberta and after reviewing previous research.

Initially, reliability of rater judgments during the FCE was studied. This project was completed in the summer of 2000 and formed an in-lieu-of Master's project for the doctoral candidate. Funding for this project was obtained through a research grant sponsored jointly by the University of Alberta's Department of Physical Therapy and the Alberta Physical Therapy Association. Workers' compensation claimants being treated in an occupational rehabilitation program at Millard Health, the major WCB-Alberta rehabilitation facility, were enrolled. Next, a series of cohort studies were undertaken with the aim of determining the predictive validity of FCE determinations. A research grant was obtained from the Alberta Heritage Foundation for Medical Research – Health Research Fund. Historical cohorts were formed by extracting archived information from administrative and clinical databases of the WCB-Alberta and Millard Health. The database created also allowed investigation of FCE construct validity.

Findings of this research have been presented nationally and internationally. Results have also been shared with the WCB-Alberta and have contributed to informed appraisal of health care and claims management process and policy. Presently, further study is ongoing on the association between FCE and one-year self-rated pain and disability outcomes and on other potentially important confounding influences on the association between FCE and recovery outcomes.

## ACKNOWLEDGEMENTS

I wish to express sincere gratitude to my doctoral supervisor, Dr. Michele Crites Battié. Her expertise, wisdom and professionalism were greatly appreciated throughout my studies and my success was dependent on her influence. I would also like to thank my other committee members, Dr. David Cassidy and Dr. Sharon Warren for their thoughtful input and perspectives on methodological issues, and my examining committee members, Dr. Ambikaipakan Senthilselvan for his statistical tutoring and Dr. Tapio Videman for hours spent in theoretical discussion.

I wish to thank Paul Hagler, Associate Dean of Graduate Studies and Research, and the Faculty administrative staff including Angela Libutti and Vicki Trombley for their many contributions to my academic success.

I would like to acknowledge and thank Lawrence Chung, Kathryn Graham, Laurie Condratow and other staff members from the Workers' Compensation Board of Alberta who greatly assisted with data acquisition.

I am grateful for the financial support for either my doctoral studies or our research projects from the Alberta Heritage Foundation for Medical Research, the Canadian Institutes of Health Research, the Physical Medicine Research Foundation, the Killam Trusts, the University of Alberta's Department of Physical Therapy and the Alberta Physiotherapy Association.

Lastly, and most importantly, I would like to express gratitude to my family for their support and love during my studies. My beautiful wife Sondi provides constant support and encouragement, and my sons Grayson, Kepler, and Everett provide a source of motivation and daily rejuvenation.

## TABLE OF CONTENTS

## LIST OF TABLES

Page

## LIST OF FIGURES

Page

# CHAPTER 1
# Introduction

## 1.1 Overview

This thesis addresses the fundamental measurement properties and theoretical foundations of a functional capacity evaluation (FCE) used internationally and by the Alberta Workers' Compensation Board (WCB-Alberta) to inform return to work decisions for injured workers. This introductory chapter provides background on the use of FCE by the WCB-Alberta, describing the context in which this research was undertaken, and identifies the specific aims of this thesis research. The next five chapters have been published, are in press or have been submitted for publication elsewhere, yet each contributes to the overall evaluation of FCE measurement properties presented. Chapter two is a book chapter in which the controversy surrounding the theoretical basis of FCE is discussed and literature related to measurement properties reviewed. Chapters three to six describe primary research studies of various FCE measurement properties. The last chapter provides a synthesis and critical discussion.

## 1.2 Functional Capacity Evaluation

FCEs are standardized batteries of physical, physiological and functional tests aimed at identifying a subject's performance potential for work.[1, 2] Safe, maximal performance on specific work-related activities tested during FCE is compared to the required physical job demands of the subject's occupation to determine readiness to safely return to work following musculoskeletal injury.[3] FCEs are based on the theoretical concepts that physical loading that taxes or exceeds individual physical capacity is responsible for injury and pain exacerbation following injury, and that safety at work can be ensured through matching an individual's functional abilities to their required physical job demands.[4-6] Avoiding tasks that exceed a worker's ability should, in theory, reduce most of the risk of injury and pain exacerbation, and allow the individual to perform sustained, productive work.

FCE is being increasingly used by workers' compensation boards and insurance companies internationally to inform return to work decisions.[7] Numerous FCE protocols are available, marketed and currently used, including the Isernhagen Work Systems' (Duluth, MN).[8, 9]

### 1.3 The Isernhagen Work Systems' FCE

The Isernhagen Work Systems' FCE protocol is made of a series of tests representing the physical demands of work outlined in the American Dictionary of Occupational Titles, and includes lifting, carrying, pushing, and pulling among other tasks (see Table 1-1 for all tasks in the protocol).[10, 11] A kinesiophysical approach to testing is used in which a subject's safe, maximal performance levels are judged by the administering clinician using physiological and biomechanical observations.[12] Functional ability is conceptualized as separate and distinct from pain intensity ratings. Performance on each task is stopped when further upgrades to subject performance are judged unsafe and maximal performance levels are determined. Isernhagen Work Systems provides a list of biomechanical and physiological criteria to observe during testing but recommends that each facility using the protocol define specific performance endpoints which are considered indicative of safe, maximal performance for each task (Table 1-2).[10] Testing occurs over 5-8 hours divided into two half-day periods, with consistency of performance judged between days of assessment to allow determination of sustainable work levels.

### 1.4 FCE Use by the WCB-Alberta

The Isernhagen FCE protocol is used internationally and throughout the WCB-Alberta authorized health care provider network to guide return to work and claims management decisions for injured workers.[12] This FCE was selected and is valued by the WCB-Alberta due to its reliance on rater observations of biomechanical and physiological signs as opposed to subject reports of acceptable

performance limits. Demonstrated functional impairment is attributed to the compensable injury if a worker is judged to have exerted maximal effort.[12, 13]

In 1998, WCB-Alberta sponsored 956 standardized Isernhagen FCEs for injured workers.[14] Of these, 567 (59%) were completed at Millard Health, the WCB-Alberta's major rehabilitation facility in Edmonton.[14] Approximately 40% of the functional assessments undertaken are performed on claimants with low back injuries, which represents the largest group assessed according to anatomical site of injury.[15] The majority of claimants assessed are English-speaking (90%), employed (70%), and male (69%). The current cost to the WCB-Alberta for completion of a formal Isernhagen FCE is $750 plus associated medical costs for determining suitability for participation in maximal performance testing. Since 1998, the number of claimants undergoing FCE each year has increased annually, with 1166 injured workers assessed in 2002 at Millard Health alone.[16]

### 1.5 Return to Work Determinations With FCE

The increasing use of FCE to inform decision-making has arisen in large part due to an inability to adequately judge recovery and ability to return to work in some claimants using traditional physical examination and diagnostic imaging studies.[17] In some injury cases, such as fractures or amputations, work-related functional abilities may not show a direct relationship with the severity of the workers' physical impairment.[18,19] Often workers with severe impairment to bodily structures are able to function at levels required for work. Additionally, physical examination and diagnostic imaging studies are frequently inconclusive or do not demonstrate abnormalities that explain the claimants' reported work-related functional limitations.[20] This includes most cases of low back pain, which forms the majority of chronic disability claims in most jurisdictions and the majority of claimants assessed with FCE by the WCB-Alberta.[21,22] In order to avoid the limitations inherent in relying on impairment-level observations for return to work decision-making, direct measurement of work-related functional ability using FCE has been advocated.[3, 8]

Within the WCB-Alberta system, complex claims management decisions are made by case managers assigned to work with individual injured workers.[23] Such decisions include whether a worker is ready to return to work or whether ongoing indemnity benefits or associated support are required. Claims management decisions are guided by formal policy and procedures outlined by the WCB-Alberta and informed by various sources of information including reporting from health care professionals who have previously examined or treated the injured worker, diagnostic imaging studies, and reporting from specialized work assessment units where FCEs are performed.[23] Therefore, FCE results are used, in part, to guide return to work determinations.

Workers are typically referred for FCE testing within the WCB-Alberta system when they have surpassed expected injury healing times, have plateaued with medical and rehabilitative treatment, yet continue to report difficulties related to their compensable condition which hinder return to work.[24] FCE is performed to identify a worker's functional tolerances and ability to safely return to work. This is completed through comparing maximum performance levels on activities tested during the FCE to the associated physical job requirements of the worker's job.[9] Determining physical job demands is accomplished through employer report and direct work site measurement or, when these are unavailable, worker self-report with a standardized questionnaire.[25] Direct work site visits are ideal and potentially most valid, but are not possible when workers are unemployed or when the remoteness of the work place makes transportation to the site unfeasible.[26] Human Resources Development Canada's National Occupational Classification ratings of physical job requirements for occupational groups including strength, limb coordination, and body position are also used at times to augment worker self reports of required duties.[27]

Workers who demonstrate functional abilities that meet or exceed job demands on all tasks in the FCE protocol are considered ready and safe to return to work.[3]

Immediate return is recommended and such return is anticipated to be free of future pain exacerbation or injury recurrence as the worker is judged capable of tolerating all required physical load exposures.[4] Suspension of indemnity benefits and claim closure are potential consequences for workers found fit for work, regardless of whether a job is available to return to.[23]

Decisions are more complicated for workers who do not demonstrate the ability to perform all required job demands.[23] For employed subjects demonstrating functional tolerances below required work levels, return to work with physical restrictions or further rehabilitation is typically recommended depending on individual circumstances. Further treatment may be undertaken if the worker has not previously participated in all potential and available rehabilitation or pain management programs. Return to work performing modified duties within the activity levels determined as safe and acceptable using FCE could take place if an employer is available and willing to accommodate the restrictions of the injured worker. These modified duties may be performed on a temporary or permanent basis. Permanent physical restrictions are recommended when the injured worker has exhausted all possible medical and rehabilitative treatment, experienced prolonged periods of disability and work loss (typically greater than two years), and have physical and functional limitations that can clearly be linked to the original injury sustained.[28, 29]

When suitable duties are not available or the worker is unemployed, FCE performance levels may be used to guide vocational rehabilitation and future job searching through identification of transferable abilities and skills.[23] Injured workers in such situations may be faced with giving up chosen trades and occupations and accepting alternative employment that has been determined appropriate given their functional limitations and restrictions. This may involve retraining or employment that does not provide the same financial earning capacity as the workers' preinjury work. Thus, major decisions are made in the

lives of injured workers based, in part, on results of FCE testing. FCE must have acceptable measurement properties and provide consistent and valid information.

### 1.6 Essential FCE Measurement Properties

Fundamental properties of measurement tools deemed especially important in the area of FCE include interrater and test-retest reliability, and predictive validity.[30,31] Reliability refers to the consistency of a measure with interrater reliability defined as the extent to which multiple raters provide consistent rating and test-retest reliability defined as the extent to which multiple applications of a test provide consistent results.[32] Validity is the extent to which a measure assesses what it is intended to measure with predictive validity being the extent to which a test is associated with a subsequent meaningful event, which in the case of FCE is sustained return to work.[32] Predictive validity is an especially important property for FCE, as return to work is influenced by multiple factors in addition to functional abilities and occupational loading requirements.[33,34] Because of the multifactorial nature of return to work following musculoskeletal injury, the adequacy of FCE's theoretical foundations has been questioned due to its focus on physical load exposures in excess of an individual's physical capacity as the primary barrier to return to work.[35, 36]

Numerous literature reviews have been published which consistently emphasize that further research is needed related to the reliability and validity of all FCE protocols, including the Isernhagen Work Systems'.[7,30,31,37] Meaningful selection of a tool such as FCE can only be performed when the usefulness and validity of the information gained through use of the tool can be weighed against the associated administrative burdens.

### 1.7 Statement of the Problem

FCE use is becoming increasingly common in Alberta and in occupational health clinics worldwide to assist in the management of injured workers. More frequently workers' compensation boards and major insurers are relying on FCE

results to inform return to work decisions. Major implications of FCE testing include personal consequences for injured workers undergoing testing and societal consequences including high administrative costs. The accuracy and validity of FCE determinations are, therefore, of fundamental importance but have been minimally studied. Additionally, the adequacy of the theoretical foundations of FCE is controversial.

### 1.8 Purpose

The overall purpose of this thesis research was to examine essential measurement properties and the theoretical foundations of the Isernhagen Work Systems' FCE. Specific research purposes included:

1) Determining the inter-rater and test-retest reliability of safe, maximal lifting determinations during kinesiophysical FCE.
2) Investigating the construct validity of kinesiophysical FCE testing.
3) Investigating the predictive validity of the Isernhagen Work Systems' FCE in determining timely and sustained return to work alone and while controlling for other factors influencing return to work.

Findings of these studies are generalizable to workers' compensation claimants with low back injuries, undergoing FCE testing for purposes of claims decision-making. The predictive validity studies represent the first investigations of the association between an indicator of the matching relationship between FCE performance and required job demands and measures of timely and sustained return to work. Results will provide important knowledge for users of FCE information, allowing informed selection of FCE as a measurement tool and appropriate interpretation of results.

## TABLE 1-1

| Tasks Included in the Formal Isernhagen Work Systems' Functional Capacity Evaluation Protocol | |
|---|---|
| **Performed on Day One** | **Performed on Day Two** |
| Floor-to-Waist Lift | Repeat Day One Lifts (and carries if variation observed) |
| Overhead Lift | Squatting |
| Horizontal Lift | Crouching |
| Push | Kneeling |
| Pull | Crawling |
| Right Side Carry | Standing |
| Left Side Carry | Sitting |
| Front Carry | Handgrip |
| Elevated Work | Hand Coordination |
| Forward Bend in Sitting | Walking |
| Forward Bend in Standing | Step Ladder Climbing |
| Rotation in Sitting | Stair Climbing |
| Rotation in Standing | Balance |

## TABLE 1-2

| **Physiological and Biomechanical Signs Observed by Raters When Judging Safe, Maximal Effort During Kinesiophysical Functional Capacity Evaluation** |
|---|
| 1) Muscle bulging of prime movers<br>2) Involuntary use of accessory muscles<br>3) Altered body mechanics including counter-balancing or use of momentum<br>4) Loss of equilibrium<br>5) Increased base of support<br>6) Decreased efficiency and smoothness of movement<br>7) Cardiovascular signs including heart rate and breathing patterns<br>8) Peripheralization of radicular or referred symptoms. |

## 1.9 References


1. Innes E, Straker L. A clinician's guide to work-related assessments: 1-purposes and problems. *Work* 1998; 11:183-189.
2. Abdel-Moty E, Fishbain DA, Khalil TM, et al. Functional capacity and residual functional capacity and their utility in measuring work capacity. *Clin J Pain* 1993; 9:168-73.
3. Isernhagen SJ. *The comprehensive guide to work injury management.* Gaithersburg, Md.: Aspen Publishers, 1995:xxxii, 821.
4. Schonstein E, Kenny DT. The value of functional and work place assessments in achieving a timely return to work for workers with back pain. *Work* 2001; 16:31-38.
5. McGill SM. The biomechanics of low back injury: implications on current practice in industry and the clinic. *J Biomech* 1997; 30:465-75.
6. Keyserling WM. Workplace risk factors and occupational musculoskeletal disorders, Part 1: A review of biomechanical and psychophysical research on risk factors associated with low-back pain. AIHAJ 2000; 61:39-50.
7. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther* 1998; 78:852-66.
8. Lechner D, Roth D, Straaton K. Functional capacity evaluation in work disability. *Work* 1991; 1:37-47.
9. Isernhagen SJ. Return to work testing: functional capacity and work capacity evaluation. *Orthop Phys Ther NA* 1992; 1:83-98.
10. Functional Capacity Evaluation Manual, Edition One. Duluth, MN: Isernhagen Work Systems, 1997.
11. United States. Employment and Training Administration. *Dictionary of occupational titles.* Indianapolis: JIST Works, 1991:2 v. in 1 (1404 ).
12. Isernhagen SJ. Functional capacity evaluation: rationale, procedure, utility of the kinesiophysical approach. *J of Occup Rehab* 1992; 2:157-168.
13. Egan DA. *Functional assessments, work hardening, and work conditioning: A review for the Workers' Compensation Board of Nova Scotia.* 1996.

14. Work Assessment Centre Program Evaluation Summary Report. Alberta Workers' Compensation Board Health Care Services, 1998.
15. Work Assessment Centre Results. Millard Rehabilitation Centre, 2003.
16. Work Assessment Centre Results. Millard Health, 2003.
17. Hunt DG, Zuberbier OA, Kozlowski AJ, et al. Are components of a comprehensive medical assessment predictive of work disability after an episode of occupational low back trouble? *Spine* 2002; 27:2715-9.
18. World Health Organization. *International Classification of Functioning, Disability and Health Final Draft.* Geneva, Switzerland, 2001.
19. Mooney V. Impairment, disability, and handicap. *Clin Orthop* 1987:14-25.
20. Frank JW, Brooker AS, DeMaio SE, et al. Disability resulting from occupational low back pain. Part II: What do we know about secondary prevention? A review of the scientific evidence on prevention after disability begins. *Spine* 1996; 21:2918-29.
21. Frank JW, Kerr MS, Brooker AS, et al. Disability resulting from occupational low back pain. Part I: What do we know about primary prevention? A review of the scientific evidence on prevention before disability begins. *Spine* 1996; 21:2908-17.
22. Work Assessment Centre Program Evaluation Summary Report. Alberta Workers' Compensation Board Health Care Services, 1998.
23. Workers' services (Returning to work). Workers' Compensation Board Alberta Website, accessed April 30, 2003.
24. Work Assessment Centres Business Procedure. Alberta Workers' Compensation Board Claimant Services, 2001.
25. Lindstrom I, Ohlund C, Nachemson A. Validity of patient reporting and predictive value of industrial physical work demands. *Spine* 1994; 19:888-93.
26. Hansson GA, Balogh I, Bystrom JU, et al. Questionnaire versus direct technical measurements in assessing postures and movements of the head, upper back, arms and hands. *Scand J Work Environ Health* 2001; 27:30-40.

27. National occupational classification, 2001: Occupational descriptions. [Ottawa]: Human Resources Development Canada, 2001.
28. Alberta Disability Duration Guidelines: Workers' Compensation Board Alberta, 2003.
29. Policies and information manual: Workers' Compensation Board Alberta: Policy: 04-04 Part 1, 2003.
30. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13:125-52.
31. Velozo CA. Work evaluations: critique of the state of the art of functional assessment of work. *Am J Occ Ther* 1993; 47:203-9.
32. Finch E, Brooks D, Stratford P, Mayo N. *Physical rehabilitation outcome measures: A guide to enhanced clinical decision making*. Toronto: Canadian Physiotherapy Association, 2002.
33. Crook J, Milner R, Schultz IZ, Stringer B. Determinants of occupational disability following a low back injury: a critical review of the literature. *J Occup Rehab* 2002; 12:277-95.
34. Truchon M, Fillion L. Biopsychosocial determinants of chronic disability and low-back pain: a review. *J Occup Rehab* 2000; 10:117-42.
35. Rudy TE, Dieber SJ, Boston JR. Functional capacity assessment: Influence of behavioural and environmental factors. *J Back Musculoskel Rehabil* 1996; 6:277-288.
36. Hadler NM. Back pain in the workplace. What you lift or how you lift matters far less than whether you lift or when. *Spine* 1997; 22:935-40.
37. Innes E, Straker L. Reliability of work-related assessments. *Work* 1999; 13:107-24.

# CHAPTER 2

## Functional Capacity: Evaluation, Theoretical Aspects and Evidence

*A version of this chapter has been submitted for publication. Gross and Battié 2002. Textbook chapter in Pain Research and Clinical Management Series - Measuring and Managing Patients, Practitioners and Therapies in Rehabilitation, Guest Ed. Maureen Simmonds*

### 2.1 Introduction

You have been asked by a major insurance carrier to independently examine an injured worker to assist in management of the case. The client is a 41-year-old male long-haul truck driver who reports developing low back pain (LBP) twelve weeks ago at work shortly after lifting a large freight container off his trailer. His pain is localized to the lumbar spine, does not radiate, and yet has not subsided. He has not worked since the accident and has been receiving indemnity benefits from the insurance company. He reports his job requires long periods of driving and heavy materials handling occasionally when loading freight. Physical examination revealed limited trunk mobility but normal neurological functioning. X-rays demonstrate moderate signs of degeneration at the L5/S1 level. The questions posed to you by the client's case manager include:

1) Is the client ready to return to regular work duties?
2) If no, what physical activity restrictions are required?

These questions delineate two of the most sought after pieces of information from health care providers in cases of work-related LBP and other injuries resulting in work absence. Decisions based on the answers given have enormously important implications for the injured worker, the employer and insurance carrier. Yet, few health care providers are confident in making such determinations. Numerous approaches exist to obtain answers to the two questions posed, from reliance on medical testing and physical examination findings to self-report measures

completed by the client. An approach commonly used in compensation and private insurance systems is to utilize results of functional capacity testing.

**2.2 Functional Capacity - What is function and how should it be measured?**

The goal of functional capacity testing is to directly measure ability and capacity for specific tasks, however, what exactly should be quantified is controversial.[1] Some authors consider testing of functional capacity to be limited to measures of specific body systems or tissue function, while others recommend measures of ability for real-life physical tasks.[2-4] This has resulted in the development of numerous testing methods reportedly measuring functional capacity including tests of joint mobility, muscle strength, general body fitness, self-report ratings of ability and batteries of such tests.[5-7] Such variations are due, in part, to an inadequate definition of the construct of functional capacity.[8]

Consideration of the definitions of function and capacity that have been presented recently by the World Health Organization and outlined in the *International Classification of Functioning, Disability and Health (ICF* - see table 1*)* may provide a basis for further clarifying the construct of functional capacity.[9] *Capacity* is described as an individual's ability to execute a task or an action. Functional capacity thus indicates the highest probable level a person may reach in a given domain at a given moment within limits imposed by physiological systems or anatomical structures. In our driver with LBP and all work-related injury scenarios, the domain and tasks are specific to the unique duties of individual jobs. This definition broadly encompasses many factors as possible limits to capacity, including but not limited to such physiological quantities as muscular strength, joint mobility, cardiovascular endurance, neuromuscular control and reaction time. These variables function together and jointly contribute to capacity within the identified domain. Adequate performance of systems and tissues contribute to maximal capacity. When structural or physiological impairment exists, restriction on a given task may or may not be evident; compensation for the impairment may take place if other body tissues have such

capability. An example is reaching an object from the floor through hip and knee flexion when trunk flexion is compromised. Measurement of functional capacity, therefore, must go beyond impairment-based measures and encompass actual ability measures to capture maximal function possible.

Physical and physiological limits may restrict function, but external environmental influences also affect performance. The World Health Organization defines *performance* as executing a task in the current environment. Capacity may change with alterations in the physical or contextual environment with potential for participation restriction if the environment is not conducive. Such factors as temperature, humidity level and quality of standing surface have been found to influence performance.[10,11] As well, environment is intimately associated with psychological state, which also has important influences on performance levels. As psychological and other contextual factors have the potential for dramatic influences on functional testing and, therefore, decisions based on such testing, these issues will be discussed in a separate section.

**2.3 Functional Capacity Testing in Rehabilitation**

The rehabilitation disciplines are uniquely suited to functional testing and have coined the term Functional Capacity Evaluation (FCE) for such testing.[12] Rehabilitation, as the name implies, aims at returning an individual to former functional levels. This typically does not entail curative treatments or goals of ridding the client of a disease state. Instead, in the case of work-related injury the rehabilitation professional's goal is to return the worker to pre-injury ability levels to allow performance of work and other duties. This involves optimizing work capacity and participation in the work environment, as well as determination of such capacity - the aim of FCE.

In the area of work injury, rehabilitation professionals commonly use FCEs to assist in case management decision-making and in expediting the return-to-work process.[12-14] Direct measures of actual or simulated work tasks are made and tests

of manual handling, sustained postures and positioning, and ambulation are commonly included. Most functional evaluations incorporate the twenty physical work functions described in the U.S. Department of Labor's *Selected Characteristics of Occupations as Defined in the Revised Dictionary of Occupational Titles* (DOT).[12,15]

Other reasons have been reported where FCE may be utilized to guide clinical decision-making.[16,17] These include setting baseline measurements in work hardening or functional restoration programs to assist in treatment planning, or as an outcome measure from such programs. FCEs are also used to guide vocational planning or retraining programs after work-related injury or in performing pre-employment and post-offer screening. Another recommended use of FCE has been to assess the extent of disability to assist in judgments of permanent impairment or determination of wage-earning potential in litigation cases.

**2.4 Theoretical Basis**

FCE assessment and decision-making processes have a theoretical basis in injury theory, which also forms the foundation of most current compensation systems.[18] Injury is thought to occur when the tolerance of a body tissue to withstand an imposed load is exceeded.[19] This overloading may potentially result in tissue damage or injury, with accompanying pain and disability. Pain is considered a direct result of injury and tissue harm. Once healing has occurred, functional testing is used in an attempt to determine safe and tolerable loading. The performance levels demonstrated on FCE are judged to be the physical exposure levels the worker is capable of experiencing without further injury occurrence.[2]

In the case of the male truck driver with LBP who is presently 12 weeks post-injury, a functional assessment may be recommended as normal healing timeframes for soft tissue injury have been surpassed. The FCE may reveal he is capable of lifting to a medium level and has difficulty with prolonged periods of sitting due to frequent position changes and pain behaviours observed on a sitting

test. Prolonged driving and pre-injury levels of heavy materials handling would be judged as unsafe and temporary alternative job placement may be recommended. If suitable modified duties were available, a rapid return to the worksite could be facilitated with a gradual return to full duties planned as the worker recovers. This idealistic scenario and management process seems intuitive and reasonable when seen in the context of an injury model. However, there are practical and theoretical issues that pose challenges to such idealistic FCE application and interpretation such as the assumption of sincerity of effort, which underlies the validity of FCE. Another thorny issue is that of determining "safe maximum" from observation of apparent effort and adherence to current biomechanical materials handling standards.

### 2.5 Evaluation Approaches

Many standardized, formal FCE protocols currently exist and are available for purchase.[16] While subtle differences exist between all protocols making each somewhat unique, the tests can be categorized into one of two general categories, psychophysical and kinesiophysical (also known as biomechanical or physiological).[1] In psychophysical testing, performance is stopped by the client once he or she perceives maximal effort or acceptable loading has been reached. The kinesiophysical approach differs in that it places maximal effort determinations in the hands of trained clinicians relying on observation of biomechanical and physiological responses to testing. Once safe, maximal performance is achieved for a particular task, as determined using an operational definition provided by the protocol developer, testing is stopped and another task attempted.

Tests of acceptable load using the psychophysical approach have been commonly reported in the peer-reviewed literature.[20-22] However, the validity and usefulness of this testing method in the management of injured workers has come into question. In a study of healthy workers by Mital, subjects asked to lift using a psychophysical approach tended to overestimate their abilities when compared to

performance over a 12-hour session.[22] None of the ten subjects tested were able to sustain the weight levels they found acceptable during the functional testing. This has potentially serious consequences when psychophysical lifting techniques are used to make decisions matching workers' capacities to job demands.

During kinesiophysical testing, operational definitions of maximal effort include such biomechanical and physiological signs as counterbalancing, use of momentum or accessory muscles and increased heart rate.[2] Cardiovascular limits for most FCEs have been adopted from the American College of Sports Medicine fitness guidelines and commonly 85% of maximal heart rate is considered an upper limit.[23] "Proper" body mechanics are required during testing aimed at minimizing load on the musculoskeletal system and performance is considered unsafe if guidelines for proper body mechanics are breeched. Safety recommendations are most commonly based on biomechanical investigations of tissue response to loading, such as pressure studies of lumbar intervertebral discs.[24] When performance on FCE approaches unsafe levels or unsafe techniques are adopted as determined by the assessing clinician, the test is stopped and another activity attempted. Maximal performance is determined, along with level of effort exerted by the client, using this approach. "Sub-maximal" performance as indicated by inconsistency or a lack of observed signs of effort is often looked upon negatively by claims adjudicators.

A psychophysical testing protocol for manual handling has been put forward by Snook and variations on this approach are still used in FCE testing[25,26], while the Isernhagen Work Systems' FCE (Duluth Minnesota) and Ergoscience's Physical Work Performance Evaluation (Birmingham, AL) are examples of tests incorporating kinesiophysical/ biomechanical methods. More recently, attempts to merge the two approaches have been made with clinician observation being combined with client report of effort.[27,28] No research has been published comparing or contrasting the various testing approaches to determine comparability or superiority of one method over another. However, in

compensation and private insurance systems the kinesiophysical/ biomechanical approach is often preferred over the psychophysical given the controversial nature of some claims and questionable motivation of some clients.[2]

### 2.6 Decision-Making Processes

As mentioned, most standardized FCE protocols incorporate testing of the physical demands of work outlined in the Dictionary of Occupational Titles. Content validity has been defined for FCE as whether or not the FCE contains all of these physical tests.[29] Safe, tolerable performance levels are determined during FCE for these various forms of loading and work restrictions are outlined. Measurements of ability on such tasks aim to outline physical capacity for work, thus answering the two important questions posed to our healthcare professional at outset. If a worker's function on an FCE exceeds or matches critical job demands, it is determined that he or she could successfully return to pre-injury work levels. If job demands are not met, work restrictions are set for use in guiding return to suitable employment including job modification or retraining if needed. Determination or prediction of ability to return to work is likely the primary use of FCE.

While most FCE protocols employ and make decisions using contemporary standards of safe lifting and materials handling based on biomechanical investigations, controversy still exists about what performance methods are "safest".[19] Most biomechanical modeling studies have been performed using cadaver or animal tissue, with questionable relevance to live human subjects. Practical difficulties related to "unsafe" performance are also identified from observation of experienced workers. Most workers prefer to use stooping techniques, identified as producing higher loading on lower lumbar spinal segments, instead of squatting when performing low-level activities.[24] This may be due to the fact squatting requires a higher metabolic expenditure.

Work restrictions, theoretically, reduce the risk of further injury and subsequent pain reporting as work performance should remain below injury thresholds if restrictions are abided by. While injury theories provide an appealing foundation from which to make decisions regarding work ability, some difficulties exist in the case of LBP and other non-specific soft tissue disorders that form the majority of compensation cases. Often no tissue insult, pathology, or physical "injury" can be identified despite numerous medical tests.[30] As in our scenario, such examples are all too often seen in clinical practice with management strategies based on injury theory being woefully inadequate.

Disability associated with most low back "injuries" is a complex, pain-mediated phenomenon. Drastic differences in levels of disability are often observed between individuals with similar physical findings. Disability in the form of work loss has been found to be poorly associated with pain levels.[31,32] Oleske et al studied 88 employees who experienced a work-related low back disorder and found the number of lost workdays did not correlate with LBP intensity, or the Oswestry or Roland Disability Questionnaires. These findings run counter to current injury theories that hold pain and disability as a direct result of injury. As well, even workers in sedentary jobs report pain and experience time loss, contrary to the concept of load-induced injury and pain. When taken in the broader context of psychological, socioeconomic and cultural influences that affect pain perceptions and reaction to pain, FCE testing based on physical performance and job characteristics may be of limited usefulness in determining readiness for return to work.

### 2.7 Psychometric Properties

Despite problematic theoretical foundations, FCEs continue to be popular measurement tools, with important decisions made based on their findings. Psychometric properties must, therefore, be evaluated.[33] Some research on reliability and validity has been published in peer-review journals. Responsiveness, an important property for FCE as it is often used pre-and post-

rehabilitation to determine outcome, has not been studied.[17,34,35] Two recent reviews have thoroughly described the available evidence dealing with reliability and validity of work-related functional assessments and will be summarized along with more recent research in these areas.[29,36]

### 2.8 Reliability

The issue of reliability, while important for all clinical tests in common use, is particularly important in the case of kinesiophysical FCE. Adequate reliability of psychophysical testing has been demonstrated, with judgment on the part of the observing clinician not required.[25] The agreement and stability of determinations related to maximal performance and level of effort based on clinical observations, as in the case of kinesiophysical FCE, are of particular interest as they may be influenced by the observing clinician.

Overall, reliability studies of effort determinations using broad categories have reported values within acceptable limits for clinical use.[37-40] One study of reliability of kinesiophysical FCE looked specifically at safe, maximal manual handling determinations as opposed to level of effort and found both inter-rater and test-retest interclass correlation coefficient values above 0.75, which is defined by Portney and Watkins as acceptable for performance-type measures in clinical use.[41, 42] Reliability has mainly been studied for manual handling tasks, which have been reported as the primary determinants for measuring a job's physical demands. However, one study did look at all FCE sub-tests for the Physical Work Performance Evaluation protocol in 50 subjects with a variety of musculoskeletal disorders and found acceptable reliability.[43] While reliability appears adequate, more research on various forms of reliability of entire FCE protocols is needed, particularly for test-retest and inter-rater with varying therapist experience, training, and disciplines.

## 2.9 Validity

Validity has been defined as whether a test measures what it is purported to measure, along with the usefulness and meaningfulness of decisions made using test results.[29] The major forms of validity, which together form the test's overall validity, include face, content, criterion-related (concurrent and predictive), and construct validity. Innes and Straker reported acceptable face and content validity, as determined by a group of experts, for most currently available functional work tests.[29] FCE protocols incorporating work-specific testing and the physical demands of work as reported in the DOT are considered to have good face and content validity.

Less formal research has been completed on other important forms of validity. Construct validity has been evaluated for kinesiophysical FCE in comparison to self-reported disability measures in a sample of 42 subjects receiving workers' compensation for work-related symptoms. Moderate associations were found between FCE, the Oswestry Low Back Disability Questionnaire and the Bodily Pain Scale, indicating these tools are measuring related but different constructs.[44] No work on concurrent validity has been performed comparing FCE findings to actual performance on real world work tasks. In fact, when concurrent validity has been examined other FCEs have been used as the gold standard of comparison rather than actual work performance.[45] Four predictive validity studies have been reported for three separate FCE protocols.[15,46-48] Two have been published in manuscript format while two were merely abstracts. All reported positive results for FCE predictive ability, however only Fishbain et al reported a statistical evaluation of other factors along with DOT-FCE information potentially contributing to prediction. They found that prediction of work status was dependent on reported pain levels and workers' compensation status. The FCE sub-tests found to contribute to return-to-work included hand dexterity testing, a counter-intuitive finding given the sample included only individuals with chronic LBP. In the three other reports, follow-up was poor with only around half of the

subjects contacted and inadequate consideration was given to other factors that may have contributed to the therapists' predictions.

Individual components of FCE protocols have been studied for predictive ability, with physical findings from musculoskeletal examination and psychophysical lifting (without comparison to work demands) consistently reported as unrelated to future pain reporting or return-to-work.[20,21,49-51] One study of individuals undergoing low back surgery reported on the ineffectiveness of prescribed work restrictions on future work ability, demonstrating a decreased likelihood of successful return-to-work when restrictions were given, however, restrictions were not set using FCE but clinician opinion on the part of a surgeon.[52] Research is needed to clarify the role and validity of using FCE information for determination of ability for and prediction of return to work.

### 2.10 Challenges in Interpreting FCE Findings

The important decisions made based on FCE necessitate not only sound test metrics but also sound interpretation of outcomes. Among the important issues that may potentially affect FCE interpretations are fluctuations in pain and disability levels, the effect of the assessment setting or context on results, and the ability to obtain valid information concerning frequency tolerances and critical job demands.

### 2.11 Fluctuations in Pain and Disability

In a test-retest reliability study of maximal manual handling determinations, despite methodological controls taken to ensure stability in function between testing occasions, a portion of the subjects with work-related low back problems did not feel they could participate in the second session due to pain.[41] This was an important finding with implications for how FCE interpretations are made in claimants with painful conditions. The marked variability in perceived function observed may have been heightened by the ease of not participating in the study, as compared to evaluations performed within compensation or insurance systems.

Yet, the findings likely represent true variability in perceptions of pain-related ability or motivation to perform over time, which may be exhibited in the work environment, as well. None of the subjects attributed their exacerbation to the first testing occasion. Such fluctuation and recurrence of pain and associated disability levels are common in cases of LBP.[30, 53] FCE attempts to take a "snap-shot" of function, which is then generalized to actual ability. Due to the short nature of the test, alterations in pain and disability may not be detected even if the assessment takes place over multiple sessions.

In the case of permanent disability determinations for such conditions as work-related idiopathic low back pain, with or without disc degeneration or pathology, FCE determinations are particularly suspect. The assumption of having reached a stable outcome and maximal recovery would seem to be prerequisite of such determinations and questionable for pain mediated disability from ill defined conditions.

**2.12 Assessment Setting**

Another difficulty in applying FCE results to real work settings arises due to constraints imposed by the assessment environment.[33,54] As mentioned earlier, functional capacity should be assessed in a standardized fashion, including a standardized setting, resulting in more accurate comparisons between occasions and individuals. However, the assessment environment itself likely influences performance on FCE in subtle, yet important ways that potentially affect interpretations of test results. An assessment performed in a workers' compensation setting (i.e. rehabilitation facility owned by insurance carrier) may have different results from one performed in a private clinic or at the worksite. Heightened anxiety over the testing site or motivation to demonstrate disability to the insurance provider may diminish the claimant's performance. Environmental influences on FCE results have not been studied and remain an important area for future research. Due to external influences on performance, standardized

measurement of functional capacity is currently recommended with testing preferably taking place at the work site for higher applicability.

### 2.13 Frequency Tolerances

Determinations made during FCE include safe, maximal ability for work but also frequency tolerances for those same activities. For example, a client may be capable of lifting to a heavy level but how long this performance intensity can be sustained must also be known prior to recommending return to work. Psychophysical assessment leaves such frequency decisions to the client being tested, however when outcomes were compared to actual full-day performance, the assessment predictions proved inadequate.[22] For kinesiophysical testing, no standard exists for making such extrapolations, with the decision often being left up to the clinical decision making ability of the therapist using observations made during the assessment.[17] A client using maximal effort at a particular level is typically judged to be able to sustain such performance on a rare or very occasional basis. The accuracy of such judgments has not been evaluated.

Some efforts to improve tolerable frequency judgments have been made. Saunders et al used multivariable regression to create models for purposes of determining frequency tolerances, but the usefulness of these estimates was reported as questionable by the authors.[55] With little justifying empirical evidence, some FCE developers recommend longer testing sessions or testing over multiple sessions to more adequately define tolerance for sustained activity. The accuracy of longer assessments over shorter (and therefore less expensive) FCEs has also not been evaluated. Until a standardized, validated method of measuring performance tolerances and sustainability for work is published, such determinations must be considered no more than well-informed guesses.

### 2.14 Critical Job Demands

One last, yet extremely important issue influencing FCE decision-making is the process employed for determining a job's critical demands. Evaluating what a

worker does at a particular worksite is important for FCE as observed ability is compared to work demands and a decision for work-readiness made based on this comparison. When a job is available, a worksite visit is recommended as the gold standard for determination of critical demand levels prior to FCE. When a worksite visit is not an option (i.e. when geographic, budgetary, political or other influences necessitate against it), reliance on either an employer or worker report is required. This is an unwanted situation as validity of these reports is unknown. Injured workers covered by workers' compensation may report higher work demands than those not so covered, profoundly influencing FCE decisions.[56,57]

Even though a work site visit is considered the gold standard, practical difficulties arise in determining all critical job tasks and their associated demand levels. Many jobs vary day to day and physical demands fluctuate accordingly. A brief work site encounter by a healthcare professional is unlikely to establish a comprehensive outline of all job demands required by the employee. In addition, no standardized work site visit protocol has been established and validated. Questions arise such as: should the physical demands be limited to only those tasks assessed on FCE or should all potential tasks be reported on? Should the visit be limited to physical work demands, or should psychological, emotional and other stressors arising in the work site be considered? These and similar questions related to the workplace environment have important implications for return to work decisions and FCE. Currently no adequately tested and standardized method of measuring critical job demands has been published.

### 2.15 Contextual Factors

A dynamic interaction exists between health conditions and contextual factors in determining functioning and disability, especially in conditions mediated by pain reporting. Contextual factors that potentially influence disability, include, but are not limited to, variables in psychological, socioeconomic, cultural, and economic realms. Demonstrated capacity is likely influenced by motivation to return to work, feelings of depression or anxiety, self-efficacy or perceptions of disability,

which are in turn influenced by economic state, potential for earnings, or cultural beliefs surrounding disability.[58-60] An injured individual off work and receiving just enough money to procure necessities of life from a compensation system may perform differently on FCE if he or she has a well-paying job to return to versus no job to return to. Variations in cultural or family support of an individual's disability also may affect performance levels. Contextual factors must be considered when making return-to-work decisions based on results of functional testing.

Psychological and socioeconomic factors have consistently been found to significantly influence outcome in studies of LBP reporting and prediction of recovery.[49,61,62] In a recent study of individuals with sub-acute LBP by van der Giezen et al, level of job satisfaction and breadwinner status more powerfully influenced recovery than physical work demands or perceived work ability.[61] Studies by Infante-Rivard and McIntosh and colleagues found the type of industry the client worked in at the time of injury reporting to be predictive of return to work, with subjects in private or construction industries likely to experience delayed recovery.[63,64] Fishbain et al have presented empiric evidence of the need to incorporate psychological and economic factors along with FCE findings when predicting return to work.[15] In a predictive validity study of their functional evaluation, 185 chronic LBP patients were followed up at 6, 12, 24, and 30-month intervals after completing a rehabilitation program and receiving a FCE. A multivariable analysis performed found that FCE results were unable to accurately predict who would be employed at follow-up without also considering reported pain levels and workers' compensation status. These last two variables were consistently the strongest predictors of outcome and FCE results played only a minor role in prediction.

Returning to our initial scenario, had our truck driver been single with few economic responsibilities to family, facing return to an unsupportive workplace environment in which he had negative associations with supervisory staff or

coworkers, he may not have been motivated to perform to his highest levels during FCE. Alternatively, had he been the sole wage earner in a family of five, without the availability of modified duties and the only options being return to full duties or job termination, the worker may have been motivated to perform to higher levels during testing regardless of pain. Such contextual issues greatly complicate interpretation of FCE outcomes and must be given consideration.

### 2.16 Summary

Functional capacity measures including formal, standardized FCE protocols are currently available and one major use is to assist with return to work decision-making. Specifically, these measures aim to answer questions related to fitness for work and whether physical restrictions are warranted. While FCE is commonly used, little rigorously conducted, peer-reviewed research has been published indicating superiority of one protocol over another, or concerning FCE in general. Some evidence has been reported indicating that FCE subtest outcomes are reliable between raters and occasions. However, less has been published regarding validity and the meaningfulness of FCE interpretations. Most studies reported have substantial methodological flaws and further research is required into concurrent, predictive, and construct validity issues. Many factors including psychological state, assessment environment, and social and economic contextual factors influence function, performance and return to work decisions and need to be acknowledged when interpreting FCE results. Clinicians performing such assessments and others using them in decision-making need to be cognizant of the important influences of psychological, social and other contextual factors on functional capacity and related testing.

## TABLE 2-1

| Definitions From the International Classification of Functioning, Disability and Health |
| --- |
| **Impairment** – Problems in body function or structure such as significant deviation or loss. |
| **Activity** - The execution of a task or action by an individual. |
| **Participation** – Involvement in a life situation. |
| **Functioning** - All body functions, activities and participation as an umbrella term. |
| **Capacity** - An individual's ability to execute a task or an action, the highest probable level of functioning that a person may reach in a given domain at a given moment. |
| **Performance** – What an individual does in his or her current environment. |
| **Domain** - A practical and meaningful set of related physiological functions, anatomical structure, actions, tasks, or areas of life. |

## 2.17 References


1. Innes E, Straker L. A clinician's guide to work-related assessments: 2-Design problems. *Work* 1998; 11:191-206.
2. Isernhagen SJ. Functional capacity evaluation: rationale, procedure, utility of the kinesiophysical approach. *J Occup Rehab* 1992; 2:157-168.
3. Rantanen P. The dynamometric trunk muscle strength test in low back trouble. Medical Faculty. Helsinki: University of Turku, PhD Thesis, 2001.
4. Takala EP, Viikari-Juntura E. Do functional tests predict low back pain? *Spine* 2000; 25:2126-32.
5. Deyo RA. Measuring the functional status of patients with low back pain. *Arch Phys Med Rehabil* 1988; 69:1044-53.
6. Waddell G, Somerville D, Henderson I, Newton M. Objective clinical evaluation of physical impairment in chronic low back pain. *Spine* 1992; 17:617-28.
7. Fairbank J. Revised Oswestry Disability questionnaire. *Spine* 2000; 25:2552.
8. Abdel-Moty E, Fishbain DA, Khalil TM, et al. Functional capacity and residual functional capacity and their utility in measuring work capacity. *Clin J Pain* 1993; 9:168-73.
9. WHO. *International Classification of Functioning, Disability and Health Final Draft.* Geneva, Switzerland, 2001.
10. Madeleine P, Voigt M, Arendt-Nielsen L. Subjective, physiological and biomechanical responses to prolonged manual work performed standing on hard and soft surfaces. *Euro J Appl Phys & Occup Phys* 1998; 77:1-9.
11. Nag PK, Bandyopadhyay P, Ashtekar SP, Kothari D, Desai H, Nag A. Human work capacity under combined stress of work and heat. *J Hum Ergo* 1996; 25:105-13.
12. *Occupational Health Physical Therapy Guidelines.* American Physical Therapy Association, Amended1997:APTA website.

13. Gibson L, Strong J. A review of functional capacity evaluation practice. *Work* 1997; 9:3-11.

14. Mooney V, Leggett S. Objective functional assessment in disability case resolution. *J Occup Rehab* 1997; 7:217-24.

15. Fishbain DA, Cutler RB, Rosomoff H, Khalil T, Abdel-Moty E, Steele-Rosomoff R. Validity of the dictionary of occupational titles residual functional capacity battery. *Clin J Pain* 1999; 15:102-10.

16. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther* 1998; 78:852-66.

17. Lechner D, Roth D, Straaton K. Functional capacity evaluation in work disability. *Work* 1991; 1:37-47.

18. Hadler NM. Back pain in the workplace. What you lift or how you lift matters far less than whether you lift or when. *Spine* 1997; 22:935-40.

19. McGill SM. The biomechanics of low back injury: implications on current practice in industry and the clinic. *J Biomech* 1997; 30:465-75.

20. Gallagher RM, Rauh V, Haugh LD, et al. Determinants of return-to-work among low back pain patients. *Pain* 1989; 39:55-67.

21. Milhous RL, Haugh LD, Frymoyer JW, et al. Determinants of vocational disability in patients with low back pain. *Arch Phys Med Rehabil* 1989; 70:589-93.

22. Mital A. The psychophysical approach in manual handling - a verification study. *Human Factors* 1983; 25:485-491.

23. American College of Sports Medicine., Franklin BA, Johnson EP. *ACSM's guidelines for exercise testing and prescription.* Philadelphia: Lippincott Williams & Wilkins, 2000:xix, 368.

24. Wilke HJ, Neef P, Caimi M, Hoogland T, Claes LE. New in vivo measurements of pressures in the intervertebral disc in daily life. *Spine* 1999; 24:755-62.

25. Alpert J, Matheson L, Beam W, Mooney V. The reliability and validity of two new tests of maximum lifting capacity. *J Occup Rehab* 1991; 1:13-29.

26. Matheson LN, Mooney V, Grant JE, et al. A test to measure lift capacity of physically impaired adults. Part 1--Development and reliability testing. *Spine* 1995; 20:2119-29.

27. Matheson LN, Mooney V, Holmes D, et al. A test to measure lift capacity of physically impaired adults. Part 2--Reactivity in a patient sample. *Spine* 1995; 20:2130-4.

28. Greenwood KM, DeNardis R. An assessment of the reliability of measurements made using the Melbourne Protocol and the Hanoun Multi-cervical Unit. *Unpublished manuscript*, 2000.

29. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13:125-52.

30. Frank JW, Kerr MS, Brooker AS, et al. Disability resulting from occupational low back pain. Part I: What do we know about primary prevention? A review of the scientific evidence on prevention before disability begins. *Spine* 1996; 21:2908-17.

31. Linton SJ. The relationship between activity and chronic back pain. *Pain* 1985; 21:289-94.

32. Oleske DM, Andersson GB, Lavender SA, Hahn JJ. Association between recovery outcomes for work-related low back disorders and personal, family, and work factors. *Spine* 2000; 25:1259-65.

33. Velozo CA. Work evaluations: critique of the state of the art of functional assessment of work. *Am J Occ Ther* 1993; 47:203-9.

34. Mayer TG, Gatchel RJ, Kishino N, et al. Objective assessment of spine function following industrial injury. A prospective study with comparison group and one-year follow-up. *Spine* 1985; 10:482-93.

35. Sachs BL, David JA, Olimpio D, Scala AD, Lacroix M. Spinal rehabilitation by work tolerance based on objective physical capacity assessment of dysfunction. A prospective study with control subjects and twelve-month review. *Spine* 1990; 15:1325-32.

36. Innes E, Straker L. Reliability of work-related assessments. *Work* 1999; 13:107-24.

37. Gardener L, McKenna K. Reliability of occupational therapists in determining safe, maximal lifting capacity. *Aus Occ Ther J* 1999; 46:110-19.

38. Reneman MF, Jaegers SMBJ, Westmaas M, Goeken LNH. The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. *Work* 2002; 18: 23-27.

39. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgements of level of lift effort in a kinesiophysical Functional Capacity Evaluation. *Work* 1999; 12:145-50.

40. Smith RL. Therapists' ability to identify safe maximum lifting in low back pain patients during functional capacity evaluation. *JOSPT* 1994; 19:277-81.

41. Gross DP, Battié MC. Reliability of safe maximal lifting determinations of a functional capacity evaluation. *Phys Ther* 2002; 82: 364-71.

42. Portney LG, Watkins MP. *Foundations of clinical research : applications to practice*. Upper Saddle River, NJ: Prentice Hall, 2000:xiv, 742.

43. Lechner DE, Jackson JR, Roth DL, Straaton KV. Reliability and validity of a newly developed test of physical work performance. *J Occup Med* 1994; 36:997-1004.

44. Hart DL. Relation between three measures of function in patients with chronic work-related pain syndromes. *J Rehab Outcome Measure* 1998; 2:1-14.

45. Dusik LA, Menard MR, Cooke C, Fairburn SM, Beach GN. Concurrent validity of the ERGOS work simulator versus conventional functional capacity evaluation techniques in a workers' compensation population. *J Occup Med* 1993; 35:759-67.

46. Lechner D, Sheffield GL, Page JJ, Jackson JR. Predictive validity of a functional capacity evaluation: The Physical Work Performance Evaluation. *Phys Ther* 1996; 76:S81.

47. Smith S, Cunningham S, Weinberg R. The predictive validity of the functional capacities evaluation. *Am J Occup Ther* 1986; 40:564-67.

48. Smith SL, Cunningham S, Weinberg R. Predicting reemployment of the physically disabled worker. *Occ Ther J Res* 1986; 3:179.

49. Bigos SJ, Battié MC, Spengler DM, et al. A prospective study of work perceptions and psychosocial factors affecting the report of back injury. *Spine* 1991; 16:1-6.

50. Biering-Sorensen F. Physical measurements as risk indicators for low-back trouble over a one-year period. *Spine* 1984; 9:106-19.

51. Troup JD, Foreman TK, Baxter CE, Brown D. 1987 Volvo award in clinical sciences. The perception of back pain and the role of psychophysical tests of lifting capacity. *Spine* 1987; 12:645-57.

52. Hall H, McIntosh G, Melles T, Holowachuk B, Wai E. Effect of discharge recommendations on outcome. *Spine* 1994; 19:2033-7.

53. Carey TS, Garrett JM, Jackman AM. Beyond the good prognosis. Examination of an inception cohort of patients with chronic low back pain. *Spine* 2000; 25:115-20.

54. Rudy TE, Lieber SJ, Boston JR. Functional capacity assessment: Influence of behavioural and environmental factors. *J Back Musculo Rehab* 1996; 6:277-288.

55. Saunders RL, Beissner KL, McManis BG. Estimates of weight that subjects can lift frequently in functional capacity evaluations. *Phys Ther* 1997; 77:1717-28.

56. Hadler NM, Carey TS, Garrett J. The influence of indemnification by workers' compensation insurance on recovery from acute backache. North Carolina Back Pain Project. *Spine* 1995; 20:2710-5.

57. Lindstrom I, Ohlund C, Nachemson A. Validity of patient reporting and predictive value of industrial physical work demands. *Spine* 1994; 19:888-93.

58. Estlander AM, Vanharanta H, Moneta GB, Kaivanto K. Anthropometric variables, self-efficacy beliefs, and pain and disability ratings on the isokinetic performance of low back pain patients. *Spine* 1994; 19:941-7.

59. Gibson L, Strong J. Assessment of psychosocial factors in functional capacity evaluation of clients with chronic back pain. *Br J Occup Ther* 1998; 61:399-404.

60. Kaplan GM, Wurtele SK, Gillis D. Maximal effort during functional capacity evaluations: an examination of psychological factors. *Arch Phys Med Rehab* 1996; 77:161-4.

61. van der Giezen AM, Bouter LM, Nijhuis FJ. Prediction of return-to-work of low back pain patients sicklisted for 3-4 months. *Pain* 2000; 87:285-94.

62. Gatchel RJ, Polatin PB, Mayer TG. The dominant role of psychosocial risk factors in the development of chronic low back pain disability. *Spine* 1995; 20:2702-9.

63. Infante-Rivard C, Lortie M. Prognostic factors for return to work after a first compensated episode of back pain. *Occup Environ Med* 1996; 53:488-94.

64. McIntosh G, Frank J, Hogg-Johnson S, Bombardier C, Hall H. Prognostic factors for time receiving workers' compensation benefits in a cohort of patients with low back pain. *Spine* 2000; 25:147-57.

# CHAPTER 3

## Reliability of Safe Maximum Lifting Determinations of a Functional Capacity Evaluation

*A version of this chapter has been published. Gross and Battié 2002. Physical Therapy. 82: 364-371.*

### 3.1 Introduction

One challenge faced by clinicians when treating individuals off work due to low back pain (LBP) is balancing early return to work recommendations with concerns of delayed recovery or pain exacerbation that could result from premature spinal loading.[1,2] Functional Capacity Evaluations (FCE) are measurement tools created to assist in determining safe, tolerable levels of function, and for predicting when an individual is ready to return to work duties.[3] In FCE, a trained clinician attempts to measure an injured worker's maximum physical abilities for job-related tasks. Tasks assessed may include manual handling, positional work, and activities requiring ambulating or hand coordination. Manual handling has been described as the primary determinant for rating a job's physical demands.[4] If the worker does not have a job to return to, the information gained is used during vocational rehabilitation or job placement services by comparing results to known demands of other occupations. The determinations of performance levels during FCE, therefore, have far reaching implications with respect to return to work and employability.

Various types of FCE exist. Two common approaches have been described as psychophysical and kinesiophysical.[5] Psychophysical FCEs place the worker in control and performance is stopped when the worker believes maximal function has been reached. The kinesiophysical approach places the administering therapist in control and tasks are stopped when biomechanical signs of maximal effort are observed, such as accessory muscle usage and counterbalancing. A set of standardized criteria for judging increased effort and maximal levels are

outlined for the kinesiophysical method. Theoretically, this ensures the safety of the injured worker as assessment is to be stopped prior to overexertion.

If FCE is to be considered a useful tool, reliability and validity must be demonstrated.[6,7,8,9] As determinations require judgments regarding safety, some variance is expected with repeated measures within individual therapists and between therapists. In addition, variations in subject performance due to wellness on the day of the evaluation, motivation, pain levels or interactions between the client and therapist conducting the evaluation may influence results. With these considerations in mind, inter-rater and test-retest reliability have been viewed as the most important forms of test reliability.[6,10,11]

Some work has been done looking at reliability of various aspects of kinesiophysical testing.[4,12,13,14] Some limitations of previous studies have been the utilization of videotaped subject performance resulting in a loss of some clinical information such as cardiovascular responses to testing used in maximal effort determination and gained during real-life observation. Secondly, studies done using real-life observation have not overcome the potential bias resulting from one rater influencing the judgment of the other when stopping the test. Lastly, all previous studies used a categorical outcome variable, rather than the interval level outcome of weight handled, as is determined in routine FCE testing.

Our goal was to determine the inter-rater and test-retest reliability of lifting determinations of maximal safe manual handling levels during kinesiophysical FCE using the Isernhagen Work Systems'* protocol in a medically stable, LBP population receiving workers' compensation.

---

* Isernhagen Work Systems,
1015E. Superior Street,
Duluth, MN 55802

## 3.2 Methods

*3.2.1 Subjects*

The study sample was one of convenience and drawn from a rehabilitation center of the Workers' Compensation Board of Alberta. Subject inclusion criteria were selected to ensure the safety of participating subjects and to enroll subjects at a point in recovery when FCE testing is routinely performed. Inclusion criteria were: off work and receiving compensation for LBP; participation in an occupational rehabilitation program (subjects had plateaued with treatment and were in the process of being discharged); medical stability as determined by a physician[15]; absence of metastatic disease, non-stable musculoskeletal conditions or uncontrolled medical disorders; and a physician's determination of suitability for FCE following review of an ECG for subjects over 45 years of age. Written informed consent was obtained from all subjects prior to enrollment. Subjects were free to stop testing or withdraw at any time.

Subjects were recruited through consultation with treating rehabilitation teams to identify eligible clients nearing the end of their treatment program. All prospective subjects were scheduled for FCE testing at discharge irrespective of study participation. Twenty-eight subjects with LBP were enrolled in the study from April to July, 2000. At a 0.05 alpha level, no significant differences were observed between our subjects and the entire group of clients with low back injuries discharged from the center during the data collection period using chi-square tests for categorical variables and independent-sample t-tests for continuous variables. Variables compared were: age, gender, National Occupation Classification (NOC) code, job attachment status, duration of injury, and length of time off work, as determined from the center's clinical database for all subjects discharged (Table 1).

Basic client characteristics and medical history data were collected at the time of enrollment and subjects were asked three proposed core outcome measure questions advocated by Deyo et al.[16] From the core outcome questions asked of

subjects, the modal bothersomeness of pain and interference with work due to pain were both moderate. However, subjects most frequently reported being very dissatisfied with their symptoms, despite having nearly completed their rehabilitation program.

Five occupational therapists were enrolled to perform testing and act as raters. All raters had previously been trained by representatives of Isernhagen Work Systems, were conducting FCEs in clinical practice, and had at least 5 years of experience using kinesiophysical observation techniques. Raters reported an average length of time trained in and performing kinesiophysical FCEs of 7.4 years (range 5-9 years). All were full-time employees and reported an average completion of 4.4 evaluations per week using kinesiophysical observation methods. Three were male, two female and the average length of time spent in professional practise was 15.4 years.

Prior to the study, kinesiophysical principles and an operational definition of maximal effort were reviewed with the raters. Raters were asked to observe the following signs of increased effort in judging when subjects had reached maximal, safe levels:

1) Muscle bulging of prime movers
2) Involuntary use of accessory muscles
3) Altered body mechanics including counter-balancing or use of momentum
4) Loss of equilibrium
5) Increased base of support
6) Decreased efficiency and smoothness of movement
7) Cardiovascular signs including heart rate and breathing patterns
8) Peripheralization of radicular or referred symptoms.

*3.2.2 Study Protocol*

A repeated measure design was used with the goal of independent, yet simultaneous observation of each subject by two raters. This occurred on two

separate occasions separated by two to four treatment days, a time period in which no significant change was expected in subject performance while allowing some time to lessen recall of the previous performance. Between occasions, raters continued to perform regular work duties including other FCEs. Time of day and place of assessment were held constant. Testing took place within the subject's last week of a rehabilitation program.

Assessment tasks of floor-to-waist, waist-to-crown and horizontal lifting, and front, right and left side carrying were completed. The specific protocol for each lift and carry was followed as outlined in the Isernhagen Work System's FCE Manual[17] with sets of five repetitions being completed for each subtest at each successive weight level.

To obtain independent yet simultaneous observation by the raters, three raters were selected randomly from the group of five for each enrolled subject. The first rater selected was referred to as the primary rater. The primary rater's responsibility was to converse with the subject, guide the subject through testing and upgrade weight in the lifting unit. Upgrades were possible in 1.1, 2.2, or 4.5kg increments or any combination of these. The primary rater was the only individual with exact knowledge of the weight lifted or carried as the other raters were not able to see into the lifting unit and did not observe weight upgrades. The primary rater documented the amount of weight lifted or carried during each set and other raters did not have access to this documentation. The primary rater also had the major responsibility for ensuring subject safety and was to stop testing if they judged safety to be obviously compromised.

The next two raters selected were referred to as secondary raters. They observed performance and prompted the primary rater through testing, but were instructed not to interact with subjects. Secondary raters were instructed not to observe or talk to each other, but were allowed to walk around the assessment area for observation angle of choice.

Secondary raters were blinded to each other's prompts and determinations to avoid any potential bias in the following manner. For each subject and subtest, the primary rater progressed testing from low to higher weight levels. Sets for each subtest were sequentially numbered on both the primary and secondary rater documentation forms. The primary rater documented the weight level and secondary raters documented their observations for each set. After observing subject performance on an individual set, secondary raters documented their observations, then were allowed to prompt the primary rater nonverbally as to whether the weight in the lifting unit should be upgraded or testing stopped because maximal levels had been determined. They did this through pointing to one of two closely placed boxes on the bottom of their documentation forms, with the words "STOP" and "UPGRADE" written in the boxes. Documentation stations were placed far enough apart for secondary raters not to see their companion's prompt. Primary raters walked between documentation stations to receive feedback. When a particular set was judged as maximal, the secondary rater pointed to the box stating "STOP", documented their observations and circled the corresponding set number. All further prompting by this secondary rater was made by indicating "STOP". Testing continued with the primary rater upgrading weight until both secondary raters indicated "STOP". At the end of testing, all raters sealed their documentation forms in envelopes and delivered them to a secure location.

Maximal weight levels in kilograms, as judged by the secondary raters, were determined through comparison of the primary rater's documentation with the corresponding set circled by each secondary rater. The factor leading to test termination for each lifting subtest also was recorded by the secondary raters. Limiting factors were categorized as physical maximum, cardiovascular limitation, non-functional time, or subject desire/ pain.

*3.2.3 Data Analysis*

Intraclass Correlation Coefficients (ICC – Shrout and Fleiss Model 1,1[18]) with 95% confidence intervals (CI) were obtained for inter-rater and test-retest reliability of secondary raters' judgments of maximal weight levels measured in kilograms. Two comparisons per subject were available for both forms of reliability. Because ICC values diminish when variance in a sample decreases, which would be the case if duplicate or repeat measures for both raters were used in analysis of test-retest data, calculations were performed separately for the first and second secondary raters' determinations.[18] Also, inter-rater ICC calculations were performed using the first session, with values from the second session used to judge stability of results.

Paired t-tests with alpha level set at 0.05 were used to compare mean differences between occasions on each subtest to determine if a testing effect existed between days of testing. Kappa values and percentage agreement were calculated for agreement on subject performance limiting factors. The statistical software package SPSS† was used for ICC, t-test and Kappa calculations.

ICC is currently the statistic of choice for reliability analyses of interval data, however classical test theory may not provide a complete understanding of this issue. Generalizability theory may provide a more effective conceptual approach, and comprehensive reviews have been published.[19,20,21] Generalizability coefficients and estimated variance components for the factors controlled for were calculated. Generalizability coefficients represent the relative generalizability of a measurement to the total range of possible scores for that measurement, with results ranging from 0 to 1 similar to ICC. Estimated variance components show the contribution made to total variance by each controlled factor. These statistics were calculated using formulas discussed elsewhere.[20]

† SPSS Inc.
233 S. Wacker Drive
Chicago, IL 60606

### 3.3 Results

Of the 28 subjects enrolled, 75% participated in both testing sessions. Three subjects did not attend on day two and three others attended, but stated they did not feel capable of any manual handling due to LBP. Partial data sets were obtained from six subjects due to rater reporting error, subject desire, primary rater overruling a decision to upgrade (1 subject each), and lack of time to complete testing (3 subjects) and these are reflected in the various numbers of subjects per subtest in the tables.

ICC values for inter-rater reliability on session one ranged from 0.95 to 0.98 (Table 2). Results were equally high for the second session. Test-retest ICC values ranged from 0.78 to 0.94 when calculated using the first secondary raters' scores and from 0.81-0.91 when using the second secondary raters' scores (Table 3). The high degree of similarity between the ICC values and confidence intervals for the duplicate measures provides an indication of the stability of the test-retest values.

Mean scores of weight lifted between days were compared for all subjects completing testing. Consistently, subjects lifted more on day two but these differences were statistically significant only for low level lifting (day one 21.8 kilograms, day two 25.7 kilograms; $p$=0.007) and front carrying (day one 32.2 kilograms, day two 34.7 kilograms; $p$=0.015).

Findings from analysis of agreement for factors limiting test performance are summarized in Table 4. Kappa values ranged from 0.47 to 1.00 and overall percentage agreement was 86.4% (235/272). Raters both rated a particular subject's performance as physical maximum on 68.8% of the comparisons. Of the 37 incidents where there were disagreements, the same weight level was judged as maximum in 30 cases, with 26 of these being judged as physical maximum versus subject desire.

Estimated variance components and generalizability coefficients were also calculated and are shown in Table 5. Estimated variance components showed the highest portion of variance consistently resulted from between-subject variability (80.3-91.4%), as expected. However, with respect to sources of measurement inconsistencies, the greatest portion of variance was explained by the *subject-occasion* interaction (4.5-16.8%). Generalizability coefficients ranged from 0.92 to 0.96.

### 3.4 Discussion

Inter-rater reliability was excellent, with all subtest ICC values above 0.9. Results were similar when values from either day of testing were used in analyses. Also of interest is that on similar subtests ICC results were similar, i.e. right and left side carrying, possibly reflecting internal consistency. When ratings of subjects completing testing on both test sessions were analyzed, ICC values for test-retest reliability were lower (0.78-0.94) than those for inter-rater reliability. Test-retest reliability results were stable between secondary raters. Good generalizability was also seen as all generalizability coefficients were over 0.9.

In the present study, three subjects returned for day two of testing but stated they did not feel capable of participating in manual handling activities due to reported pain exacerbation. The ease with which subjects could withdraw or terminate testing may have led to more subjects declining testing during the second session than would have occurred under normal FCE test conditions. However, the subjects' beliefs and perceptions of pain, disability and physical capacity that led them to decline testing may represent valid influences on FCE results. The first test session was not cited as the reason for increased pain in any of the subjects declining.

The testing interval was selected to minimize functional change. Also, return to work was imminent in this group of subjects deemed medically stable, yet the

performance of some subjects varied between occasions. This is especially true of those subjects unwilling to participate on the second occasion. Variations in subjects' performance between days may be due to the reasons discussed previously such as wellness, motivation, pain level, etc. Another potential contribution to the observed variability is a testing effect in subjects participating in both days. Comparison of means between days, with significant increases on the second occasion for low level lifting and front carry, indicates that a testing effect likely did exist. It was not great enough, however, to diminish test-retest ICC values below acceptable levels.

Estimated variance components for subjects participating on both days clarify what factors were responsible for the variance observed. Consistently, *subjects* were responsible for the greatest variance, a desirable finding supporting the acceptable ICCs. The *subject-occasion* interaction, defined by Shavelson and Webb[20] as variance arising due to inconsistencies between occasions in particular subjects' performance, was consistently the second leading source of variance. The minimal residual variance in maximal ratings was made of various combinations of other factors depending on the subtest, but these factors contributed little to the total variance.

Due to the variability observed between days and the fact three subjects felt they could not participate on the second occasion, manual handling is recommended over a two-day period. The Isernhagen Work System's FCE protocol acknowledges client performance may vary between days and recommends a two-day session of manual handling ability.

Raters agreed substantially or perfectly on the performance-limiting factor for test termination on most subtests according to the Landis and Koch categorization for Kappa values.[22] Front and left side carrying agreement was moderate.

No previous study has looked specifically at the reliability of determinations of maximal levels using actual weight lifted, but other aspects of reliability of the kinesiophysical approach have been examined. When Isernhagen et al. studied inter-rater reliability of gross judgments of lifting effort, raters were able to accurately discriminate between "light" and "heavy" lifting effort (Kappa = 0.81).[4] Their study used videotape of the subjects' performance, therefore some clinical detail would have been lost. Smith studied the ability of trained and experienced therapists to reliably judge whether low back-injured patients can lift from the floor to waist with 'safe body mechanics' as operationally defined by the author.[14] Inter-rater Kappa values ranged from 0.62 to 0.64. In Smith's study, as in the Isernhagen study and a study by Gardener, videotape was used for viewing subject performance.[12] The present study's design allowed clinically realistic observation and gave access to all information gained during typical FCE while allowing simultaneous observation of subjects. The slightly higher reliability we found may be due to added information available to our raters such as subject cardiovascular responses, symptoms, and three-dimensional viewing.

In a study by Lechner et al, inter-rater reliability of maximal effort during another FCE protocol was examined.[13] In this assessment, maximal effort was determined through observation of body mechanics and lifting technique. Inter-rater Kappa values found for manual handling determinations within Dictionary of Occupational Titles' categories ranged from 0.62 to 0.88. These findings of substantial to almost perfect reliability are similar but slightly lower than our findings. As the FCE under study was newly developed, raters had minimal experience with total training time being approximately 20-24 hours. Conversely, raters in our study had at least five years of experience. As well, the study protocol by Lechner et al did not achieve independent observation between raters, resulting in a potential bias of one rater by the primary rater responsible for test termination.

One limitation of the present study affecting evaluation of test-retest reliability, in particular, was subject mortality. As noted previously, three subjects felt incapable of participating on day two of testing. Also, only partial data sets were obtained from six subjects due to rater reporting error, subject lack of desire to perform all subtests, primary rater overruling a decision to upgrade, or lack of time to complete testing. A diminished sample size resulted and may have altered reliability calculations had all subjects been tested on all subtests. Yet, the consistency seen when alternate rater or occasion ICC values were calculated, indicate the stability of the findings in the subjects tested. The effect on reliability when altering factors such as therapist discipline, level of therapist experience and setting, remains unknown.

### 3.5 Conclusions

Inter-rater reliability of kinesiophysical lifting and carrying determinations as conducted by experienced raters on a sample of low back-injured WCB claimants was excellent. Test-retest reliability, although lower, was generally good in subjects completing testing. A subgroup of subjects was unwilling to participate in the second day of maximal testing due to a reported increase in symptoms unrelated to FCE testing. Assessment of manual handling over more than one occasion is, therefore, recommended to capture variability in function between occasions.

## TABLE 3-1

| Characteristics of WCB Clients with Low Back Pain | | |
|---|---|---|
| **Characteristic** | **Subjects (n=28)** | **Eligible Clients (n=172)** |
| Gender (% male) | 71 | 71 |
| Age in Years (mean, range) | 41 (23-62) | 41 (19-65) |
| Occupation: | | |
| Truck Drivers (%) | 21 | 14 |
| Labourers (%) | 18 | 5 |
| Job Attached (%) | 71 | 61 |
| Median Duration of Injury (days) | 123 (71-584) | 136 (52-2921) |
| Mean Duration | 165 | 213 |
| Median Time Off Work (days) | 112 (54-255) | 114 (24-579) |
| Mean Time Off Work | 125 | 152 |

## TABLE 3-2

| Interrater Reliability for Session One | | | |
|---|---|---|---|
| TASK | ICC | 95% CI | N |
| Floor to Waist Lift | **0.98** | 0.96-0.99 | 27 |
| Waist to Overhead Lift | **0.96** | 0.92-.0.98 | 27 |
| Horizontal Lift | **0.96** | 0.91-0.98 | 27 |
| Front Carry | **0.96** | 0.90-0.98 | 25 |
| Right Carry | **0.96** | 0.91-0.98 | 24 |
| Left Carry | **0.95** | 0.90-0.98 | 23 |

## TABLE 3-3

**Test-retest Reliability ICC Values**

| TASK | Secondary Rater One | | | Secondary Rater Two | | |
|---|---|---|---|---|---|---|
| | ICC | 95% CI | N | ICC | 95% CI | N |
| Floor to Waist Lift | **0.78** | 0.51-0.91 | 18 | **0.83** | 0.60-0.93 | 18 |
| Waist to Overhead Lift | **0.84** | 0.63-0.93 | 18 | **0.81** | 0.56-0.92 | 18 |
| Horizontal Lift | **0.86** | 0.67-0.95 | 18 | **0.88** | 0.71-0.95 | 18 |
| Front Carry | **0.90** | 0.75-0.96 | 17 | **0.87** | 0.68-0.95 | 17 |
| Right Carry | **0.94** | 0.85-0.98 | 16 | **0.91** | 0.76-0.97 | 16 |
| Left Carry | **0.86** | 0.65-0.95 | 15 | **0.83** | 0.57-0.94 | 15 |

**TABLE 3-4**

**Rater Agreement on Performance Limiting Factor**

| Task | Kappa | Percent Agreement | Comparisons |
|---|---|---|---|
| Floor to Waist Lift | **0.64** | 79.2% | 48 |
| Waist to Crown Lift | **0.62** | 83.0% | 47 |
| Horizontal Lift | **0.77** | 97.5% | 48 |
| Front Carry | **0.47** | 82.2% | 45 |
| Right Carry | **1.00** | 100% | 43 |
| Left Carry | **0.56** | 87.8% | 41 |

## TABLE 3-5

**Generalizability Calculations**

| Task | Factor | Estimated Variance | Percent of Total Variance | Generalizability Coefficient |
|---|---|---|---|---|
| Floor to Waist Lift | Subject (S) | 435.8 | **83.6%** | 0.95 |
| | Rater (R) | 0.0 | 0.0% | |
| | Occasion (O) | 24.7 | 4.7% | |
| | S x R[a] | 1.8 | 0.3% | |
| | S x O[b] | 34.5 | **6.6%** | |
| | R x O[c] | 0.0 | 0.0 % | |
| | S x R x O[d] | 24.2 | 4.7% | |
| Waist to Overhead Lift | Subject | 127.2 | **80.3%** | 0.90 |
| | Rater | 0.0 | 0.0% | |
| | Occasion | 0.2 | 0.1% | |
| | S x R | 1.6 | 1.0% | |
| | S x O | 26.6 | **16.8%** | |
| | R x O | 0.5 | 0.3% | |
| | S x R x O | 2.3 | 1.5% | |

a - S x R = Subject-Rater Interaction

b - S x O = Subject-Occasion Interaction

c - R x O = Rater-Occasion Interaction

d - S x R x O = Residual, Error

Only two tasks are shown. Variance components from tasks not shown were similar.

## 3.6 REFERENCES


1. Abenhaim L, Rossignol M, Valat JP, et al. The role of activity in the therapeutic management of back pain. Report of the International Paris Task Force on Back Pain. *Spine.* 2000; 25(4 Suppl): 1S-33S.
2. Waddell G. A new clinical model for the treatment of LBP. *Spine.* 1987; 12:632-644.
3. Gibson L, Strong J. A review of functional capacity evaluation practice. *Work.* 1997; 9:3-11.
4. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgments of level of lift in kinesiophysical Functional Capacity Evaluation. *Work.* 1999; 12:145-150.
5. Isernhagen SJ. Functional capacity evaluation: rationale, procedure, utility of the kinesiophysical approach. *J Occup Rehabil.* 1992; 2:157-168.
6. Innes E, Straker L. Reliability of work-related assessments. *Work.* 1999; 13: 107-124.
7. Innes E, Straker L. Validity of work-related assessments. *Work.* 1999; 13: 125-152.
8. Sheikh K. Disability scales: assessment of reliability. *Arch Phys Med Rehabil.* 1986; 67: 245-249.
9. Velozo CA. Work evaluations: critique of the state of the art of functional assessment of work. *Am J of Occ Ther.* 1993; 47: 203-209.
10. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther.* 1998; 78:852-865.
11. Lechner D, Roth D, Straaton K. Functional capacity evaluation in work disability. *Work.* 1991; 1:37-47.
12. Gardener L, McKenna K. Reliability of occupational therapists in determining safe, maximal lifting capacity. *Aust Occ Ther J.* 1999; 46: 110-119.
13. Lechner DE, Jackson JR, Roth DL, Straaton KV. Reliability and validity of a newly developed test of physical work performance. *J Occup Med.* 1994;36: 997-1004.

14. Smith RL. Therapist's ability to identify safe maximum lifting in LBP patients during functional capacity evaluation. *JOSPT.* 1994;19: 277-281.
15. Hart DL, Isernhagen SJ, Matheson LN. Guidelines for functional capacity evaluations of people with medical conditions. *JOSPT.* 1993; 18:682-686.
16. Deyo RA, Battié MC, Beurskens AJHN, et al. Outcome measures for LBP research: A proposal for standardized use. *Spine.* 1998; 23: 2003-2013.
17. Isernhagen Work Systems *Functional Capacity Evaluation Manual*, Edition One, Duluth, MN, 1997.
18. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practise*, Second Edition. New Jersey: Prentice Hall, 2000.
19. Roebrock ME, Harlaar J, Lankhorst GJ. The application of generalizability theory to reliability assessment: An illustration using isometric force measurements. *Phys Ther.* 1993; 73: 386-395.
20. Shavelson RJ, Webb NM. *Generalizability Theory: A Primer.* London, New Delhi: Sage Publications, 1991.
21. Stratford PW, Norman GR, McIntosh JM. Generalizability of grip strength measurements in patients with tennis elbow. *Phys Ther.* 1989; 69: 276-281.
22. Landis RJ and Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33: 159-174.

# CHAPTER 4

## The Construct Validity of a Kinesiophysical Functional Capacity Evaluation Administered Within a Workers' Compensation Environment

*A version of this chapter has been accepted for publication. Gross and Battié 2003. Journal of Occupational Rehabilitation.*

### 4.1 Introduction

No gold standard exists for the measurement of functional capacity, despite a major need for such a tool in the area of compensated work-related injury. The inability to accurately measure functional capacity poses a challenge to treating clinicians responsible for monitoring treatment effectiveness, insurers responsible for maximizing recovery of injured workers to pre-injury states, and to injured workers experiencing functional deficit and work loss.

Functional Capacity Evaluations (FCE) are standardized batteries of physical performance and functional measures that are commonly used to determine a subject's ability to perform work related activities.[1] During FCE, an injured worker's performance on job-related tasks is measured and compared to his or her physical job demand levels. Recommendations are made based on FCE results regarding employability, including whether the worker can safely return to pre-injury or modified work. Many formal FCE protocols have been developed and are currently marketed.[2] Yet, recent reviews of the scientific literature have revealed a lack of peer-reviewed studies exploring the psychometric properties of work related functional assessment.[3, 4]

The Isernhagen Work Systems' (IWS, Duluth, MN) assessment protocol is one example of a kinesiophysical FCE.[5] In kinesiophysical testing, administering clinicians rely on observation of physiological and biomechanical signs of effort to determine safe, maximum performance levels.[6] Functional capacity is theorized as being separate and distinct from pain intensity. Observational criteria for

kinesiophysical judgments of safe, maximum effort published previously contain recommendations to monitor for peripheralization of radiating pain, however, pain intensity ratings are not considered.[5, 7] Elements of the IWS protocol have been studied in subjects with low back pain for some facets of reliability, with both inter-rater and test-retest reliability of determinations of lifting and carrying performance levels made by therapists using operational definitions reported as acceptable.[7,8] The construct validity of kinesiophysical FCE testing, however, has been minimally examined.

Construct validity is the extent to which a test behaves in accordance with hypotheses concerning how it should behave.[9] Hypotheses are generated through theoretical consideration of the magnitude and direction of expected relationships with other constructs. Because construct validity studies concern associations between abstract concepts, the magnitude of a relationship is judged to be low if correlation coefficients vary between 0 and 0.29, moderate between 0.3 and 0.59, and a strong relationship is judged if coefficients are above 0.6.[3,10] This interpretation of correlation magnitude is less stringent than that used in studies of concurrent validity in which a new measure is compared to an accepted gold standard and stronger associations are anticipated.

For kinesiophysical FCE, where administering clinicians judge maximum performance levels solely on biomechanical and physiological criteria, the underlying theory posits that functional capacity is separate from pain severity. Thus, performance on the FCE would be expected to correlate moderately with other measures of function or disability, but poorly with measures of pain intensity. Two previous studies have examined the relationship between FCE performance and perceived disability. Hart reported moderate correlations between maximum performance levels on various FCE lifting and carrying tasks and the Oswestry Disability Questionnaire in a study of 42 subjects with chronic low back pain.[11] Reneman et al also found moderate relationships in 64 chronic low back pain subjects when performance on FCE (categorized according to

Dictionary of Occupational Titles work levels[12]) was compared to the Oswestry Disability Questionnaire and Quebec Back Pain Disability Questionnaire.[13] No study could be located which examined the relationship between maximum performance on kinesiophysical FCE and ratings of pain intensity.

In the present study, we examined the construct validity of the IWS FCE by examining the pattern of relationships between pain intensity, perceived disability, and performance on kinesiophysical FCE administered on subjects with compensable low back conditions. We hypothesized that increasing functional performance levels on FCE would correlate moderately (r=0.3-0.6) with decreasing levels of perceived disability measured with the Pain Disability Index (PDI), as the overall construct of measurement is function in both cases. In addition, we hypothesized that increasing performance levels on kinesiophysical FCE would correlate poorly (r=0-0.29) with increasing ratings of pain severity on a pain visual analogue scale (VAS).

## 4.2 Methods

### *4.2.1 Design*

A cross-sectional study design was employed. Information on FCE, perceived disability, and pain severity were obtained in a clinical setting during application of a work assessment battery. Information obtained from the work assessment battery was being used for purposes of claims decision-making, including determination of fitness to return to work. Data on primary measures of interest and descriptive characteristics, including subject age, gender, diagnosis, duration of injury and employment status (whether or not they had a job to return to), were extracted from a clinical database. Data were obtained as part of a larger study approved by the University of Alberta's Health Research Ethics Board.

### *4.2.2 Subjects*

Subjects included all claimants meeting study inclusion criteria and seen for FCE between January 1, 1999 and December 31, 2000, at the major workers'

compensation rehabilitation facility in Alberta, Canada. Specific inclusion criteria, were: compensated work-related injuries with ICD-9 codes indicative of injury to the low back; FCE conducted at least 6 weeks after the date of accident; medical clearance to participate in FCE, including ECG testing on all subjects over 45 years of age; and not discharged for further medical investigation or treatment. Claimants referred for FCE testing within the Alberta workers' compensation environment are usually those who have met or surpassed expected injury healing times, have plateaued with medical and rehabilitative interventions, yet report ongoing difficulties related to their compensable condition. Such claimants are evaluated with FCE for purposes of determining fitness to return to work and employability.

*4.2.3 Measures*

**Functional Capacity Evaluation** - The IWS FCE protocol is used throughout the WCB-Alberta health care provider network. IWS representatives train all clinicians performing FCE at the centre from which subject data was obtained. This kinesiophysical FCE protocol includes a battery of physical performance tests representing the physical demands of work outlined in the Dictionary of Occupational Titles.[12] At the rehabilitation centre from which subjects came, FCE performance levels are compared to physical job demands on all tasks making up the protocol and a rating of meeting or failing the job demand is given for each task. The observational criteria for determining maximum performance levels on manual handling at the centre have been previously published.[7] For our analysis, the following two FCE performance measures were used:

a) *Average Maximum Weight Lifted* - The maximum amount of weight lifted or carried on all six manual handling tasks in the FCE protocol (low level, overhead and horizontal lifting, and front, left, and right side carrying) was averaged for each subject. These tasks were selected as they have been found to have good inter-rater and test-retest reliability.[7,8] Higher scores were indicative of better performance.

b) *Number of Failed Tasks* - The number of FCE tasks in the protocol rated as failed by the administering clinician was recorded. A failure is judged when subject performance does not meet or exceed the physical demand levels required at his or her employment. The number of failed tasks (out of 25) was summed and an overall FCE outcome score obtained for each subject, with higher scores indicative of worse performance. A clinician rating of the number of failed tasks as an indicator of FCE performance has not been studied previously for aspects of reliability or validity, yet such ratings are central to individual determinations of readiness to return to work.

The process for determining physical job demands for comparison with FCE performance at the rehabilitation facility varies depending on individual case characteristics, but is accomplished through employer report and direct work site measurements or, when these are unavailable, worker self-report using a standardized questionnaire. Significant challenges exist when attempting to accurately measure physical job demands, with direct work site measurements of physical exposure being preferable.[14] However, direct measurements are often impossible or impractical and worker self-report is the only source of information. Reliability and validity of the specific methods used at the centre to determine physical job demands have not been tested, however, the results are used routinely for making administrative decisions.

**Pain Disability Index** - The PDI is a measure of perceived disability due to pain. It has been tested in a variety of patient populations including those with chronic back pain.[15-17] The PDI asks patients to rate their level of disability on a 0-10 scale on seven areas of activity: family/home responsibility, recreation, social activity, occupation, sexual behaviour, self-care, and life-support activity. The subject's scores on these seven items are summed for a total score between 0 and 70, with higher scores indicative of higher levels of perceived disability. At times subjects omit answering some PDI items and previous researchers have recommended using a percentage PDI, calculated as the subject's total score

divided by the total possible score for only the items completed, to maximize available data. Gronblad et al found both overall PDI score and percentage score correlated strongly and nearly equivalently with the Oswestry Disability Questionnaire.[18] Both strategies were employed in the present analysis.

Previous investigators have reported good test-retest reliability, high internal consistency and good concurrent validity for the PDI when tested against the Oswestry Disability Questionnaire in a group of subjects with low back pain.[18,19] The PDI has been shown to accurately discriminate between patients who had just undergone back surgery and patients with low back pain maintaining full-time work.[20] Three published reports were located in which the PDI was compared to physical performance measures in samples of chronic low back pain patients.[19,21,22] Physical performance tests in these studies included squatting, spinal range of motion measures, arch-ups, and repetitive sit-ups, along with isometric lifting. The PDI consistently correlated moderately with these tests (r=0.30-0.61).

**Pain Visual Analogue Scale** - The pain VAS is a reliable and valid measure of perceived pain severity.[9] The scale used in the present study is anchored at both ends, with 0 "No Pain" at one end and 10 "Unbearable Pain" at the other. Nine ticks are evenly distributed between these anchors. The subject is asked to place a mark on the line to represent the pain they experience most of the time. When graded, the tick closest to the subject's mark is selected and thus a numeric score between 0 and 10 is obtained. Both the VAS and PDI measures were administered during the FCE while the client performed either the 30-minute stand or sit portions of the protocol, which occurs after completion of some assessment tasks but prior to a verbal debriefing in which the clinician discusses with the client their overall performance on the FCE.

*4.2.4 Analysis*

A multitrait Pearson correlation matrix was created to determine the pattern of relationships between the two indicators of FCE performance, the PDI, and the VAS. Pearson correlation was used as these variables had continuous level outcomes with approximately normal distributions. The multitrait matrix was modified from the multitrait-multimethod matrix, which has been considered the gold standard for examining the construct validity of a measure.[23] The multitrait matrix allows hypothesized patterns of relationships between variables to be observed. The correlation between individual PDI item scores and other variables were also analyzed for subjects with no missing PDI data, as the item related to occupational disability may be more closely associated with FCE than overall PDI score. An alpha level of 0.05 was chosen to judge significance. All calculations were performed using the computer application SPSS (Chicago, IL).

**4.3 Results**

Of the 381 subjects meeting inclusion criteria, 321 (84%) had complete information on all measures except the PDI and were included in subsequent analyses. Forty-one of the 321 subjects (13%) had missing PDI item scores (39 missing one item, 2 missing two items) and percentage PDI scores were used to maximize available data. Subjects missing PDI items were more likely to be female, employed and have lower levels of perceived disability. The item inquiring about sexual behaviour was the most frequently omitted, with 32 subjects not answering, followed by the recreation question with 4 missing scores.

The majority of subjects were men with a non-specific diagnosis, primarily sprain or strain, who were evaluated more than nine months following their injury (Table I). Mean, median and standard deviation scores for the FCE, PDI, and VAS measures indicate these subjects had moderate levels of disability and pain (Table II).

Moderate correlations were observed between the two FCE performance measures, and between FCE and the PDI and VAS (Table III). Greater weight

lifted on FCE was associated with fewer failed FCE tasks. Higher performance levels on both FCE indicators were associated with lower levels of perceived disability. Lower pain severity ratings on the VAS correlated moderately with higher FCE performance levels, and strongly with lower ratings of perceived disability on the PDI. Total PDI score and percentage PDI score were similarly associated with the other variables. The PDI item having the highest correlation with both FCE indicators was that concerning occupational disability in those subjects with no missing PDI item responses (Table IV).

**4.4 Discussion**

Both IWS FCE performance measures studied correlated moderately with the self-report measure of perceived disability, with better FCE performance related to lower levels of perceived disability as hypothesized. This was seen, in particular, for the PDI item specifically related to occupational disability. These findings indicate that the FCE and the PDI measure similar or related constructs. As the PDI is considered a valid measure of functional ability following musculoskeletal injury, this finding supports the use of the IWS FCE as an indicator of function. Yet the moderate correlations also suggest that unique information is provided from the physical performance and self-report measures, which may warrant the use of both in claimants with sub-acute and chronic low back pain undergoing evaluations of functional capacity and disability.

The correlation coefficients between FCE performance and perceived disability were similar to those seen in previous studies incorporating self-report measures. Hart reported a Pearson correlation of 0.46 between the Oswestry Disability Questionnaire administered prior to patient examination and amount of weight lifted on FCE.[11] Reneman et al found an inverse Spearman rank correlation of 0.52 when categorized manual handling scores from the IWS FCE were compared to the Oswestry Disability Questionnaire administered prior to the FCE.[13] As the PDI is highly correlated with the Oswestry Disability Questionnaire, our finding of a moderate correlation ($r = -0.51$) between weight lifted on IWS FCE and the

PDI was anticipated. We also included a previously unstudied indicator of FCE outcome, the number of failed FCE tasks. This variable considers the subject's performance on FCE in comparison to their physical demands of work, as FCE is interpreted clinically. Both FCE measures correlated similarly with the PDI, but correlated only moderately with each other indicating both variables address unique components of clinically administered FCE outcomes. The moderate correlation of FCE measures could be anticipated, since one measure reflects maximum physical performance while the other represents maximum physical performance related to the individuals' specific job demands, which can vary substantially.

While the relationship between FCE and VAS was not as strong as the PDI and VAS, both indicators of physical capacity on the FCE correlated moderately with the measure of pain intensity. This indicates maximum kinesiophysical FCE performance is not independent of pain ratings as purported (shared variance ($r^2$) of 10-20%).[5] Such independence is not realistic for volitional tests of physical capacity in cases of pain-mediated musculoskeletal disability. Multiple factors including physiological functioning, pain intensity, perceptions of ability, and other personal and contextual variables influence functional ability.[24]

The study findings are limited to workers' compensation claimants with diagnoses related to the low back, undergoing testing with the IWS FCE protocol for purposes of claims decision-making. Claimants undergoing FCE testing within the Alberta workers' compensation system represent a group of individuals for whom FCE is most commonly requested, those who have met or surpassed expected musculoskeletal injury healing times and are no longer progressing with treatment, yet report ongoing difficulties related to their compensable condition. The magnitude of the associations observed may not apply to all claimants on total temporary disability or those with other musculoskeletal injuries.

Another limitation of our findings relates to one indicator of performance on the IWS FCE, the number of failed tasks. This indicator correlated similarly with pain intensity and perceived disability as did average weight lifted on the FCE and is central to FCE determinations and associated decision-making, but has not been studied for reliability. Such reliability, in great part, rests on the reliability and validity of the methods used for determining physical job demands for comparison with physical performance on FCE. Finally, while the study results support the construct validity of the IWS FCE administered clinically, further knowledge of other forms of validity, such as predictive validity for safe return to work, are needed to support the use of FCE.

### 4.5 Conclusion

The pattern of relationships observed between the IWS FCE and PDI supports the construct validity of both indicators of FCE performance as measures of function in individuals with work-related low back pain. However, kinesiophysical FCE performance was not unrelated to ratings of pain intensity as purported.

## TABLE 4-1

| SUBJECT CHARACTERISTICS (n=321) | |
|---|---|
| **Variable** | **Percentage or Mean (SD)** |
| Male | 72% |
| Employed | 63% |
| Age (year) | 42 (9.9) |
| Days From Injury to FCE | 737 (1361), Median 307 |
| Diagnosis | |
| Non-specific (sprain/ strain/ pain) | 83% |
| Disc Pathology | 12% |
| Fracture/ Dislocation | 5% |
| National Occupational Classification | |
| Limited (Less than 5kg) | 7% |
| Light (5-9kg) | 30% |
| Medium (10-20kg) | 38% |
| Heavy (>20kg) | 25% |

## TABLE 4-2

**DISTRIBUTION OF VARIABLES WITHIN THE WORK ASSESSMENT BATTERY**

| Variable | Mean (SD) |
|---|---|
| Average Weight Lifted During Functional Evaluation (kg) | 15 (8.6) |
| Number of Failed Tasks on Functional Capacity Evaluation | 9 (5.3) |
| Pain Disability Index (out of 70, n=280) | 40 (15.0) |
| % Pain Disability Index | 56 (21.6) |
| Pain Visual Analogue Scale | 5 (2.1) |

**TABLE 4-3**

| MULTITRAIT PEARSON CORRELATION MATRIX | | | | |
|---|---|---|---|---|
| | FCE(2) | PDI | %PDI | Pain VAS |
| Average Weight Lifted (FCE1) | -0.53 | -0.52 | -0.51 | -0.45 |
| Number of Failed Tasks (FCE2) | | 0.44 | 0.45 | 0.34 |
| Pain Disability Index (PDI) | | | | 0.77 |
| %PDI | | | | 0.79 |
| All correlations significant at 0.05 level. | | | | |

## TABLE 4-4

**RELATIONSHIPS BETWEEN FUNCTIONAL CAPACITY EVALUATION, PAIN VISUAL ANALOGUE SCALE AND PAIN DISABILITY INDEX ITEMS**

**(All values represent Pearson coefficients, n=280)**

| | Pain Disability Index Items | | | | | | |
|---|---|---|---|---|---|---|---|
| | Family/ Home | Recreation | Social Activities | Occupation | Sexual Behaviour | Self Care | Life Support |
| FCE (1)[1] | -0.43 | -0.44 | -0.43 | -0.46 | -0.41 | -0.38 | -0.41 |
| FCE (2)[2] | 0.41 | 0.35 | 0.36 | 0.48 | 0.29 | 0.34 | 0.30 |
| Pain VAS[3] | 0.66 | 0.59 | 0.70 | 0.55 | 0.59 | 0.70 | 0.68 |

[1] FCE (1) = Average Weight Lifted

[2] FCE (2) = Number of Failed Tasks

All Correlations significant at 0.05 level

## 4.6 References


1. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther* 1998; 78:852-66.
2. Gibson L, Strong J. A review of functional capacity evaluation practice. *Work* 1997; 9:3-11.
3. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13:125-52.
4. Innes E, Straker L. Reliability of work-related assessments. *Work* 1999; 13:107-24.
5. Isernhagen SJ. Functional capacity evaluation: rationale, procedure, utility of the kinesiophysical approach. *J Occup Rehab* 1992; 2:157-168.
6. Isernhagen SJ. *The comprehensive guide to work injury management.* Gaithersburg, Md.: Aspen Publishers, 1995:xxxii, 821.
7. Gross DP, Battié MC. Reliability of safe maximum lifting determinations of a functional capacity evaluation. *Phys Ther* 2002; 82:364-71.
8. Reneman MF, Dijkstra PU, Westmaas M, Goeken LN. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehabil* 2002; 12:269-75.
9. Finch E, Brooks D, Stratford P, Mayo N. *Physical Rehabilitation Outcome Measures: A Guide to Enhanced Clinical Decision Making.* Toronto: Canadian Physiotherapy Association, 2002.
10. Domholdt E. *Physical therapy research : principles and applications.* Philadelphia, Pa.: W.B. Saunders, 2000:xiv, 522.
11. Hart DL. Relation between three measures of function in patients with chronic work-related pain syndromes. *J Rehab Outcome Measure* 1998; 2:1-14.
12. United States. *Employment and Training Administration. Dictionary of occupational titles.* Indianapolis: JIST Works, 1991:2 v. in 1 (1404 ).
13. Reneman MF, Jorritsma W, Schellekens JM, Goeken LN. Concurrent validity of questionnaire and performance-based disability measurements

in patients with chronic nonspecific low back pain. *J Occup Rehabil* 2002; 12:119-29.

14. Hansson GA, Balogh I, Bystrom JU, et al. Questionnaire versus direct technical measurements in assessing postures and movements of the head, upper back, arms and hands. *Scand J Work Environ Health* 2001; 27:30-40.
15. Chibnall JT, Tait RC. The Pain Disability Index: factor structure and normative data. *Arch Phys Med Rehabil* 1994; 75:1082-6.
16. Gronblad M, Jarvinen E, Airaksinen O, Ruuskanen M, Hamalainen H, Kouri JP. Relationship of subjective disability with pain intensity, pain duration, pain location, and work-related factors in nonoperated patients with chronic low back pain. *Clin J Pain* 1996; 12:194-200.
17. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: psychometric properties. *Pain* 1990; 40:171-82.
18. Gronblad M, Hupli M, Wennerstrand P, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain* 1993; 9:189-95.
19. Gronblad M, Jarvinen E, Hurri H, Hupli M, Karaharju EO. Relationship of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) with three dynamic physical tests in a group of patients with chronic low-back and leg pain. *Clin J Pain* 1994; 10:197-203.
20. Pollard CA. Preliminary validity study of the pain disability index. *Percept Mot Skills* 1984; 59:974.
21. Gronblad M, Hurri H, Kouri JP. Relationships between spinal mobility, physical performance tests, pain intensity and disability assessments in chronic low back pain patients. *Scand J Rehabil Med* 1997; 29:17-24.
22. Kang SW, Lee WN, Moon JH, Chun SI. Correlation of spinal mobility with the severity of chronic lower back pain. *Yonsei Med J* 1995; 36:37-44.

23. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000; 81:S15-20.

24. Kaplan GM, Wurtele SK, Gillis D. Maximal effort during functional capacity evaluations: an examination of psychological factors. *Arch Phys Med Rehabil* 1996; 77:161-4.

# CHAPTER 5

## The Prognostic Value of Functional Capacity Evaluation in Patients with Chronic Low Back Pain: Part 1 Timely Return

*A version of this chapter has been accepted for publication. Gross, Battié, and Cassidy, Spine (accepted May 2003).*

### 5.1 Introduction

Return to work has been considered a key outcome for judging resolution of work-related low back disability.[1] Predicting this outcome has proven especially complex and multifaceted with psychological or socioeconomic factors, such as employment status or perceptions of health and disability, most closely related to return to work.[2-4] However, judgments of fitness to return to work and determinations of permanent disability are most frequently made using information from physical examination or diagnostic imaging, despite the fact such observations are consistently reported as being poorly associated with future recovery.[5-7]

The use of Functional Capacity Evaluations (FCE) as an alternate or adjunct method of making judgments of performance potential and readiness for work following musculoskeletal injury is becoming increasingly common in occupational health care clinics worldwide.[8, 9] FCEs are standardized batteries of clinical tests purported to measure a subject's safe physical ability for work-related activity.[10] Performance on tasks in the FCE protocol is compared to required physical job demands of the subject's occupation and a decision regarding ability to return to work is made. If performance meets or exceeds all job requirements, the injured individual is deemed ready to return.[11] Performance that does not match all job demands may be used to guide rehabilitative interventions or modified work programs. If the worker does not have a job to return to, results are used to guide vocational rehabilitation and job placement.

While FCE is a clinical tool used to guide return to work recommendations made by health care professionals, administrative decisions are also made, in part, based on their results. Determinations are made regarding ability to safely perform various work activities and consequences of benefit suspension and claim closure exist in some jurisdictions for subjects declared ready for work. Decisions based on FCE, therefore, have important implications related to employability and personal livelihood for the injured individuals undergoing such testing. As well, the process of FCE testing can cost as much as advanced diagnostic procedures, as typical protocols take hours to complete. FCE consequently has important financial implications for individuals undergoing testing, compensation systems and society. The ability of FCE to accurately determine future work ability is of great importance to the integrity of the process and predictive validity has been considered a key psychometric property.[12]

The predictive ability of some commonly performed FCE tasks has been examined recently by Matheson et al.[13] Higher levels of weight lifted during a floor-to-waist lift task were associated with higher likelihood of return to work six-months following the FCE in clients with chronic musculoskeletal disorders. As well, Fishbain et al have reported that performance on various FCE activities including acceptable weight lifted were related to future return to work.[14] However, both of these studies indicated other personal and contextual factors, such as gender, pain intensity, compensation status and duration of disability, were more closely related to recovery. As well, the amount of weight lifted is a somewhat artificial indicator of FCE performance as typically the relationship between performance and job demands forms the basis of return to work recommendations. Subjects lifting less on FCE may also have lower physical job demands and be considered as safe to return as subjects lifting to higher levels. The comparison between FCE and job demands, although standardly used for FCE interpretation and decision-making, has not been studied for its association with return to work.

We investigated the predictive validity, or prognostic value, of the Isernhagen Work Systems' (Duluth, MN) FCE in determining timely recovery among workers' compensation claimants with low back pain conditions. We hypothesized better performance on FCE would be associated with faster time to return to work, measured using the proxy indicator of suspension of total temporary disability (TTD) benefits, and faster time to claim closure.

## 5.2 Materials and Methods

### *5.2.1 Design*

Determining the prognostic value of a health indicator is a staged process, which Altman and Lyman have described using the concepts of exploratory and confirmatory studies.[15] Exploratory studies aim to evaluate the association between an individual prognostic indicator and the outcome of interest alone and after adjusting for the confounding effect of other important predictors. Confirmatory studies seek to verify previous exploratory findings as well as establish more accurately the magnitude of predictability added by the indicator of interest. The present study reports on exploratory and confirmatory investigations undertaken relative to the prognostic value of FCE.

A historical cohort design was used. Data on variables of interest were extracted from clinical and administrative databases of the Alberta Workers' Compensation Board (WCB-Alberta) and were merged using a common unique identifier. The study plan was approved by the University of Alberta's Health Research Ethics Board.

### *5.2.2 Subjects*

Clients undergoing FCE at the major WCB-Alberta rehabilitation facility between January 1, 1999 and December 31, 2000, who had open claims for work-related low back injury formed our sample. Separate cohorts were formed for subjects seen in 1999 and 2000. Specific inclusion criteria were chosen to ensure medical stability and included: compensated work-related injuries with ICD-9 codes

indicative of injury to the lower back; FCE conducted at least six weeks after the date of accident; and medical clearance to participate in FCE with no further medical investigation or treatment recommended. Claimants undergoing FCE within the Alberta workers' compensation environment are usually those who have met or surpassed expected injury healing times and have plateaued with medical and rehabilitative interventions, yet report ongoing difficulties related to their compensable condition. Such claimants are evaluated with FCE for purposes of determining fitness to return to work and administrative decisions are made, in part, based on results.

*5.2.3 Measures*

**Functional Capacity Evaluation** - The Isernhagen Work Systems' FCE protocol is used throughout the WCB-Alberta health care provider network. During administration of this FCE, the clinician relies on observation of biomechanical and physiological signs of effort to determine safe, maximum performance levels.[16] The protocol includes activities representing the physical demands of work outlined in the Dictionary of Occupational Titles, such as lifting, carrying, pushing, pulling, and other tasks.[17] Acceptable inter-rater and test-retest reliability have been reported for determinations of maximum performance on the lifting and carrying tasks.[18, 19] Representatives of Isernhagen Work Systems trained all clinicians performing FCE at the centre from which data were obtained.

At the rehabilitation facility, claimant performance on each of 25 tasks in the FCE is compared to related physical job demands and is given either a pass or fail rating. Determining physical job demands at the rehabilitation facility is accomplished through employer report and direct work site measurement or, when these are unavailable, worker self-report with a standardized questionnaire. Formal work-site evaluation had been previously performed for 17% of the subjects. Our plan was to create a dichotomous variable of whether or not any FCE tasks were failed for use as a prognostic indicator, as the decision that a

claimant is safe to return to work is made only when all tasks have been successfully passed.[10, 16] In addition, to replicate the study by Matheson et al, the maximum weight lifted during the floor-to-waist lift task along with the rating of whether this task was rated as failed were considered independently as predictors.

**Potential Confounding Variables** - Other variables were extracted for analysis based on their potential for a confounding influence on the relationship between FCE and our outcomes. We selected variables from the administrative databases that had been reported as predictive of recovery in previously published studies or where a sound theoretical rationale for considering the variable existed. The variables considered included gender; age; diagnosis; employment status; days from injury to FCE; scores from the Pain Disability Index (a reliable and valid measure of perceived disability[20, 21]) and pain Visual Analogue Scale completed by subjects at time of FCE, clinician recommendation following FCE administration; job physical demands rating from Human Resources Development Canada's National Occupational Classification[22]; pre-injury annual salary; number of health care visits preceding the FCE for the compensable back condition; and number of previous WCB back claims.

**Outcomes** - As FCE is thought to determine ability to return to work, our goal was to examine indicators of timely return to work following FCE. Among subjects receiving TTD benefits, time in days from FCE to suspension of TTD in the year following FCE administration served as the primary proxy measure of return to work and functional recovery. One year was judged to be a suitable follow-up time as events occurring after this period are unlikely to be related to FCE testing. Claimants receive TTD benefits when an entire day of work loss is being subsidized and suspension of benefits typically indicates return. However, some claimants undergoing FCE are not receiving TTD benefits and may be working modified hours while receiving temporary partial disability benefits, working full-time yet reporting ongoing difficulties, or receiving reemployment assistance payments or more permanent economic loss subsidies for the open

claim. We therefore also analyzed the value of FCE in predicting subsequent time to claim closure in the year following FCE as a secondary indicator of functional recovery.

*5.2.4 Analysis*

Descriptive statistics were calculated including Kaplan-Meier curves for time receiving TTD benefits and until claim closure. Ongoing benefit reception and open claims were censored at one year. Significant differences between subject characteristics in the separate cohorts were determined using Chi square, log rank and t tests. A 0.05 alpha level was chosen to judge significance.

Exploratory analysis of the prognostic value of FCE was performed using data from the cohort of subjects undergoing FCE in 1999. Initially, crude relationships between the FCE performance indicators and time to TTD suspension and claim closure were determined using Cox proportional hazards regression.[23] Then, the crude effect of each FCE indicator was adjusted using a risk factor modeling strategy.[24] Each FCE variable was added separately to a multivariable Cox regression model along with other potential confounding variables found crudely related to outcomes at a 0.20 alpha level or which altered a FCE regression coefficient by 20% or greater.[24] This modeling approach allowed for control of variables most likely to have a confounding influence while avoiding the addition of too many independent variables to the multivariable regression. The partial and marginal amount of variation explained by the FCE variables were also calculated using the technique described by Schemper.[25,26] The proportional hazards and linearity assumptions were evaluated.

To confirm the results of the exploratory analyses, all analyses described above were repeated on a separate cohort of subjects seen for FCE in 2000. This included fitting the same multivariable risk factor models to the new cohort, allowing determination of the stability of the estimated hazard rate ratios (HRR) and estimated proportions of variation explained by FCE.[27]

The recommended sample size of 10 events per independent variable was adhered to for all regressions.[28] All calculations were performed using the computer application SPSS (Chicago, IL).

## 5.3 Results

### *5.3.1 Exploratory Analysis*

Of the 150 subjects seen in 1999 and meeting inclusion criteria, 114 (76%) had complete data and were included in subsequent analyses. Subjects with missing data had a significantly longer time between injury and FCE (553 vs. 1293 days), but were similar with respect to the other eleven factors examined (Table 1). The majority of subjects were men with non-specific diagnoses (primarily sprain or strain), evaluated more than six months following their injury.

Only 5 subjects (4%) seen in 1999 passed all FCE tasks. It was not possible, therefore, to dichotomize FCE as whether or not job demands were met or exceeded on all 25 tasks and this variable was modeled as a continuous predictor.

**Time to TTD Suspension** - At the time of FCE, 77 subjects (68%) were receiving TTD benefits. The only significant difference on available data between subjects receiving benefits and those not was a higher likelihood of a fit to work recommendation by the administering clinician following FCE testing in those not receiving TTD benefits. The median time receiving TTD benefits following FCE was 32 days and 4 subjects (5%) received benefits for the entire follow-up year.

A higher number of failed FCE tasks was crudely related to delayed time until TTD suspension (Table II). After controlling for potential confounders, subjects were approximately 9% less likely to experience TTD suspension at any time over the follow-up year for each additional task rated as failed (HRR 0.91; 95% confidence interval (CI) 0.86-0.96). The partial, or adjusted, explanatory effect of this indicator was 14.8% of the variation in time to TTD suspension. As well,

those subjects lifting to higher levels on the floor-to-waist lift (adjusted HRR 1.48; 95% CI 1.14-1.92) and having this task rated as passed (adjusted HRR 2.83; 95% CI 1.49-5.35, shown graphically in Figure I) experienced TTD suspension sooner. Both weight lifted and the pass/fail rating on the floor-to-waist lift explained comparable amounts of variation in time to TTD suspension as the number of failed tasks on the entire protocol (10.5% and 11.3% respectively).

**Time to Claim Closure** - The median time to claim closure following FCE for the 1999 cohort was 97 days and 25 claims (22%) remained open for the entire follow-up year. Consistent with time to TTD suspension, a higher number of failed FCE tasks, a higher amount of weight on the floor-to-waist lift, and a passed rating on the floor-to-waist lift were crudely associated with faster claim closure (Table III). Hazard rate ratios changed minimally after controlling for potential confounders and the FCE variables explained small amounts of variation in time to claim closure (<8%).

*5.3.2 Confirmatory Analysis*

Descriptive information was statistically compared between the 1999 and 2000 cohorts and no significant differences were observed with the exception of an increase of 1.5 extra FCE tasks failed on average in the 2000 cohort (Table I). Only 6 of 132 (5%) subjects passed all FCE tasks in 2000.

Associations between all FCE performance indicators and outcomes were similar in direction and magnitude between the 1999 and 2000 cohorts (Table II and III). The pass/fail rating on the floor-to-waist lift appeared slightly more predictive in the 2000 cohort, however, the proportions of variation explained by the other FCE variables were consistent in the confirmatory cohort. Performance on the floor-to-waist lift explained as much variation as the pass/fail ratings for the entire FCE.

## 5.4 Discussion

A recommendation to return to work based on FCE results has been advocated only in cases where a subject has demonstrated physical performance at levels matching or exceeding all required job demands. Yet, only 4% of subjects (11 of 246 in both cohorts) deemed medically stable and referred for FCE were rated as meeting all job demands. A subsequent decision to block return to work in the remaining 96% of subjects would have hindered return to work, as indicated through suspension of TTD benefits, in the majority of subjects evaluated. Nearly all subjects (95%) had TTD suspended during the follow-up year.

As anticipated, indicators of better performance on the Isernhagen Work Systems' FCE administered within a workers' compensation system to assist in claims decision-making were related to faster time to TTD suspension and claim closure. Claimants were approximately 9% less likely to experience TTD suspension at any time in the follow-up year for each FCE task rated as failed. Subjects were also approximately 50% more likely to experience TTD suspension for every extra 10kg lifted on the floor-to-waist lift and three times more likely if this task was rated as passed.

The prognostic ability of FCE was anticipated to be high within an administrative system in which decisions are made, in part, based on FCE results. However, the magnitude of the association between FCE and outcomes, determined by hazard rate ratios and proportion of variation in the outcomes explained, was quite low. Within the confirmation analyses, which likely represent the most accurate estimates of the adjusted relationship between FCE and outcomes given that risk factor models were being tested on this cohort, the FCE indicators added little more than 10% to prognostic accuracy after controlling for potential confounders. Given that return to work recommendations are made for individual claimants based, in part, on results of FCE testing, explaining an additional 10% of the variation in TTD suspension or claim closure appears trivial. This finding, combined with the findings of Matheson et al and Fishbain et al who have

reported that other personal and contextual factors were more closely related to recovery than FCE performance, makes it questionable whether the high administrative burden associated with FCE testing is worth the small increase in predictability obtained.[13, 14]

If costs prohibit testing with an entire FCE protocol, a reasonable alternative for the purpose of making predictions of future recovery may be testing only performance on the floor-to-waist lift. Matheson et al reported this indicator was associated with six month return to work, and in the present study it explained as much variation in outcomes as information from the entire FCE protocol regarding the matching relationship between performance and job demands. This simple test could easily be incorporated into a comprehensive physical examination for chronic low back pain patients and may warrant further study.

While the addition of a confirmatory analysis on a separate cohort provides confidence in our results, limitations include the reliance on administrative databases. While claim and benefit status are important personal and societal outcomes, other health related outcomes including self-rated pain and disability would provide a richer description of the claimant recovery experience. In addition, information on all potentially confounding factors was not available in the databases. Such variables include claimant expectations of recovery and perceptions of workplace support, among others. Another potential source of bias is measurement error, as the measurement properties of the methods used for determining job demands for comparison to FCE performance have not been studied.

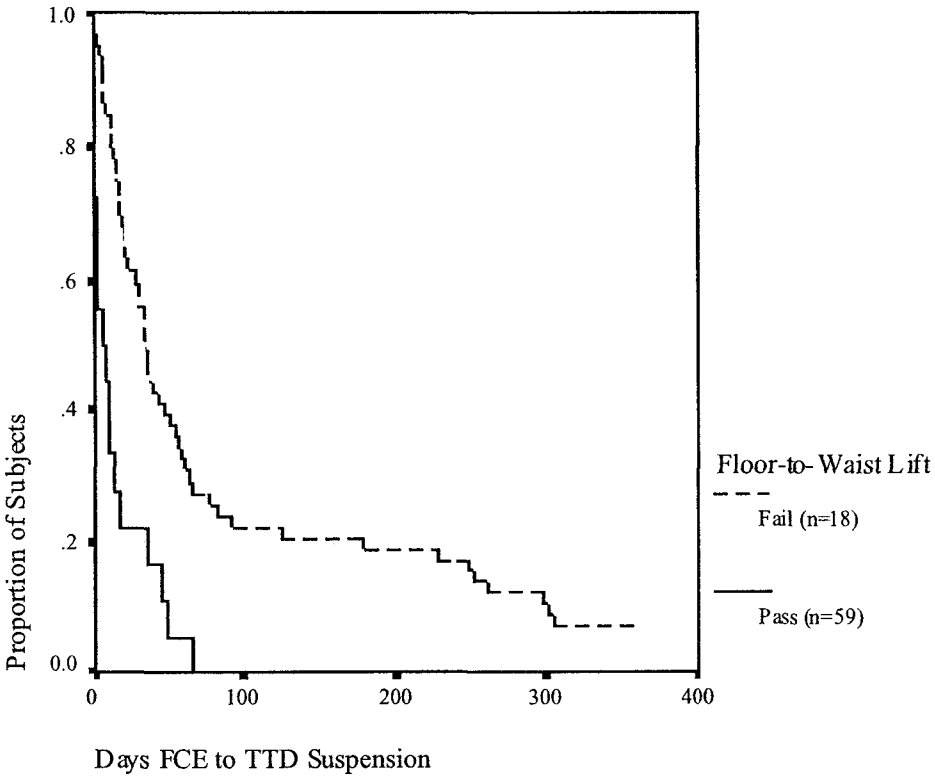
### 5.5 Conclusion

Indicators of better performance on the Isernhagen Work Systems' FCE administered within a workers' compensation system to assist in claims decision-making were related to faster time to TTD suspension and claim closure. The number of failed FCE tasks was weakly associated with both outcomes and the

amount of variation explained was low. Performance on the floor-to-waist lift appears to predict as well as information from the entire FCE protocol.

## FIGURE 5-1

**Kaplan-Meier Estimates of the Time to Suspension of Total Temporary Disability Benefits (1999 Cohort, n=77)**



## Table 5-1

| SUBJECT CHARACTERISTICS | | |
|---|---|---|
| | **1999 Cohort** (n=114) | **2000 Cohort** (n=132) |
| **Variable** | **Mean (Std dev) or Percentage** | |
| Age (yr) | 41 (9.5) | 40 (9.4) |
| Days From Injury to Functional Capacity Evaluation (FCE) | 553* (1522) | 599** (813) |
| Pain Disability Index (%) | 55 (20) | 54 (22) |
| Pain Visual Analogue Scale | 5 (2.0) | 5 (2.2) |
| Failed FCE Tasks (out of 25) | 7.9 (7.0) | 9.4 (4.9) |
| FCE Maximum Floor-to-Waist Lift (kg) | 14.7 (10.6) | 12.2 (10.1) |
| Annual Salary (thousands of dollars) | 35.6 (21.0) | 31.2 (13.7) |
| Number of Pre-FCE Health Care Visits | 35 (33.4) | 34 (28.0) |
| Previous WCB Back Claims | 1 (1.8) | 1 (2.2) |
| Male (%) | 74 | 71 |
| Employed (%) | 66 | 71 |
| Diagnosis (%) | | |
| Non-specific (Sprain/ strain) | 81 | 80 |
| Disc Pathology | 12 | 16 |
| Fracture/ Dislocation | 7 | 5 |
| National Occupational Classification (%) | | |
| Limited (<5kg) | 4 | 8 |
| Light (5-10kg) | 34 | 29 |
| Medium (10-20kg) | 35 | 39 |
| Heavy (>20kg) | 26 | 24 |
| Receiving TTD Benefits (%) | 68 | 65 |

* Median 212

** Median 244

## TABLE 5-2

**The Relationship Between Indicators of Performance on Functional Capacity Evaluation and Time to Total Temporary Disability Benefit Suspension**

| | **1999 Cohort** (n=77, 73 events) | | **2000 Cohort** (n=86, 84 events) | |
|---|---|---|---|---|
| *Variable* | *HRR* (95% CI)* | *PVE*** | *HRR (95% CI)* | *PVE* |
| Number of Failed Tasks | | | | |
| Crude | 0.89 (0.85-0.94) | 24.5% | 0.91 (0.87-0.96) | 16.3% |
| Adjusted‡ | 0.91 (0.86-0.96) | 14.8% | 0.92 (0.87-0.97) | 10.9% |
| Floor-to-Waist Lift (10 kg units) | | | | |
| Crude | 1.49 (1.19-1.87) | 14.1% | 1.43 (1.14-1.79) | 9.9% |
| Adjusted | 1.48 (1.14-1.92) | 10.5% | 1.43 (1.09-1.89) | 7.3% |
| Floor-to-Waist Lift (Pass/Fail Rating) | | | | |
| Crude | 3.22 (1.83-5.67) | 16.5% | 4.19 (2.12-8.29) | 14.0% |
| Adjusted | 2.83 (1.49-5.35) | 11.3% | 3.74 (1.81-7.71) | 11.4% |

*HRR – Hazard Rate Ratio

**PVE – Proportion of Variation Explained or R square for Cox regression (partial and marginal effects shown)

‡ The effect of each FCE indicator was adjusted for the Pain Disability Index, clinician recommendation following FCE, age, pre-injury annual salary, and number of health visits preceding the FCE.

## TABLE 5-3

**The Relationship Between Indicators of Performance on Functional Capacity Evaluation and Time to Claim Closure**

| | **1999 Cohort** (n=114, 89 events) | | **2000 Cohort** (n=132, 107 events) | |
|---|---|---|---|---|
| *Variable* | *HRR* (95% CI)* | *PVE*** | *HRR (95% CI)* | *PVE* |
| Number of Failed Tasks | | | | |
| Crude | 0.92 (0.88-0.96) | 12.5% | 0.91 (0.86-0.95) | 12.6% |
| Adjusted‡ | 0.92 (0.88-0.98) | 7.2% | 0.92 (0.87-0.97) | 7.2% |
| Floor-to-Waist Lift (10 kg units) | | | | |
| Crude | 1.31 (1.08-1.59) | 6.2% | 1.37 (1.12-1.66) | 6.9% |
| Adjusted | 1.17 (0.91-1.50) | 1.3% | 1.29 (1.02-1.64) | 3.4% |
| Floor-to-Waist Lift (Pass/Fail Rating) | | | | |
| Crude | 2.26 (1.41-3.63) | 8.4% | 2.04 (1.16-3.59) | 3.8% |
| Adjusted | 2.18 (1.26-3.77) | 6.3% | 4.01 (2.10-7.64) | 10.3% |

*HRR – Hazard Rate Ratio

**PVE – Proportion of Variation Explained or R square for Cox regression (partial and marginal effects shown)

‡ The effect of each FCE indicator was adjusted for the Pain Disability Index, gender, salary, employment status, age, number of health visits preceding FCE, NOC strength category, and time between injury and FCE

## 5.6 References


1. Amick BC, 3rd, Lerner D, Rogers WH, Rooney T, Katz JN. A review of health-related work outcome measures and their uses, and recommended measures. *Spine* 2000; 25:3152-60.
2. Crook J, Milner R, Schultz IZ, Stringer B. Determinants of occupational disability following a low back injury: a critical review of the literature. *J Occup Rehabil* 2002; 12:277-95.
3. Truchon M, Fillion L. Biopsychosocial determinants of chronic disability and low-back pain: a review. *J of Occup Rehab* 2000; 10:117-42.
4. van der Giezen AM, Bouter LM, Nijhuis FJ. Prediction of return-to-work of low back pain patients sicklisted for 3-4 months. *Pain* 2000; 87:285-94.
5. McIntosh G, Frank J, Hogg-Johnson S, Bombardier C, Hall H. Prognostic factors for time receiving workers' compensation benefits in a cohort of patients with low back pain. *Spine* 2000; 25:147-57.
6. Cocchiarella L, Andersson GBJ, *American Medical Association. Guides to the evaluation of permanent impairment.* Chicago, Ill.: American Medical Association, 2001:xxii, 613.
7. Hunt DG, Zuberbier OA, Kozlowski AJ, et al. Are components of a comprehensive medical assessment predictive of work disability after an episode of occupational low back trouble? *Spine* 2002; 27:2715-9.
8. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther* 1998; 78:852-66.
9. Wyman DO. Evaluating patients for return to work. *Am Fam Phys* 1999; 59:844-8.
10. Innes E, Straker L. A clinician's guide to work-related assessments: 1-purposes and problems. *Work* 1998; 11:183-189.
11. Isernhagen SJ. *The comprehensive guide to work injury management.* Gaithersburg, Md.: Aspen Publishers, 1995:xxxii, 821.
12. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13:125-52.

13. Matheson LN, Isernhagen SJ, Hart DL. Relationships among lifting ability, grip force, and return to work. *Phys Ther* 2002; 82:249-56.

14. Fishbain DA, Cutler RB, Rosomoff H, Khalil T, Abdel-Moty E, Steele-Rosomoff R. Validity of the dictionary of occupational titles residual functional capacity battery. *Clin J Pain* 1999; 15:102-10.

15. Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Res Treat* 1998; 52:289-303.

16. Isernhagen SJ. Functional capacity evaluation: rationale, procedure, utility of the kinesiophysical approach. *J of Occup Rehab* 1992; 2:157-168.

17. United States. Employment and Training Administration. *Dictionary of occupational titles*. Indianapolis: JIST Works, 1991:2 v. in 1 (1404 ).

18. Gross DP, Battié MC. Reliability of safe maximum lifting determinations of a functional capacity evaluation. *Phys Ther* 2002; 82:364-71.

19. Reneman MF, Dijkstra PU, Westmaas M, Goeken LN. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehabil* 2002; 12:269-75.

20. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: psychometric properties. *Pain* 1990; 40:171-82.

21. Gronblad M, Hupli M, Wennerstrand P, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain* 1993; 9:189-95.

22. Canada. Human Resources Development Canada. *National occupational classification, 2001 : Occupational descriptions.* [Ottawa]: Human Resources Development Canada, 2001:1 v. (loose-leaf).

23. Kleinbaum DG. *Survival analysis : a self-learning text. Statistics in the health sciences*. New York: Springer, 1996:xii, 324.

24. Hosmer DW, Lemeshow S. *Applied survival analysis : regression modeling of time to event data*. Wiley series in probability and statistics. Texts and reference section. New York: Wiley, 1999:xiii, 386.

25. Schemper M. The relative importance of prognostic factors in studies of survival. *Stat Med* 1993; 12:2377-82.

26. Schemper M, Stare J. Explained variation in survival analysis. *Stat Med* 1996; 15:1999-2012.

27. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15:361-87.

28. Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep* 1985; 69:1071-77.

# CHAPTER 6

## The Prognostic Value of Functional Capacity Evaluation in Patients with Chronic Low Back Pain: Part 2 Sustained Recovery

*A version of this chapter has been accepted for publication. Gross and Battié, Spine (accepted May 2003).*

### 6.1 Introduction

Low back pain is a recurring phenomenon.[1] While the anticipated duration of most acute episodes is relatively short, a moderate probability exists of recurrence in the year after recovery.[2] A recent review has reported the likelihood of pain recurrence varies between 35 and 82%, and the likelihood of recurrence of compensated disability varies between 18 and 44% when all future compensated work absences are considered.[3] Predicting which individuals will experience recurrent back pain has proven difficult. The most promising prognostic indicators include a higher number of previous episodes and missed workdays, along with higher pain severity and lower self-rated functional ability during a current episode.[4-6]

Functional Capacity Evaluations (FCE) are standardized batteries of clinical tests which are commonly used to determine readiness or ability for safe return to work following musculoskeletal injury, implying low risk of future recurrence.[7] This determination is made by comparing demonstrated performance on FCE to the required physical job demands of the worker's occupation.[8] If an injured worker's performance on FCE meets or exceeds all physical job requirements, the individual is considered 'safe' to return to work. Immediate return is recommended and such return is anticipated to be uncomplicated by future injury occurrence as the worker is not at risk of exposure to intolerable physical loading.[9,10] FCEs are, therefore, used as prognostic tools for determining future functional tolerances and potential for injury recurrence.

FCEs are used widely in workers' compensation and other insurance systems to assist in making return to work and case management decisions. Determinations made with FCE have important implications related to employability and personal livelihood for the injured individuals undergoing testing, as well as for the employers accepting the worker back to their workplace. FCE is expected to be a valid tool to accurately identify individuals ready to return to work, who have low risk of future injury or recurrence of their back pain problem.[11]

In part one of this series, we examined the predictive ability of the Isernhagen Work Systems' (Duluth, MN) FCE in determining timely return to work. Better performance on the FCE indicated by a lower number of failed tasks, weight lifted during the floor-to-waist lifting task and whether this task was rated as failed were weakly related to shorter time to administrative recovery in workers' compensation claimants with chronic low back pain. This is consistent with the findings of Matheson et al and Fishbain et al who have reported that various individual FCE tasks were associated with future return to work, but less predictive than other personal and contextual factors such as gender, pain ratings, compensation status, and duration of disability.[12,13] No studies of the ability of FCE to predict successful and sustained recovery have yet been reported. In the present study, we investigated the ability of the Isernhagen Work Systems' FCE to predict sustained recovery, or the absence of future work-related recurrent events, among workers' compensation claimants with chronic low back pain.

## 6.2 Materials and Methods

### *6.2.1 Study Design*

We conducted an exploratory analysis, as described by Altman and Lyman, of the prognostic ability of FCE in predicting sustained recovery.[14] A historical cohort study design was used. Data on variables of interest were extracted from clinical and administrative databases of the Alberta Workers' Compensation Board (WCB-Alberta) and were merged using a common unique identifier. The University of Alberta Health Research Ethics committee approved this study.

*6.2.2 Subjects*

Our study cohort was composed of workers' compensation claimants with low back injuries undergoing FCE in the years 1999 or 2000 at the major WCB-Alberta rehabilitation facility, who subsequently experienced administrative recovery in the form of total temporary disability (TTD) benefit suspension or claim closure in the year following FCE. Other specific inclusion criteria were chosen to ensure medical stability and included: compensated work-related injuries with ICD-9 codes indicative of injury to the lower back; FCE conducted at least six weeks after the date of accident; medical clearance to participate in FCE; and not discharged for further medical investigation or treatment. Claimants undergoing FCE testing within the Alberta workers' compensation environment are usually those who have met or surpassed expected injury healing times and have plateaued with medical and rehabilitative interventions, yet report ongoing difficulties related to their compensable condition. Such claimants are evaluated with FCE for purposes of determining readiness to return to work.

*6.2.3 Measures*

**Functional Capacity Evaluation** - The Isernhagen Work Systems' FCE protocol is used throughout the WCB-Alberta health care provider network. During administration of this FCE, the clinician relies on observation of biomechanical and physiological signs of effort to determine safe, maximum performance levels.[15] The protocol includes activities representing the physical demands of work outlined in the Dictionary of Occupational Titles, such as lifting, carrying, pushing, pulling, and other tasks.[16] Acceptable inter-rater and test-retest reliability have been reported for determinations of maximum performance on the lifting and carrying tasks.[17-19] Representatives of Isernhagen Work Systems trained all clinicians performing FCE at the centre from which data were obtained.

At the rehabilitation facility, performance on each task in the FCE is compared to related physical job demands and is given either a pass or fail rating. The process

for determining physical job demands at the rehabilitation facility varies depending on individual case characteristics, but is accomplished through employer report and direct work site measurement or, when these are unavailable, worker self-report in the form of a standardized questionnaire. Our plan was to create a dichotomous variable of whether or not any FCE tasks were rated as failed as the decision that a claimant is safe to return to work is made only when all tasks have been successfully passed. In addition, as maximum weight lifted on the floor-to-waist lift task in the protocol and whether or not this task was rated as failed were previously found to be associated with timely recovery[13,20], these variables were analyzed independently as indicators of FCE performance.

**Potential Confounding Variables** - Other variables were extracted for our analysis based on their potential for a confounding influence on FCE. We selected variables from the administrative databases that had been reported as predictive of recurrence in previously published scientific literature or where a sound theoretical rationale for considering the variable existed. The variables considered included: previous back claims; days between FCE and initial TTD suspension and claim closure; gender; age; diagnosis; employment status; days from injury to FCE; scores on the Pain Disability Index (a valid and reliable measure of perceived disability[21, 22]) and pain Visual Analogue Scale completed by subjects at time of FCE; clinician recommendation regarding fitness or readiness to return to work following FCE administration; physical job demand ratings from Human Resources Development Canada's National Occupational Classification[23]; pre-injury annual salary; and number of health care visits preceding the FCE for the compensable condition.

**Outcomes** - FCE is thought to determine ability to safely return to work. Thus, theoretically, a strong FCE performance should be associated with fewer exacerbations or recurrences of work-related low back pain as reflected through sustained return to work, than would be the case for a poor performance on FCE. Our goal was to examine indicators of sustained recovery within the year

following FCE. Sustained recovery was considered to be the avoidance of future recurrent events. A recurrent event was defined for subjects who had experienced TTD benefit suspension as re-starting TTD in the year following FCE. For subjects who had experienced claim closure, recurrence was defined as re-opening of the subject's back claim or the filing of a new back claim. Factors predicting whether any recurrent events occurred (whether TTD restarted, the claim was re-opened, or a new claim filed) were also analyzed for the entire cohort, as relying solely on benefit or claim status was more likely to underestimate overall recurrence of back injury claim-related events.[3] For example, some claimants may have open claims for the entire follow-up year yet experience recurrent episodes in the form of TTD restarting. All sustained recovery outcomes were dichotomous variables.

*6.2.4 Analysis*

To determine relationships between the FCE performance indicators and the dichotomous sustained recovery outcomes, logistic regression was used.[24] The crude effects of the FCE performance indicators were adjusted using a risk factor modeling strategy by entering them separately into multivariable logistic regression models along with other factors prognostic at a 0.20 alpha level or which altered an FCE regression coefficient by 20% or more.[25] This allowed control for potentially confounding variables while diminishing the number of independent variables in the multivariable regressions. Tests for linearity were performed. All calculations were performed using the computer application SPSS (Chicago, IL).

## 6.3 Results

Of the 278 subjects meeting inclusion criteria, 226 (81%) had complete data on all measures of interest and were included in subsequent analyses. Subjects with missing data had a significantly longer time between injury and FCE and more frequently spoke a language other than English but were similar on the other ten traits examined (Table I). Subjects were predominantly employed males, with

non-specific diagnoses (primarily sprain or strain). The majority of subjects (95%) were rated as failing one or more FCE tasks, therefore, the number of failed FCE tasks was considered a continuous variable for subsequent analyses.

Of the 226 subjects in the cohort, 157 (69%) were receiving TTD benefits at the time of the FCE and had benefits suspended within a year whereas 196 subjects (87%) experienced claim closure. Thirty subjects in the cohort had open claims for the entire follow-up year, yet had the potential for recurrent TTD episodes.

Of the 157 subjects experiencing TTD suspension, 19 (12%) restarted TTD benefits within the same year. The average amount of time between the FCE and TTD restarting was 117 days, and most subjects restarted within 3 months of initial suspension (median 37 days). Of the 196 subjects experiencing claim closure in the year following FCE, 28 (14%) had their claim reopened or filed a new back related claim. Overall, 46 of the 226 subjects (20%) experienced a recurrent claim-related event defined as either restarting TTD, reopening their claim or filing a new back-related claim, within the year following FCE. When the number of FCE tasks was dichotomized at the median of eight failed tasks for purposes of description, 16% percent of those with a higher number of failed tasks had recurrent events, as compared to 25% of those with fewer failed tasks. None of the 11 subjects who passed all FCE tasks experienced a future recurrent event.

Logistic regression modeling indicated a higher number of failed FCE tasks was consistently related to a lower likelihood of experiencing any recurrent event, after controlling for the number of previous back claims, the Pain Disability Index, gender, and clinician recommendation after administration of the FCE (Table II). Adjusting for the number of days between FCE and TTD suspension and days to claim closure did not alter the direction of this relationship and only minimally changed (<3%) the magnitude of the odds ratios. Subjects were approximately 6% (odds ratio 0.94; 95% confidence interval 0.86-1.02) less likely

to experience any future recurrent event for each additional FCE task failed. The maximum amount of weight lifted during the floor-to-waist lift task and rating of whether this task was rated as failed were not consistently associated with recurrence outcomes.

### 6.4 Discussion

In this sample of WCB claimants experiencing TTD suspension or claim closure within a year following FCE administration, the majority of subjects (95%) were not rated as meeting or exceeding physical job demands for all FCE tasks. For this reason, we were unable to dichotomize the FCE variable at this point as originally planned and the number of failed tasks was considered a continuous variable in logistic regression modeling. While none of the 11 subjects passing all FCE tasks experienced a future recurrence, it is likely that this small subgroup is unrepresentative of back injury claimants passing all FCE tasks and, therefore, this may be a chance finding. Only two subjects within this subgroup would have been expected to report a recurrent claims event at the overall recurrence rate of 20% observed for the entire group studied. A larger sample of claimants with long standing back problems who pass all FCE tasks would be needed to clarify the true rate of recurrence in this group.

While better performance on FCE has previously been reported as mildly associated with quicker time to return to work as judged through TTD suspension, the present analyses found better performance on FCE was not associated with fewer recurrences or, alternatively, 'safe' return to work as theorized. An underlying theoretical basis of FCE is that subjects identified as capable of handling physical load exposures at work should have less likelihood of future injury or pain exacerbation. Contrary to this, a higher number of failed FCE tasks was consistently associated with a lower likelihood of recurrence after controlling for potential confounders. Additionally, performance on the floor-to-waist lift was not associated with any sustained recovery outcome. Therefore, within a workers' compensation environment, the validity of sustainable recovery

determinations based on results of the Isernhagen Work Systems' FCE protocol administered is suspect.

One potential reason for the unexpected relationship observed between FCE performance and recurrence includes a premature and therefore riskier return to work in those subjects failing fewer tasks. However, adjusting for measures of time to initial TTD suspension and claim closure did not substantially alter the direction or magnitude of the association between FCE and recurrence. Consequently, it is doubtful that these subjects were at higher risk of recurrence due to earlier return to work.

Another potential explanation is that those subjects failing more FCE tasks alter their work behaviour based on FCE results and are thus at less risk of recurrence. However, given that few subjects passed all FCE tasks, it is likely that most subjects received some recommendations regarding how their work behaviour and activities could be altered to minimize risk. Alternatively, the prognostic effect of FCE observed within this study may be due to unique aspects of clinical or case management decisions and policy within the workers' compensation system from which subjects came and relationships may differ in alternate jurisdictions.

The decision to use a measurement tool to assist in the management of health conditions must be made after weighing the validity and usefulness of the information obtained against the feasibility and administrative burdens associated with using the tool.[26] In the case of FCE, predictive validity has been considered a key psychometric property, as predicting ability to resume and sustain future work is a primary use. Our earlier findings and those of others indicate FCE adds modestly to the prediction of initial return to work beyond that provided by other prognostic factors.[12,13,20] Our current results are the first related to prediction of sustained recovery and are contrary to FCE's theoretical basis regarding which clients are likely to avoid future recurrence. Case managers, health care professionals and other users of FCE information will have to decide whether the

limited validity and usefulness of the information obtained with FCE outweighs the administrative burden and costs associated with such testing.

Low back pain is a recurring phenomenon and our estimates of the rate of recurrence for compensated low back pain are similar to those reported previously in the literature.[3] When focusing on re-opened or new claims only, our recurrence rate of 14% is similar to that calculated by MacDonald et al of individuals indemnified under a large workers' compensation insurer.[27] As well, our finding of a 20% rate of all future recurrent events was consistent with findings of Watson et al and Rossignol et al who have reported rates of 18 and 36% respectively when payment for lost working days was considered.[28,29] Variability in the estimated rate of recurrence between studies is likely due to varied length of follow-up and different definitions of what constitutes recurrence within the various jurisdictions examined.

While we have presented the first data related to the prognostic value of FCE in predicting sustained recovery, our findings are limited by the fact they are reliant on archived database information. While claim and benefit status are important personal and societal indicators, other health related outcomes including future self-rated pain and disability would provide a more detailed description of claimant recovery. In addition, all potential confounders were not available within the databases and other important factors may yet be identified which alter or explain more fully the observed relationship between FCE and recurrence. Other limitations of our data are a relatively small sample size and number of events, which likely diminished our ability to observe significant relationships, and possible measurement error as the reliability and validity of the methods of determining physical job demands for comparison to FCE performance have not previously been studied.

## 6.5 Conclusion

Within a sample of workers' compensation claimants with low back pain who had underwent FCE and subsequently experienced recovery as judged administratively through claims related outcomes, contrary to our hypothesis, better performance on FCE as defined by a lower number of failed FCE tasks was associated with higher risk of recurrence. The validity of FCE's purported ability to identify claimants who are 'safe' to return is not supported.

## Table 6-1

| SUBJECT CHARACTERISTICS (n=226) | |
|---|---|
| **Variable** | **Mean (St Dev) or Percentage** |
| Age (yr) | 41 (9.4) |
| Days From Injury to Functional Capacity Evaluation (FCE) * | 555 (1205) |
| Failed FCE Tasks (out of 25) | 8.5 (5.0) |
| Pain Disability Index (%) | 54 (21) |
| Pain Visual Analogue Scale | 5 (2.1) |
| Annual Salary (thousands of dollars) | 33.3 (18.1) |
| Number of Pre-FCE Health Care Visits | 33 (30) |
| Pre-Injury WCB Back Claims | 1 (2) |
| Male (%) | 71 |
| Employed (%) | 69 |
| Diagnosis (%) | |
| Non-specific (Sprain/ strain) | 80 |
| Disc Pathology | 14 |
| Fracture/ Dislocation | 6 |
| National Occupational Classification (%) | |
| Limited (<5kg) | 6 |
| Light (5-10kg) | 31 |
| Medium (10-20kg) | 37 |
| Heavy (>20kg) | 26 |
| Receiving Total Temporary Disability Benefits (%) | 69 |
| * Median 190 | |

## TABLE 6-2

**The Relationship Between Indicators of Performance on Functional Capacity Evaluation and Sustained Recovery**

| *Variable* | *Crude OR* (95% CI)* | *Adjusted‡ OR (95% CI)* |
|---|---|---|
| **New or Recurrent Claim** | | |
| *n=196, 28 (14%) re-opened* | | |
| Number of Failed Tasks | 0.95 (0.87-1.03) | 0.96 (0.87-1.06) |
| Floor-to-waist Lift (10kg units) | 1.06 (0.73-1.55) | 0.74 (0.44-1.22) |
| Floor-to-waist Lift (Pass/Fail) | 0.66 (0.21-2.02) | 0.43 (0.11-1.64) |
| **TTD Restart** | | |
| n=157, 19 (12%) restarted | | |
| Number of Failed Tasks | 1.01 (0.92-1.11) | 0.96 (0.86-1.08) |
| Floor-to-waist Lift (10kg units) | 1.10 (0.71-1.72) | 1.13 (0.66-1.95) |
| Floor-to-waist Lift (Pass/Fail) | 1.21 (0.37-3.94) | 1.92 (0.52-7.13) |
| **All Recurrent Events** | | |
| n=226, 46 (20%) recurred | | |
| Number of Failed Tasks | 0.95 (0.89-1.02) | 0.94 (0.87-1.02) |
| Floor-to-waist Lift (10kg units) | 1.14 (0.84-1.54) | 0.92 (0.62-1.38) |
| Floor-to-waist Lift (Pass/Fail) | 1.13 (0.49-2.56) | 1.19 (0.46-3.05) |

* Odds Ratio

‡ Adjusted for the number of previous back claims, the Pain Disability Index, gender, and clinician recommendation following FCE administration.

## 6.6 References


1. de Vet HC, Heymans MW, Dunn KM, et al. Episodes of low back pain: a proposal for uniform definitions to be used in research. *Spine* 2002; 27:2409-16.
2. Carey TS, Garrett JM, Jackman A, Hadler N. Recurrence and care seeking after acute back pain: results of a long-term follow-up study. North Carolina Back Pain Project. *Med Care* 1999; 37:157-64.
3. Wasiak R, Pransky GS, Webster BS. Methodological challenges in studying recurrence of low back pain. *J of Occup Rehab* 2003; 13:21-31.
4. Infante-Rivard C, Lortie M. Relapse and short sickness absence for back pain in the six months after return to work. *Occup Environ Med* 1997; 54:328-34.
5. Seferlis T, Nemeth G, Carlsson AM. Prediction of functional disability, recurrences, and chronicity after 1 year in 180 patients who required sick leave for acute low-back pain. *J Spinal Disord* 2000; 13:470-7.
6. van den Hoogen HJ, Koes BW, Deville W, van Eijk JT, Bouter LM. The prognosis of low back pain in general practice. *Spine* 1997; 22:1515-21.
7. Innes E, Straker L. A clinician's guide to work-related assessments: 1-purposes and problems. *Work* 1998; 11:183-189.
8. Isernhagen SJ. Functional capacity evaluation: rationale, procedure, utility of the kinesiophysical approach. *J of Occup Rehab* 1992; 2:157-168.
9. McGill SM. The biomechanics of low back injury: implications on current practice in industry and the clinic. *J Biomech* 1997; 30:465-75.
10. Isernhagen SJ. *The comprehensive guide to work injury management.* Gaithersburg, Md.: Aspen Publishers, 1995:xxxii, 821.
11. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13:125-52.
12. Fishbain DA, Cutler RB, Rosomoff H, Khalil T, Abdel-Moty E, Steele-Rosomoff R. Validity of the dictionary of occupational titles residual functional capacity battery. *Clin J Pain* 1999; 15:102-10.

13. Matheson LN, Isernhagen SJ, Hart DL. Relationships among lifting ability, grip force, and return to work. *Phys Ther* 2002; 82:249-56.
14. Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Res Treat* 1998; 52:289-303.
15. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgements of level of lift effort in a kinesiophysical Functional Capacity Evaluation. *Work* 1999; 12:145-50.
16. United States. Employment and Training Administration. *Dictionary of occupational titles*. Indianapolis: JIST Works, 1991:2 v. in 1 (1404 ).
17. Gardener L, McKenna K. Reliability of occupational therapists in determining safe, maximal lifting capacity. *Aus Occ Ther J* 1999; 46:110-19.
18. Gross DP, Battié MC. Reliability of safe maximum lifting determinations of a functional capacity evaluation. Phys Ther 2002; 82:364-71.
19. Reneman MF, Dijkstra PU, Westmaas M, Goeken LN. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehabil* 2002; 12:269-75.
20. Gross DP, Battié MC. The Prognostic Value of Functional Capacity Evaluation in Patients with Chronic Low Back Pain: Part 1. Timely Return. *Spine* (*in press*)
21. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: psychometric properties. *Pain* 1990; 40:171-82.
22. Gronblad M, Hupli M, Wennerstrand P, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain* 1993; 9:189-95.
23. Canada. Human Resources Development Canada. *National occupational classification, 2001 : Occupational descriptions*. [Ottawa]: Human Resources Development Canada, 2001:1 v. (loose-leaf).

24. Hosmer DW, Lemeshow S. *Applied logistic regression.* Wiley series in probability and statistics. Texts and references section. New York: Wiley, 2000:xii, 373.

25. Hosmer DW, Lemeshow S. *Applied survival analysis : regression modeling of time to event data.* Wiley series in probability and statistics. Texts and reference section. New York: Wiley, 1999:xiii, 386.

26. Finch E, Brooks D, Stratford P, Mayo N. *Physical rehabilitation outcome measures: A guide to enhanced clinical decision making.* Toronto: Canadian Physiotherapy Association, 2002.

27. MacDonald MJ, Sorock GS, Volinn E, Hashemi L, Clancy EA, Webster B. A descriptive study of recurrent low back pain claims. *J Occup Environ Med* 1997; 39:35-43.

28. Watson PJ, Main CJ, Waddell G, Gales TF, Purcell-Jones G. Medically certified work loss, recurrence and costs of wage compensation for back pain: a follow-up study of the working population of Jersey. *Br J Rheumatol* 1998; 37:82-6.

29. Rossignol M, Suissa S, Abenhaim L. The evolution of compensated occupational spinal injuries. A three-year follow-up study. *Spine* 1992; 17:1043-7.

# CHAPTER 7

## General Discussion and Conclusions

### 7.1 Introduction

Functional capacity evaluations (FCE) are standardized batteries of physical performance tests aimed at identifying an injured worker's ability for work-related activity.[1, 2] FCE use is becoming increasingly common and both workers' compensation boards and major insurers are relying more frequently on FCE results to inform return to work decisions.[3] However, the choice to use a measurement tool must be made through weighing the usefulness and validity of the information obtained against the feasibility, cost and consequences of using the tool.[4] In the case of FCE, high administrative burden results as testing is time consuming and expensive, and major decisions are made in the lives of injured workers based on testing. Thus, acceptable measurement properties and adequate theoretical foundations are essential, yet few related studies have been published.[5,6] Four studies were therefore undertaken of the reliability and validity of the Isernhagen Work Systems' FCE administered on workers' compensation claimants with low back injuries. The individual properties studied will be discussed and results synthesized with a focus on findings that have implications on the use, interpretation, and theoretical basis of FCE.

### 7.2 Reliability

Reliability is the consistency of a measure.[4] Individual forms of test consistency that have been judged most important in the case of FCE include interrater and test-retest reliability.[5,7] During FCE, raters must agree with other raters viewing the same subject's performance and determinations should not vary substantially when subjects are tested over multiple occasions separated by intervals in which no functional change is anticipated. If results are too variable, determinations of subject ability will not be meaningful.

Chapter 3 describes a study of the interrater and test-retest reliability of rater judgments of safe, maximal performance using a kinesiophysical approach to testing during the lifting and carrying subtests of the Isernhagen FCE in subjects with low back pain. For subjects willing to participate in both testing occasions, the level of rater agreement was high.[8] This is consistent with findings of Reneman et al, Gardner and McKenna, Isernhagen et al, and Smith who have also reported acceptable levels of rater agreement on kinesiophysical FCE lifting determinations after provision of an operational definition of what constitutes a safe lift.[9-13] In summary, the reliability of rater determinations of safe, maximal performance using kinesiophysical criteria appears acceptable for clinical use.

While rater judgments and most subjects' FCE results appear stable, the two reliability studies involving repeated observation of actual subject performance as opposed to videotaped performance highlight the variability in some subjects' abilities and the important contribution of perceptions, beliefs and motivations to this variability. Back pain is by nature a fluctuating and recurring phenomenon, with alterations in pain and function reported from day to day and at times from hour to hour.[14-16] In a reliability study of rater maximal lifting determinations in patients with low back pain, Reneman et al observed improved performance (increased 6-9% on average) on a second testing occasion potentially due to subject familiarity with the testing situation.[10] In our study, a portion of enrolled subjects (11%) who participated in the first day of testing did not believe they were capable of performing any manual handling on the second occasion due to increased low back pain which was not attributed to day one activities. While the non-participation rate may be lower when FCE is performed within environments with consequences for non-compliance, the subjects' perceptions of pain and disability represent valid beliefs concerning personal potential for physical activity. The overall effect of variable subject perceptions of ability on FCE performance is currently unknown.

Due to the fluctuating nature of low back pain and its associated disability, brief physical performance tests such as FCE, even those administered with reliable observing methods over multiple occasions, are unlikely to capture all of the variability in all subjects' perceived or demonstrated functional abilities. Additionally, subject perceptions of inability to participate create a floor effect inherent in FCE and all performance-based measures due to the volitional nature of testing. Reluctant or cautious subjects cannot be forced to participate or work to rater-determined maximum levels, and in such situations self-report tests of functional ability may potentially have higher compliance and discriminatory power.

### 7.3 Construct Validity

Currently, the Isernhagen FCE with its kinesiophysical approach to testing is used within the WCB-Alberta system as an 'objective' and impartial measure of functional recovery following injury.[17] It is valued due to its reliance on rater observations of biomechanical and physiological signs.[18,19] Functional ability is conceptualized as separate and distinct from experienced pain, and operational definitions of maximum performance during kinesiophysical testing do not consider increasing pain intensity ratings a valid indicator of maximum performance.[8, 9, 17] Functional impairment is attributed to the compensable injury if a subject is judged to have exerted maximal effort.

While it appears raters agree on what constitutes safe, maximal performance, it is not known whether rater determinations are truly independent from levels of pain intensity as purported, especially in cases of pain-mediated disability such as low back pain.[20] Potentially, rater determinations and subject performance during volitional FCE testing are influenced consciously or subconsciously by the interaction between the rater and subject, which includes subject complaints of pain during testing.[21] If subject complaints during testing modify rater perceptions during kinesiophysical FCE, even though an external rater is making an

evaluation of the workers' functional ability, the judgments would be rendered less than 'objective'.

To examine this issue we investigated the construct validity of kinesiophysical FCE, described in detail in Chapter 4.[22] Relationships between two indicators of FCE performance, a valid self-report disability questionnaire and a measure of pain intensity, were examined in a sample made largely of individuals with non-specific low back pain. While other researchers have examined the relationship between FCE and perceived disability, our study was the first to examine the relationship between kinesiophysical FCE performance and pain intensity ratings. Results showed that the FCE performance indicators correlated moderately with both the perceived disability questionnaire and rating of pain intensity.

The moderate association between FCE and the disability questionnaire is supportive of the construct validity of FCE as a measure of function and is consistent with the results of previous investigations.[23] Reneman et al studied a group of subjects with chronic low back pain, and reported a moderate Spearman rank correlation between categorical rating of lifting performance during kinesiophysical FCE and the Oswestry Disability Questionnaire, and between FCE and the Quebec Back Pain Disability Questionnaire.[24] Hart also studied individuals with chronic low back pain and reported a moderate Pearson correlation between amount of weight lifted during FCE and the Oswestry Disability Questionnaire.[25]

While the FCE performance indicators we studied appear to be valid indicators of function, both FCE indicators also correlated to a moderate degree with the pain intensity measure. This finding is contrary to the theoretical basis of kinesiophysical testing which purports functional ability is unrelated and separate from subject reports of pain intensity, but is consistent with the conceptualization of low back pain as a non-specific, pain-mediated disability condition.[26] In most cases of compensated low back pain, physical examination and diagnostic

imaging results are unremarkable and pain is reported as the major factor limiting activity and return to work. In such instances, pain is also likely to influence demonstrated ability during functional testing, which may explain the moderate association observed between pain intensity ratings and FCE performance. In summary, kinesiophysical FCE does not appear to be a completely impartial and 'objective' measure, capable of identifying subjects' functional abilities independent of pain experiences in disability cases resulting from low back pain.

### 7.4 Predictive Validity

Kinesiophysical FCE appears to be a reliable indicator of function in patients with low back pain. However, its relationship with important clinical outcomes is unknown. Predicting ability to safely return to work, devoid of future pain exacerbation or injury reporting, is a predominant purpose of FCE. Determination of FCE predictive validity is important, especially in the cases in which FCE is most frequently used, as returning to sustainable work is influenced by environmental and personal factors in addition to functional ability consistent with a multifactorial model of disability.[27-31] Overall, accurate prediction of return to work following low back injury has proven very difficult and variables found important in prognosis seem to vary from jurisdiction to jurisdiction.[32-34] Consistently psychosocial and contextual factors, such as perceptions of health and disability and employment status, have been reported more closely associated with recovery than physical findings, especially in more chronic samples.[35, 36]

Chapters 5 and 6 described investigations of the ability of the Isernhagen FCE to predict timely and sustained return to work when administered on workers' compensation claimants with low back pain problems.[37, 38] We investigated the FCE alone and while controlling for other factors potentially influencing return to work (see Appendix A for further information on other variables studied and Appendix B for our multivariable modeling strategy). No previous investigations of the predictive ability of the matching relationship between FCE performance

and job demands have been reported, nor has FCE's ability to predict sustained recovery been studied.

The decision that a claimant is safe to return to preinjury work levels is made only when his or her performance meets or exceeds required physical job demands on all tasks in the FCE protocol.[18] However, we found that only a small portion of subjects (4%) was rated as passing all tasks, yet most experienced suspension of total temporary disability (TTD) benefits and claim closure in the follow-up year. The finding that few subjects pass all FCE tasks has important clinical and administrative implications. Case managers requesting FCEs should not expect many low back injured claimants to be found completely fit to return to work, while clinicians performing testing must consider alternate methods for making determinations and prognostications regarding fitness to return to work.

In our study, better performance on the three FCE indicators investigated (the number of failed tasks, maximum weight lifted during the floor-to-waist lift, and pass/fail rating on the floor-to-waist lift) was found weakly related to faster time to TTD suspension and claim closure, with each indicator predicting comparably. However, the magnitude of predictive accuracy was small given that administrative decisions related to return to work within this system are made, in part, based on FCE results. Two previous investigations of individuals with back pain have examined the issue of FCE predictive ability while controlling for potential confounders and reported higher prognostic ability for environmental and personal factors as compared to the functional assessment variables.[39, 40]

When investigating sustained return and the likelihood of recurrent problems in subjects experiencing administrative recovery during the follow-up year, we found a higher number of failed FCE tasks related to a lower likelihood of future recurrence. As the aim of FCE is to identify what exposures a worker is capable of tolerating safely by comparing performance levels to required physical job demands, the finding of a higher risk of recurrence in subjects failing fewer tasks

was unexpected.[17, 41] This observation is contrary to the theoretical basis of FCE, which purports that the risk of experiencing pain or injury is minimized through identifying and avoiding intolerable physical exposures.[42, 43]

Some promising exploratory research looking at the concept of minimizing risk through matching a worker's ability to required job demands has indicated that strength testing might be successfully used to identify workers at risk of musculoskeletal pain or injury due to working in jobs that exceed capacities.[44-48] However, these studies were methodologically limited in that potentially confounding factors, such as worker perceptions of ability and previous experience with disabling back pain, were not controlled for. Other prospective studies with controls implemented for confounding factors have found strength testing and physical job demands to be inconsistently related to onset of low back pain and subsequent duration of disability, with those researchers observing significant relationships consistently reporting associations that were small in magnitude.[49-55] Our study of FCE's ability to predict safe recovery following injury is the first to examine the association between future recurrence and the matching relationship between functional abilities and required physical job demands, and does not support the notion that safe physical work levels can be identified for injured workers.

The causes of low back pain and its associated disability are multifactorial. Numerous potential reasons, including intolerable physical load exposures, have been implicated in why some individuals report work-related low back pain or experience long periods of back pain-related work loss while others do not.[56,57] Low job satisfaction, previous episodes of disability, low perceptions of support in the workplace, high psychological job demands, perceptions of inadequate income, along with other factors have also been reported as risk factors for experiencing back pain, delayed recovery from the condition, or recurrence of the problem.[30,53,58-60] As our results maintain, it is unlikely that tests aimed solely at identifying physical risk factors will provide adequate information for identifying

workers who will avoid future low back pain problems. In fact, predictive indices including prognostic indicators from multiple domains, including physical, occupational, psychological and social factors, have been unable to explain large amounts of the variability in any outcome.[57,58] Accurate prediction continues to be difficult but should be performed cautiously within a theoretical framework that incorporates workers' unique personal and environmental circumstances.

Potential reasons for the counter-intuitive relationship observed in our study between subject performance on FCE and future recurrence were discussed in chapter six and include a premature and therefore riskier return in subjects failing fewer tasks, a decreased risk of recurrence in subjects failing more FCE tasks due to alterations in work behaviour based on FCE results, and unique aspects of clinical or case management decisions and policy within the workers' compensation system from which subjects came. Another potential explanation is that those subjects failing numerous FCE tasks and subsequently experiencing TTD benefit termination and claim closure do not in reality return to work and are therefore not at risk of recurrence. Potentially, these subjects are no longer supported by the WCB-Alberta but rely on alternative sources of income replacement such as familial assistance, employment assurance or government welfare aid.[61] We are currently undertaking a prospective study with one-year follow-up contact of a comparable cohort with administration of self-report employment and disability outcomes that will hopefully shed further light on this unexpected finding. In addition to employment and disability outcome measures, measures of other potentially predictive factors (expectations of recovery and perceptions of workplace support) have been incorporated into the WCB-Alberta system that will allow us to investigate the confounding influence of these factors on the relationship between FCE and return to work.

When examining timely return to work, we found performance on the floor-to-waist lift to be as predictive as information on the matching relationship between performance and physical job demands for the entire Isernhagen FCE. This

finding substantiated an earlier result by Matheson et al who reported that weight lifted during the floor-to-waist lift was one of the strongest predictors of return to work within the Isernhagen FCE.[39] It appears some activities in the protocol may be irrelevant for claimants with low back pain. Tasks such as handgrip strength, hand coordination, and crawling, among others, may increase the time and cost of administration but provide little useful or potentially extraneous information for purposes of determining fitness or readiness to return to work. Considering this along with the small level of prognostic accuracy we found for the entire Isernhagen FCE after controlling for confounders, it is doubtful that FCE administrative burdens are currently outweighed by the usefulness of the information obtained. Potentially, administering only the floor-to-waist lift test incorporated into a comprehensive physical examination for back-injured workers could reduce administrative burdens while providing comparable predictive power as the entire FCE and thus warrant the use of such performance-based testing. However, any physical performance test used for purposes of prediction must be considered within the subject's broader personal and environmental context.

### 7.5 Conclusions

Important measurement properties of the Isernhagen FCE were investigated. While rater judgments while using this tool appear to have acceptable levels of reliability for clinical use, the validity of determinations made based on results and the theoretical foundations of functional testing are questionable. Subject performance during FCE appears to be influenced by perceptions of pain and disability. Indicators of performance on the FCE were weakly associated with timely recovery, and performance on the floor-to-waist lift in the FCE protocol was as predictive as the number of failed tasks in the entire protocol. Furthermore, the validity of judgments of readiness and safe ability to return to work based on FCE results are suspect. Users of FCE information will have to decide whether the usefulness of the information obtained with FCE outweighs the associated administrative burden.

## 7.6 References


1. King PM, Tuckwell N, Barrett TE. A critical review of functional capacity evaluations. *Phys Ther* 1998; 78:852-66.
2. Isernhagen SJ. Return to work testing: functional capacity and work capacity evaluation. *Orthop Phys Ther NA* 1992; 1:83-98.
3. Wyman DO. Evaluating patients for return to work. *Am Fam Phys* 1999; 59:844-8.
4. Finch E, Brooks D, Stratford P, Mayo N. *Physical rehabilitation outcome measures: A guide to enhanced clinical decision making.* Toronto: Canadian Physiotherapy Association, 2002.
5. Innes E, Straker L. Reliability of work-related assessments. *Work* 1999; 13:107-24.
6. Innes E, Straker L. Validity of work-related assessments. *Work* 1999; 13:125-52.
7. Velozo CA. Work evaluations: critique of the state of the art of functional assessment of work. *Am J Occ Ther* 1993; 47:203-9.
8. Gross DP, Battié MC. Reliability of safe maximum lifting determinations of a functional capacity evaluation. Phys Ther 2002; 82:364-71.
9. Gardener L, McKenna K. Reliability of occupational therapists in determining safe, maximal lifting capacity. *Aust Occ Ther J* 1999; 46:110-19.
10. Reneman MF, Dijkstra PU, Westmaas M, Goeken LN. Test-retest reliability of lifting and carrying in a 2-day functional capacity evaluation. *J Occup Rehab* 2002; 12:269-75.
11. Reneman MF, Jaegers S, Westmaas M, Goeken L. The reliability of determining effort level of lifting and carrying in a functional capacity evaluation. *Work* 2002; 18:23-27.
12. Smith RL. Therapists' ability to identify safe maximum lifting in low back pain patients during functional capacity evaluation. *JOSPT* 1994; 19:277-81.

13. Isernhagen SJ, Hart DL, Matheson LM. Reliability of independent observer judgements of level of lift effort in a kinesiophysical Functional Capacity Evaluation. *Work* 1999; 12:145-50.

14. Spenkelink CD, Hutten MM, Hermens HJ, Greitemann BO. Assessment of activities of daily living with an ambulatory monitoring system: a comparative study in patients with chronic low back pain and nonsymptomatic controls. *Clin Rehabil* 2002; 16:16-26.

15. Carey TS, Garrett JM, Jackman AM. Beyond the good prognosis. Examination of an inception cohort of patients with chronic low back pain. *Spine* 2000; 25:115-20.

16. Frank JW, Kerr MS, Brooker AS, et al. Disability resulting from occupational low back pain. Part I: What do we know about primary prevention? A review of the scientific evidence on prevention before disability begins. *Spine* 1996; 21:2908-17.

17. Isernhagen SJ. Functional capacity evaluation: rationale, procedure, utility of the kinesiophysical approach. *J Occup Rehab* 1992; 2:157-168.

18. Isernhagen SJ. *The comprehensive guide to work injury management.* Gaithersburg, Md.: Aspen Publishers, 1995:xxxii, 821.

19. Egan DA. *Functional assessments, work hardening, and work conditioning: A review for the Workers' Compensation Board of Nova Scotia.* 1996.

20. Council JR, Ahern DK, Follick MJ, Kline CL. Expectancies and functional impairment in chronic low back pain. *Pain* 1988; 33:323-31.

21. Rudy TE, Dieber SJ, Boston JR. Functional capacity assessment: Influence of behavioural and environmental factors. *J Back Musculoskel Rehabil* 1996; 6:277-288.

22. Gross DP, Battié MC. The construct validity of a functional capacity evaluation administered within a workers' compensation environment. *J Occup Rehab*; (*in press*).

23. Domholdt E. *Physical therapy research : principles and applications.* Philadelphia, Pa.: W.B. Saunders, 2000:xiv, 522.

24. Reneman MF, Jorritsma W, Schellekens JM, Goeken LN. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. *J Occup Rehab* 2002; 12:119-29.

25. Hart DL. Relation between three measures of function in patients with chronic work-related pain syndromes. *J Rehab Outcome Measure* 1998; 2:1-14.

26. Deyo RA. Diagnostic evaluation of LBP: reaching a specific diagnosis is often impossible. *Arch Intern Med* 2002; 162:1444-7; discussion 1447-8.

27. Waddell G. 1987 Volvo award in clinical sciences. A new clinical model for the treatment of low-back pain. *Spine* 1987; 12:632-44.

28. Truchon M, Fillion L. Biopsychosocial determinants of chronic disability and low-back pain: a review. *J Occup Rehab* 2000; 10:117-42.

29. Carey TS, Garrett JM, Jackman A, Hadler N. Recurrence and care seeking after acute back pain: results of a long-term follow-up study. North Carolina Back Pain Project. *Med Care* 1999; 37:157-64.

30. Seferlis T, Nemeth G, Carlsson AM. Prediction of functional disability, recurrences, and chronicity after 1 year in 180 patients who required sick leave for acute low-back pain. *J Spinal Disord* 2000; 13:470-7.

31. Crook J, Milner R, Schultz IZ, Stringer B. Determinants of occupational disability following a low back injury: a critical review of the literature. *J Occup Rehabil* 2002; 12:277-95.

32. Schultz IZ, Crook JM, Berkowitz J, et al. Biopsychosocial multivariate predictive model of occupational low back disability. *Spine* 2002; 27:2720-5.

33. McIntosh G, Frank J, Hogg-Johnson S, Bombardier C, Hall H. Prognostic factors for time receiving workers' compensation benefits in a cohort of patients with low back pain. *Spine* 2000; 25:147-57.

34. Gatchel RJ, Polatin PB, Mayer TG. The dominant role of psychosocial risk factors in the development of chronic low back pain disability. *Spine* 1995; 20:2702-9.

35. Hunt DG, Zuberbier OA, Kozlowski AJ, et al. Are components of a comprehensive medical assessment predictive of work disability after an episode of occupational low back trouble? *Spine* 2002; 27:2715-9.

36. van der Giezen AM, Bouter LM, Nijhuis FJ. Prediction of return-to-work of low back pain patients sicklisted for 3-4 months. *Pain* 2000; 87:285-94.

37. Gross DP, Battié MC. The prognostic value of functional capacity evaluation in patients with chronic low back pain: Part 2. Sustained recovery. *Spine (in press)*.

38. Gross DP, Battié MC, Cassidy JD. The prognostic value of functional capacity evaluation in patients with chronic low back pain: Part 1. Timely return. *Spine (in press)*.

39. Matheson LN, Isernhagen SJ, Hart DL. Relationships among lifting ability, grip force, and return to work. *Phys Ther* 2002; 82:249-56.

40. Fishbain DA, Cutler RB, Rosomoff H, Khalil T, Abdel-Moty E, Steele-Rosomoff R. Validity of the dictionary of occupational titles residual functional capacity battery. *Clin J Pain* 1999; 15:102-10.

41. Schonstein E, Kenny DT. The value of functional and work place assessments in achieving a timely return to work for workers with back pain. *Work* 2001; 16:31-38.

42. Keyserling WM. Workplace risk factors and occupational musculoskeletal disorders, Part 1: A review of biomechanical and psychophysical research on risk factors associated with low-back pain. *AIHAJ* 2000; 61:39-50.

43. McGill SM. The biomechanics of low back injury: implications on current practice in industry and the clinic. *J Biomech* 1997; 30:465-75.

44. Snook SH, Campanelli RA, Hart JW. A study of three preventive approaches to low back injury. *J Occup Med* 1978; 20:478-81.

45. Pedersen DM, Clark JA, Johns RE, Jr., White GL, Hoffman S. Quantitative muscle strength testing: a comparison of job strength requirements and actual worker strength among military technicians. *Mil Med* 1989; 154:14-8.

46. Keyserling WM, Herrin GD, Chaffin DB, Armstrong TJ, Foss ML. Establishing an industrial strength testing program. *Am Ind Hyg Assoc J* 1980; 41:730-6.

47. Keyserling WM, Herrin GD, Chaffin DB. Isometric strength testing as a means of controlling medical incidents on strenuous jobs. *J Occup Med* 1980; 22:332-6.

48. Reimer DS, Halbrook BD, Dreyfuss PH, Tibiletti C. A novel approach to preemployment worker fitness evaluations in a material-handling industry. *Spine* 1994; 19:2026-32.

49. Bigos SJ, Battié MC, Fisher LD, Hansson TH, Spengler DM, Nachemson AL. A prospective evaluation of preemployment screening methods for acute industrial back pain. *Spine* 1992; 17:922-6.

50. Smedley J, Egger P, Cooper C, Coggon D. Prospective cohort study of predictors of incident low back pain in nurses. *BMJ* 1997; 314:1225-8.

51. Mostardi RA, Noe DA, Kovacik MW, Porterfield JA. Isokinetic lifting strength and occupational injury. A prospective study. *Spine* 1992; 17:189-93.

52. Nordin M, Skovron ML, Hiebert R, et al. Early predictors of delayed return to work in patients with low back pain. *J Musculo Pain* 1997; 5:5-27.

53. Krause N, Dasinger LK, Deegan LJ, Rudolph L, Brand RJ. Psychosocial job factors and return-to-work after compensated low back injury: a disability phase-specific analysis. *Am J Ind Med* 2001; 40:374-92.

54. Feyer AM, Herbison P, Williamson AM, et al. The role of physical and psychological factors in occupational low back pain: a prospective cohort study. *Occup Environ Med* 2000; 57:116-20.

55. Milhous RL, Haugh LD, Frymoyer JW, et al. Determinants of vocational disability in patients with low back pain. *Arch Phys Med Rehabil* 1989; 70:589-93.

56. Hoogendoorn WE, Bongers PM, de Vet HC, Ariens GA, van Mechelen W, Bouter LM. High physical work load and low job satisfaction increase

the risk of sickness absence due to low back pain: results of a prospective cohort study. *Occup Environ Med* 2002; 59:323-8.

57. Bigos SJ, Battié MC, Spengler DM, et al. A longitudinal, prospective study of industrial back injury reporting. *Clin Orthop* 1992:21-34.

58. Adams MA, Mannion AF, Dolan P. Personal risk factors for first-time low back pain. *Spine* 1999; 24:2497-505.

59. Papageorgiou AC, Macfarlane GJ, Thomas E, Croft PR, Jayson MI, Silman AJ. Psychosocial factors in the workplace--do they predict new episodes of low back pain? Evidence from the South Manchester Back Pain Study. *Spine* 1997; 22:1137-42.

60. Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine* 2002; 27:E109-20.

61. Workers' services (Returning to work). Workers' Compensation Board Alberta Website, accessed April 30, 2003.

## APPENDIX A

## Potential Confounding Variables Extracted for Prediction Studies

Besides information on the Functional Capacity Evaluation (FCE), other variables were extracted from WCB-Alberta databases for analysis in the predictive validity studies based on their potential for a confounding influence on the relationship between FCE results and our outcomes. We selected variables from the clinical and administrative databases that had been reported as predictive of recovery in previously published studies or where a sound theoretical rationale existed for considering the variable. The clinical database was the Client Admissions Tracking System from Millard Health, the major WCB-Alberta rehabilitation facility serving injured workers through occupational rehabilitation programs, work assessment, and vocational rehabilitation. The main WCB-Alberta administrative database is named the Client Information System. Millard Health and the WCB-Alberta routinely collect data in a standard fashion on all of the variables included in these databases. The databases have previously been used successfully for research and are commonly used by WCB management and policy-makers for purposes of decision-making and report generation to stakeholders within the Alberta compensation system.[1]

Independent variables selected from the databases for the proposed study include the following:

*Pain Disability Index (PDI)* - The PDI is a commonly used self-report questionnaire measuring perceived disability due to pain. Perceived disability has previously been found to be a predictor of return-to-work in persons with LBP.[2-6] Perceived disability has also been found to influence results of functional testing in individuals with chronic back pain and to correlate weakly with FCE.[7-9] It was initially developed by Pollard and has since been tested extensively in a variety of patient populations including those with chronic back pain and good psychometric properties have consistently been reported.[10-16] It asks patients to rate their level of disability on a 0-10 scale on seven areas of activity: family/home

responsibility, recreation, social activity, occupation, sexual behaviour, self-care, and life-support activity. The patient's scores on these seven items are summed and a total out of 70 is obtained. The higher the patient's score out of 70, the higher the individual's level of perceived disability. The PDI has good test-retest reliability, high internal consistency and good concurrent validity when tested against the Oswestry Disability Questionnaire in a group of subjects with LBP. It was also found to have higher test-retest reliability and was more responsive to change than the Oswestry Low Back Disability Scale. The PDI has been shown to accurately differentiate between patients who had just undergone back surgery from patients with LBP but working full-time. The PDI is standardly administered at Millard Health at the beginning and end of each client's rehabilitation program, and each time an FCE is performed.

*Pain Intensity Rating* - Pain is rated at Millard Health on a visual analogue scale in which the client places a mark on a line, anchored by 0 and 10, to measure their level of pain most of the time. Zero corresponds to no pain and 10 to the worst pain imaginable. Analogue pain scales have been previously shown to have adequate reliability and validity and to be inconsistently associated with future return-to-work and recurrence.[3-5, 17-21]

*Clinician Recommendation* – This variable was measured as whether the claimant was rated by the clinician administering the FCE as ready to return to work (preinjury or modified levels) or not at time of discharge. Recommendations are made, in part, based on FCE results but also incorporate all other information available to the clinician that may enhance prognostic accuracy.

*Duration of injury* - This variable was measured as the number of days between injury and FCE. Measures of duration of episode have consistently been associated with subsequent time to return-to-work and future recurrence in prior studies.[2, 3, 5, 22-25]

*Occupational physical loading classification* - Physical loading classifications were determined from the National Occupational Classification (NOC) code for each subject's occupation and standardly categorized as limited (up to 5kg), light (5kg but less than 10kg), medium (between 10kg and 20kg), and heavy (over

20kg) for manual handling. A second physical demand variable was created from information within the NOC on the use of other body positions including sitting, standing, walking, or extreme positions including bending, stooping, kneeling, and crouching. Occupational load exposures have been inconsistently related to future return to work.[6, 19, 24, 26]

*Diagnosis* - ICD9 codes were used to create a dichotomous variable indicating whether the subject's injury was diagnosed as a sprain/strain/ or simple back pain, intervertebral disc-related diagnosis or other diagnosis including fracture/ dislocation. Subjects sustaining more severe injuries such as fracture or dislocation may be at risk of delayed return to work.[1, 19, 27]

*Health Care Utilization* – This variable represented the total number of visits sponsored by the WCB-Alberta for medical and allied healthcare, including medical, physical therapy, or chiropractic visits between the date of injury and FCE. Subjects undergoing excessive visits to health care providers may be at risk of delayed recovery theoretically due to increased severity of the problem, pathological coping strategies or potentially due to iatrogenesis.[28, 29]

*Salary* – This variable was measured as the claimant's gross annual pre-injury salary. We hypothesized those subjects with salaries approaching but below the WCB-Alberta maximum annual insurable earnings limit ($45,600 in 1999) may be less motivated to return to work sooner. Diminished ability to 'get along' on income has been found related to report of back injury and breadwinner status has been associated with faster recovery.[30, 31]

*Language* - A dichotomous variable was created indicating whether or not English was the subject's primary language. Potentially language poses a barrier to effective communication and therefore delayed recovery. Alternatively cultural differences may lead to observed variations in recovery rates.

*Marital status* - There are conflicting reports of the role of marital status in disability in other studies of outcome prediction.[2,3,18,32] Marital status was categorized as single, divorced/separated, married/ common-law or widowed.

*Age* – Age in years at time of FCE. Conflicting findings are present in previous studies with respect to age.[2,3,19,27] Subjects with older age may have closer linkages

to employers and therefore return to work sooner, however, physiologically older tissues may heal at a slower rate.

*Gender* - Conflicting findings have been reported in the scientific literature regarding the role of gender in work-related low back problems.[3, 19, 24, 33]

*Previous Claims* - The total number of previous workers' compensation claims and the total previous back-related claims. A higher number of previous back pain episodes has been found associated with increased risk of recurrence.[5, 20]

## Appendix A References


1. Voaklander DC, Beaulne AP, Lessard RA. Factors related to outcome following a work hardening program. *J Occup Rehab* 1995; 5:71-85.
2. Haldorsen EM, Indahl A, Ursin H. Patients with low back pain not returning to work. A 12-month follow-up study. *Spine* 1998; 23:1202-7; discussion 1208.
3. Gatchel RJ, Polatin PB, Mayer TG. The dominant role of psychosocial risk factors in the development of chronic low back pain disability. *Spine* 1995; 20:2702-9.
4. Seferlis T, Nemeth G, Carlsson AM. Prediction of functional disability, recurrences, and chronicity after 1 year in 180 patients who required sick leave for acute low-back pain. *J Spinal Disord* 2000; 13:470-7.
5. Crook J, Milner R, Schultz IZ, Stringer B. Determinants of occupational disability following a low back injury: a critical review of the literature. *J Occup Rehabil* 2002; 12:277-95.
6. Nordin M, Skovron ML, Hiebert R, et al. Early predictors of delayed return to work in patients with low back pain. *J Musculo Pain* 1997; 5:5-27.
7. Gibson L, Strong J. Assessment of psychosocial factors in functional capacity evaluation of clients with chronic back pain. *Br J Occ Ther* 1998; 61:399-404.
8. Reneman MF, Jorritsma W, Schellekens JM, Goeken LN. Concurrent validity of questionnaire and performance-based disability measurements in patients with chronic nonspecific low back pain. *J Occup Rehab* 2002; 12:119-29.
9. Hart DL. Relation between three measures of function in patients with chronic work-related pain syndromes. *J Rehab Outcome Measure* 1998; 2:1-14.
10. Pollard CA. Preliminary validity study of the pain disability index. *Percept Mot Skills* 1984; 59:974.

11. Gronblad M, Hupli M, Wennerstrand P, et al. Intercorrelation and test-retest reliability of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) and their correlation with pain intensity in low back pain patients. *Clin J Pain* 1993; 9:189-95.

12. Gronblad M, Hurri H, Kouri JP. Relationships between spinal mobility, physical performance tests, pain intensity and disability assessments in chronic low back pain patients. *Scand J Rehabil Med* 1997; 29:17-24.

13. Gronblad M, Jarvinen E, Airaksinen O, Ruuskanen M, Hamalainen H, Kouri JP. Relationship of subjective disability with pain intensity, pain duration, pain location, and work-related factors in nonoperated patients with chronic low back pain. *Clin J Pain* 1996; 12:194-200.

14. Gronblad M, Jarvinen E, Hurri H, Hupli M, Karaharju EO. Relationship of the Pain Disability Index (PDI) and the Oswestry Disability Questionnaire (ODQ) with three dynamic physical tests in a group of patients with chronic low-back and leg pain. *Clin J Pain* 1994; 10:197-203.

15. Tait RC, Chibnall JT, Krause S. The Pain Disability Index: psychometric properties. *Pain* 1990; 40:171-82.

16. Tait RC, Pollard CA, Margolis RB, Duckro PN, Krause SJ. The Pain Disability Index: psychometric and validity data. *Arch Phys Med Rehabil* 1987; 68:438-41.

17. Council JR, Ahern DK, Follick MJ, Kline CL. Expectancies and functional impairment in chronic low back pain. *Pain* 1988; 33:323-31.

18. Gatchel RJ, Polatin PB, Kinney RK. Predicting outcome of chronic back pain using clinical predictors of psychopathology: a prospective analysis. *Health Psychol* 1995; 14:415-20.

19. McIntosh G, Frank J, Hogg-Johnson S, Bombardier C, Hall H. Prognostic factors for time receiving workers' compensation benefits in a cohort of patients with low back pain. *Spine* 2000; 25:147-57.

20. Carey TS, Garrett JM, Jackman A, Hadler N. Recurrence and care seeking after acute back pain: results of a long-term follow-up study. North Carolina Back Pain Project. *Med Care* 1999; 37:157-64.

21. Infante-Rivard C, Lortie M. Relapse and short sickness absence for back pain in the six months after return to work. *Occup Environ Med* 1997; 54:328-34.

22. Haldorsen EM, Wormgoor ME, Bjorholt PG, Ursin H. Predictors for outcome of a functional restoration program for low back pain patients - a 12-month follow-up study. *Euro J Phys Med & Rehab* 1998; 8:103-9.

23. Hadler NM, Carey TS, Garrett J. The influence of indemnification by workers' compensation insurance on recovery from acute backache. North Carolina Back Pain Project. *Spine* 1995; 20:2710-5.

24. Hunter SJ, Shaha S, Flint D, Tracy DM. Predicting return to work. A long-term follow-up study of railroad workers after low back injuries. *Spine* 1998; 23:2319-28.

25. Rossignol M, Suissa S, Abenhaim L. The evolution of compensated occupational spinal injuries. A three-year follow-up study. *Spine* 1992; 17:1043-7.

26. Hazard RG, Haugh LD, Reid S, Preble JB, MacDonald L. Early prediction of chronic disability after occupational low back injury. *Spine* 1996; 21:945-51.

27. Infante-Rivard C, Lortie M. Prognostic factors for return to work after a first compensated episode of back pain. *Occup Environ Med* 1996; 53:488-94.

28. McPhillips-Tangum CA, Cherkin DC, Rhodes LA, Markham C. Reasons for repeated medical visits among patients with chronic back pain. *J Gen Intern Med* 1998; 13:289-95.

29. Cote P, Cassidy JD, Carroll L. The treatment of neck and low back pain: who seeks care? who goes where? *Med Care* 2001; 39:956-67.

30. Papageorgiou AC, Macfarlane GJ, Thomas E, Croft PR, Jayson MI, Silman AJ. Psychosocial factors in the workplace--do they predict new episodes of low back pain? Evidence from the South Manchester Back Pain Study. *Spine* 1997; 22:1137-42.

31. van der Giezen AM, Bouter LM, Nijhuis FJ. Prediction of return-to-work of low back pain patients sicklisted for 3-4 months. *Pain* 2000; 87:285-94.

32. Milhous RL, Haugh LD, Frymoyer JW, et al. Determinants of vocational disability in patients with low back pain. *Arch Phys Med Rehabil* 1989; 70:589-93.

33. Cheadle A, Frankline G, Wolfhagen C, et al. Factors influencing the duration of work-related disability: A population-based study of Washington state worker's compensation. *Am J Public Health* 1994; 84:190-196.

# APPENDIX B

# RISK FACTOR MODELING STRATEGY FOR PREDICTION STUDIES

## Introduction

Logistic regression is a statistical technique used to determine the relationship between an independent variable and a dichotomous outcome.[1] Survival analysis is a collection of statistical techniques for time-to-event data.[2, 3] Both of these techniques have been used previously in attempts to predict recovery from low back pain (LBP), with specific predictor variables analyzed for their relationships with the outcome return to work or time to return. Logistic regression has also been used to determine risk of recurrence. The multivariable risk factor modeling strategy used for determination of the association between the FCE indicators and outcomes in the predictive validity studies using both Cox and logistic regression will be presented.

## Multivariable Prognostic Modeling Strategy

Hosmer and Lemeshow have described two types of model building strategies, prediction modeling and risk factor modeling.[3] Risk factor modeling differs from prediction modeling in that only one risk factor is being studied for its relationship with outcome, and other variables are included to control for a potential confounding influence. The risk factor is entered into the model first and all other variables are added later, resulting in a more accurate estimate of the risk factor's effect. In prediction modeling, all variables selected for study are considered equally, and all have the same probability of ending up in the final model. As our main concern in the FCE predictive validity studies was the relationship between the FCE predictors and outcome, a risk factor modeling strategy with purposive variable selection was used and will be presented based on strategies proposed by Hosmer and Lemeshow.

## Stages of Risk Factor Modeling

We aimed to perform prognostic modeling of FCE's relationship with outcomes in two main stages based on Altman and Lyman's concepts of exploratory analysis and confirmation.[4] Exploratory studies aim to evaluate the association between an individual prognostic indicator and the outcome of interest alone and after adjusting for the confounding effect of other important predictors. Confirmatory studies seek to verify previous exploratory findings as well as establish more accurately the magnitude of predictability added by the indicator of interest.[5] Exploratory analysis was performed on a cohort of subjects undergoing FCE in the year 1999, and confirmation was planned on a cohort of subjects seen in 2000. The confirmation stage was performed with both time to event outcomes, however, due to sample size limitations only exploratory analysis of FCE's relationship with recurrence outcomes was performed on a merged cohort of subjects seen in 1999 and 2000.

The recommended sample size of 10 events per independent variable was adhered to for all regressions.[6-8] All calculations were performed using the computer application SPSS (Chicago, IL).

## Exploratory Analysis

*Bivariate Screening*

Initially, univariate relationships were determined between the individual FCE predictors and dependent variables. Other potential confounding variables were also bivariately screened to determine variables important in prediction. Potential confounding variables not significant at a 0.20 alpha level were removed unless they were found later have an important confounding effect. This initial screening allowed avoidance of overfitting a model and multivariate testing of too many predictor variables for the limited sample size available.

R square for Cox regression statistics were also calculated bivariately for time to event outcomes to determine the magnitude of the variation in the time to

recovery outcomes explained crudely by the FCE indicators. This was performed using the technique described by Schemper and the likelihood statistics generated by SPSS.[9, 10]

*Multivariable Control for Potential Confounders*

The next step in risk factor modeling was to control for a potential confounding effect of other predictor variables on the relationship between the FCE variables and outcomes. This was performed by entering the FCE variables along with all other variables significant from the bivariate screen into a multivariable Cox or logistic regression. The FCE indicators were analyzed in separate multivariable regressions, as they were not considered independent. This procedure allowed odds and hazard rate ratios to be generated for each FCE indicator, adjusted for the other variables important from bivariate analysis. Next, the confounding effect of removed variables was tested by adding each one at a time to the reduced model, and the effect on the FCE indicators observed. However, no removed variable altered the regression coefficient of a FCE variable by 20% or greater. R square for Cox regression statistics were also calculated for FCE variables after controlling for other important predictors using the change in the likelihood statistic after the FCE variable was dropped from the model.

*Evaluating Assumptions*

Logistic and Cox regression models make two main assumptions about the relationships between predictor variables and outcomes that were evaluated. These are the linearity and additivity assumptions. In addition, in Cox regression variables must also meet the proportional hazards assumption. Numerous methods have been described for checking the validity of each. The following discussion should not be considered exhaustive of all potentially useful techniques.

Linearity - All variables measured on a continuous scale were evaluated for the linearity of their relationship with outcomes.[11] A simple yet effective method of

evaluation is performed through the use of dummy variables separating the variable into quartiles.[3] Once four equal divisions are made, the hazard rate ratio for each group is plotted on a graph against the covariate at the midpoint of each group. If the assumption is met, a polygon connecting the four points will approximate a straight line. If a deviation is observed, a transformation of the variable may be employed as indicated by the shape of the polygon. If transformations are considered, the likelihood ratio test should be used to determine if the transformation predicts more effectively than the original variable.

Using the Cox model to determine relationships between predictors and time to total temporary disability benefit suspension, the linearity assumption was judged sufficiently upheld for the number of failed FCE tasks (Figure B-1), and weight lifted on the low level lift. The assumption was questionably met for the number of health care visits preceding the FCE, where the shape of the curve (Figure B-2) seemed to indicate a curvilinear relationship. A quadratic transformation did slightly enhance the predictive ability of this variable and future regressions were performed with this variable transformed.

Additivity - Additivity implies no effect modification takes place between covariates. To evaluate the additivity of regression coefficients, all clinically plausible interactions were first listed. Each interaction was added along with their corresponding main effect variables to a multivariable regression and evaluated for significance. No significant interactions were observed in any of our models, potentially due to limited sample size to detect an important effect.

Proportional Hazards - The function underlying Cox's model is partitioned into a product of the baseline hazard involving time and an exponential expression involving the covariates of interest. This separation assumes an important quality of the data being analyzed that must be met for valid model generation. The assumed quality is that the effect of each covariate is independent of time, or that

the hazard ratios do not change/ are *proportional* over time. Variables not meeting the proportional hazards assumption must be handled through stratified or extended Cox procedures. Both procedures have some disadvantages as stratified Cox procedures do not allow effect sizes to be determined for those variables not meeting the proportional hazards assumption, and extended Cox models are more complex computationally. Evaluation of the proportional hazards assumption was performed through two methods: a) graphical, or b) time-dependent methods

a) The graphical approach uses the log-minus-log survival curve created by transforming the original estimated survival curve by taking the natural log of estimated survival probabilities twice (log-minus-log curves in SPSS). These transformations were performed for various categories of nominal variables. The resulting step functions were then plotted on the same graph and observed visually for parallelism. If the proportional hazards assumption was met, the two step functions were seen to travel together in an approximately parallel fashion towards the top right of the graph. If the variable did not meet the proportional hazards assumption, the step functions would have crossed or markedly diverged. All categorical variables including the pass/fail low level lift rating were judged to meet the proportional hazards assumption (Figure B-3).

Graphical approaches are very effective for gross departures from proportionality, however, they are limited in cases of slighter deviation. Being reliant on visual observation, graphical approaches are subject to human errors in judgment and lack statistical clarity valued in analysis. For example, log-log survival curves that are *approximately* parallel over their length may leave the observer wondering about acceptable limits of parallelism. In these cases, statistical testing involving the creation and evaluation of time-dependent variables is available to more rigorously appraise the assumption.

b) Time-dependent variables are created by multiplying each covariate by time. The resulting interaction terms are thus products of important covariates and the

outcome variable they were predicting. Time-dependent variables were added to the fitted model one at a time and significance of coefficients determined. If the time-dependent term was significantly related to outcome, the proportional hazards assumption was considered unmet. The statistical testing allowed through creation of time-dependent covariates provided an effective method of handling uncertainties arising from graphical methods. The number of health visits preceding the FCE was considered to breech the proportional hazards assumption in relation to time to total temporary disability benefits suspension and was analyzed as a time-dependent variable in multivariable analyses. All other variables were judged to meet the assumption.

**Confirmation**

Unstable regression estimates may be generated in any multivariable regression technique, but especially may arise from analysis on a sample of small size. Investigators must be cautious of "over-fitting" and should take steps to evaluate final models using either regression diagnostics involving residuals, data splitting techniques, bootstrapping or separate external validation studies in which a model is tested for stability in data collected from a separate sample.[4,11] External confirmation on a separate cohort was performed as it was judged the most effective method of validation as results are not dependent upon the original data on which exploratory modeling was performed.

After the adjusted relationships between FCE and time to recovery outcomes were determined and all assumptions accounted for, the stability of the observed relationship was evaluated in a separate cohort formed from subjects undergoing FCE in 2000. This included fitting the same multivariable risk factor models to the new cohort, allowing determination of the stability of the estimated hazard rate ratios (HRR) and estimated proportions of variation explained by FCE. Results were quite consistent between cohorts, implying ours results were upheld and trustworthy.

## FIGURE B-1

**Testing the Linearity of the Relationship Between the Number of Failed Tasks and Time to Suspension of Total Temporary Disability Benefits**

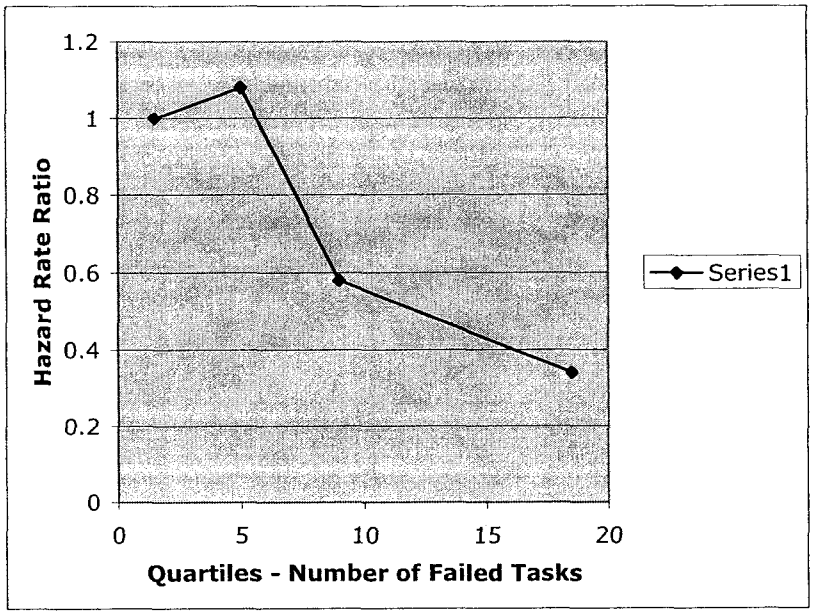


**FIGURE B-2**

**Testing the Linearity of the Relationship Between the Number of Failed Tasks and Time to Suspension of Total Temporary Disability Benefits**

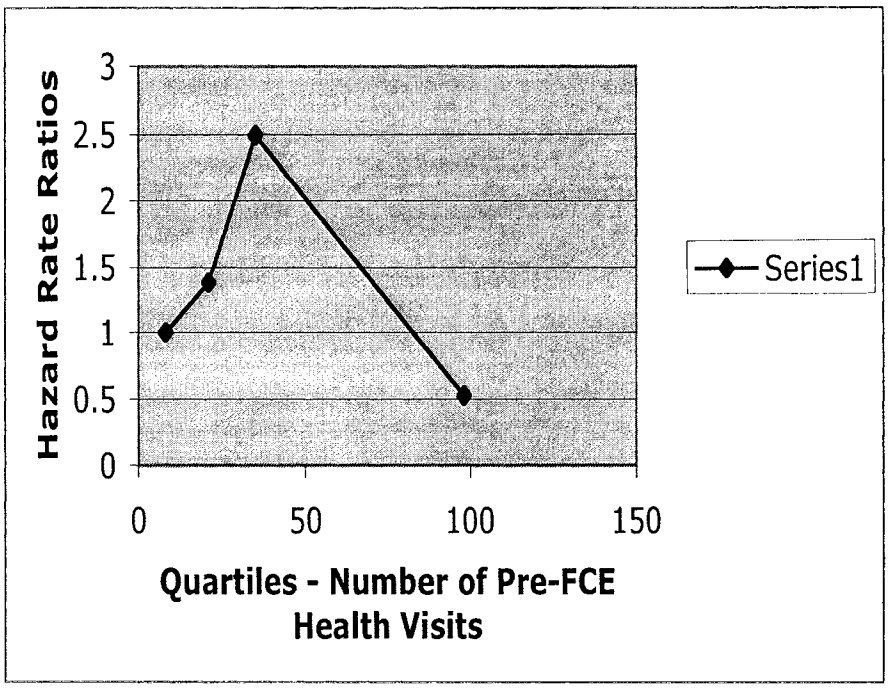
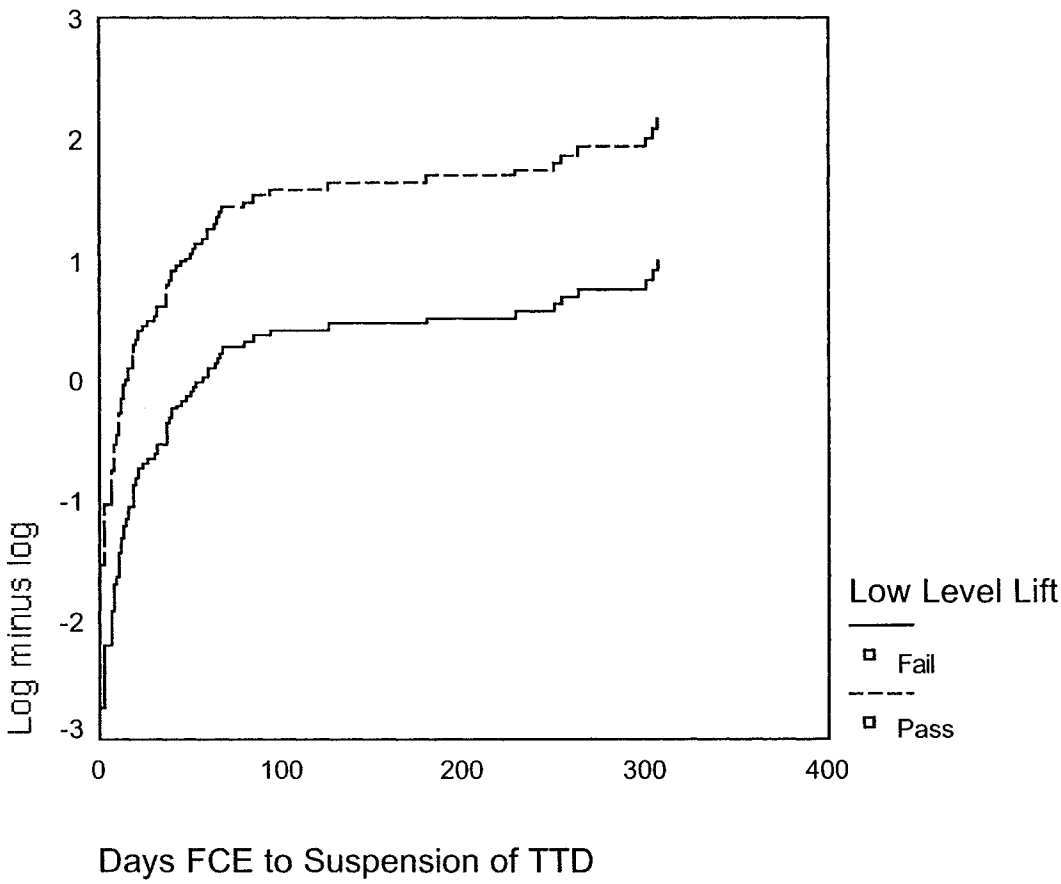


**FIGURE B-3**

**Testing the Proportional Hazards Assumption of the**

**Low Level Lift Pass/ Fail Rating with a Log-Minus-Log Graph**



## REFERENCES


1. Hosmer DW, Lemeshow S. *Applied logistic regression.* Wiley series in probability and statistics. Texts and references section. New York: Wiley, 2000:xii, 373.
2. Kleinbaum DG. *Survival analysis: a self-learning text.* Statistics in the health sciences. New York: Springer, 1996:xii, 324.
3. Hosmer DW, Lemeshow S. *Applied survival analysis: regression modeling of time to event data.* Wiley series in probability and statistics. Texts and reference section. New York: Wiley, 1999:xiii, 386.
4. Altman DG, Lyman GH. Methodological challenges in the evaluation of prognostic factors in breast cancer. *Breast Cancer Res Treat* 1998; 52:289-303.
5. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15:361-87.
6. Fayers PM, Machin D. Sample size: How many patients are necessary? *Br J of Cancer* 1995; 72:1-9.
7. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996; 49:1373-9.
8. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995; 48:1503-10.
9. Schemper M. The relative importance of prognostic factors in studies of survival. *Stat Med* 1993; 12:2377-82.
10. Schemper M, Henderson R. Predictive accuracy and explained variation in Cox regression. *Biometrics* 2000; 56:249-55.
11. Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. *Cancer Treat Rep* 1985; 69:1071-77.