A multivariate statistical model for whole-body related musculoskeletal disorders
by Harish Yerneni

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Industrial and Management Engineering
Montana State University

Abstract:
The incidence of work-related musculoskeletal disorders (MSDs) continues to be a key concern for occupational safety and health care professionals. Several factors such as repetition, forceful exertion, and awkward postures have been linked to their development. While these links have been well established, valid and reliable techniques for measuring MSD risk are lacking, particularly for jobs in non-manufacturing industries or non-repetitive jobs in general.

Marley, et. ah, (1997) examined such jobs in the power distribution industry with a goal of better understanding which work factors may be associated with MSDs. Injury data from over 2000 workers in one company were tabulated by job classification (12 total categories). Three representative categories, electric line crews, gas line crews, and meter readers were identified as having high, medium, and low risk for injury respectively, based on the recorded rate of MSDs in these categories. An ergonomic/work-methods analysis was then performed upon 5 key activities within these jobs. Activities were further broken down into 31 required tasks (e.g., climb pole, make connection, shovel, cut pipe, etc.) and even further into 18 fundamental work elements (e.g., various body postures, grasp type, force level, duration, terrain condition, etc.).

Cluster analysis involving the work element measures resulted in five clusters. Two clusters generally represented upper and lower part of the upper extremities, two clusters generally represented lower extremities and one contained miscellaneous ergonomic variables. All the coefficients of the cluster variable weights in the five clusters resulted in the same sign from principal component analysis I, signifying that the increase in values of any cluster variable in turn increases cluster score and hence the risk level. The clusters are modeled and validated using ordinal logistic regression technique. The model accurately predicted 92% and 76.5% of training and testing data sets respectively. A user-friendly web application of this model targeting the novice user has been developed.

The model needs should be trained with larger data sets for better prediction and more robust applications. However, the current model may be useful for predicting the whole-body related MSDs in the utility industry and comparable non-repetitive jobs. The identical clusters may also be useful in the understanding of physical job stress in these environments.

A MULTIVARIATE STATISTICAL MODEL FOR WHOLE-BODY RELATED MUSCULOSKELETAL DISORDERS

by

Harish Yerneni

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Industrial and Management Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2003

APPROVAL

of a thesis submitted by

Harish Yerneni

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Robert Marley ______________________ 5-16-03
(Signature) Date

Approved for the Department of Mechanical and Industrial Engineering

Dr. Vic Cundy ______________________ 5-19-03
(Signature) Date

Approved for the College of Graduate Studies

Dr. Bruce McLeod Bruce R. McLeod 5-19-03
(Signature) Date

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University – Bozeman, I agree that the library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis (paper) in whole or in parts may be granted only by the copyright holder.

Signature Y. Harish
Date 5-16-03

ACKNOWLEDGEMENTS

I would like to thank my advisor and mentor, Dr. Robert Marley for the outstanding encouragement, guidance and expertise that kept me focused on the goals of this research, for the financial support throughout my program, and more importantly for the faith in my ability to complete this task properly.

I would like to express my deep respect and profound gratitude to Dr. Robert Boik for his invaluable assistance throughout the research. His extreme patience and support are gratefully acknowledged. Dr. Boik straightened out my problems associated with understanding and implementing several techniques in multivariate statistics.

Dr. Ed Mooney, I cannot thank him enough for his expertise in teaching the principles of constructing a dynamic web application to the model. His academic and personal advice over the years is greatly appreciated. I thank Dr. Paul Schillings for his encouragement and professional wisdom. Dr. Boyd, I cannot thank him enough for the hours of work he has put in editing the report. I am indebted to him for his suggestions in the oral and written presentation of this work. He has wealth of knowledge at his fingertip and always found time in his schedule to share that with me.

I would like to thank all my fellow graduate students especially Hasan and Manimay for their scholarly support. I am thankful to MSU library services for their prompt and quick attention to my needs.

Finally, I wish to dedicate this work to my graduate committee, my lovely parents and my brother. They are simply the best.

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The incidence of work-related musculoskeletal disorders (MSDs) continues to be a key concern for occupational safety and health care professionals. Several factors such as repetition, forceful exertion, and awkward postures have been linked to their development. While these links have been well established, valid and reliable techniques for measuring MSD risk are lacking, particularly for jobs in non-manufacturing industries or non-repetitive jobs in general.

Marley, et. al., (1997) examined such jobs in the power distribution industry with a goal of better understanding which work factors may be associated with MSDs. Injury data from over 2000 workers in one company were tabulated by job classification (12 total categories). Three representative categories, electric line crews, gas line crews, and meter readers were identified as having high, medium, and low risk for injury respectively, based on the recorded rate of MSDs in these categories. An ergonomic/work-methods analysis was then performed upon 5 key activities within these jobs. Activities were further broken down into 31 required tasks (e.g., climb pole, make connection, shovel, cut pipe, etc.) and even further into 18 fundamental work elements (e.g., various body postures, grasp type, force level, duration, terrain condition, etc.).

Cluster analysis involving the work element measures resulted in five clusters. Two clusters generally represented upper and lower part of the upper extremities, two clusters generally represented lower extremities and one contained miscellaneous ergonomic variables. All the coefficients of the cluster variable weights in the five clusters resulted in the same sign from principal component analysis I, signifying that the increase in values of any cluster variable in turn increases cluster score and hence the risk level. The clusters are modeled and validated using ordinal logistic regression technique. The model accurately predicted 92% and 76.5% of training and testing data sets respectively. A user-friendly web application of this model targeting the novice user has been developed.

The model needs should be trained with larger data sets for better prediction and more robust applications. However, the current model may be useful for predicting the whole-body related MSDs in the utility industry and comparable non-repetitive jobs. The identical clusters may also be useful in the understanding of physical job stress in these environments.

CHAPTER 1

INTRODUCTION

Cumulative Trauma Disorders (CTDs) is defined as physical injuries that develop over a period of time as a result of repeated biomechanical or physiological stresses on a specific body part. In short, CTDs are disorders of softer tissue due primarily to repeated use. CTDs are often considered to be work-related. Assessing the risk or determining the onset of a CTD is very difficult (Naderi and Ayoub, 1989). CTDs occur because of a single overexertion event or frequent exertion over a period of time.

Cumulative Trauma is often referred to in the literature by a number of different terms. Other terms used to describe the same condition are repetitive trauma injuries (RTI), repetitive strain injuries (RSI), musculoskeletal disorders (MSDs), occupational overuse syndrome, osteoarthroses and degenerative joint disease (Armstrong, et. al, 1986; Salter, 1970; Silverstein, et. al, 1986). CTDs are commonly reported in the tendons, and in the nerves of upper extremities, including the fingers, the wrist, the forearm and the upper arm, and the shoulder. Vern Putz-Anderson (1988) identifies three major types of disorders according to an anatomical view: tendon disorders, neurovascular disorders, and nerve disorders.

A majority of the occupational factors causing CTDs can be characterized as involving one or more of the following components: awkward postures of the wrist or shoulders, excessive manual force, and high rates of manual repetition (Putz-Anderson, 1988). It is generally accepted that force, repetition, posture, recovery time and type of grasp are important factors in the causation of distal upper extremity disorders (Moore and Garg, 1995). Some other job factors that increase risk in combination with the other factors include cold temperature, use of gloves, use of vibrating tools, etc. (Moore and Garg, 1995). Even though not studied in detail with regard to distal upper extremity disorders, duration of exposure, static muscular work, and use of the hand as a tool are also generally accepted as risk factors (Moore and Garg, 1995).

CTDs have become a prevalent form of injury in modern industry. The Bureau of Labor Statistics (BLS, 2002), US department of Labor, states that in 2000 when looking specifically at work-related musculoskeletal disorders, 66.7% (241,800) of all illness cases were due to disorders associated with repeated trauma.

Evaluation of assessment methods plays an important role in strategy to reduce and control MSDs. There are certain techniques to aid the ergonomist in understanding and identifying CTDs problem areas. They can be classified primarily into two categories: trailing and leading indicators. Trailing indicators are defined as measures that document injuries after the fact. Examples include injury rate statistics, lost time statistics, cost data, etc. Trailing indicators should be viewed as benchmark data by which system design will ultimately be judged. Trailing indicators are not, by definition, predictive. By contrast, "leading indicators" are measures that aid the ergonomist in assessing *potential* ergonomic concern. Leading indicator methodologies are useful for regular monitoring or auditing for CTDs risk. One such methodology is self-report, often used for inter and intra task comparisons. These data can be correlated with other statistical trend data. One such technique known as the "Body Map" was developed by Marley and Kumar in 1996 and has been shown to be a reliable "leading indicator" of CTD risk for the whole-body. Another well-known technique is Rapid Upper Limb Assessment (RULA), which is a survey method for the investigation of work-related upper limb disorders (McAtamney and Corlett, 1993). Both these methods take repetition into account.

These models revealed that MSD risk is likely due to some combination of force application and awkward postures. Most knowledge has been derived from examination of repetitive manufacturing or office environments and with one variable at a time constraint. Thus, valid and reliable evaluation techniques for MSD risk are lacking though some reasonable attempts have been made. This is particularly true for jobs in non-manufacturing industries or otherwise classified as non-repetitive. Thus, the main objective of this study is to develop a model for whole-body related MSD for non-repetitive jobs or otherwise known as jobs in non-manufacturing industry.

CHAPTER 2

REVIEW OF THE LITERATURE

Cumulative Trauma Disorders

This chapter is devoted to exploring the literature dealing with cumulative trauma disorders. It discusses in detail, different types of CTDs and occupational risk factors causing them. Finally, currently available statistics relating to CTDs are provided.

Cumulative Trauma Disorders (CTDs) is defined as physical injuries that develop over a period of time as a result of repeated biomechanical or physiological stresses on a specific body part. CTDs is a collective term for syndromes characterized by discomfort, impairment, disability or persistent pain in joints, muscles, tendons and other soft tissues (Kroemer, 1989). The major distinction between a CTD and sprain or strain injuries is that CTDs cannot typically be traced to a single incident, i.e., a slip or fall resulting in an acute trauma.

However, it is true that a significant stressful event may trigger diagnosis of the condition. Thus, assessing the risk or determining the onset of a CTD is very difficult (Naderi and Ayoub, 1989). It is clear that CTDs occur because of a single overexertion event or frequent exertion over a period of time. Figure 1, adapted from Chaffin and Anderson (1999) describes the spinal motion segment failure for both over exertion and frequent exertion cases in top and bottom graphs, respectively. However, the same concept of CTDs can be extended to all body areas without loss of generality.

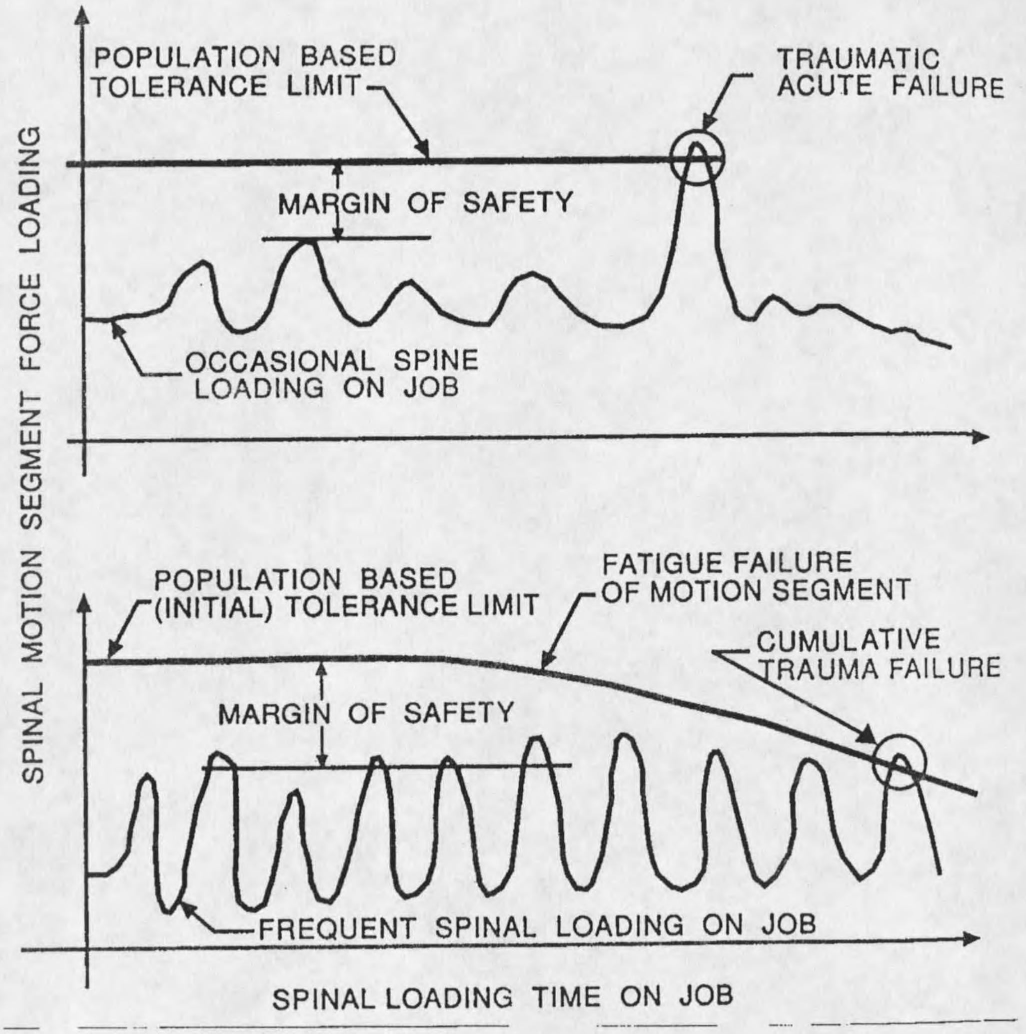


Figure 1 Depiction of spinal motion segment failure

Cumulative Trauma is often referred to in the literature by a number of different terms varying from discipline to discipline and from country to country. Other terms used to describe these disorders include repetitive trauma injuries (RTI), repetitive strain injuries (RSI), musculoskeletal disorders (MSDs), occupational overuse syndrome, osteoarthroses and degenerative joint disease (Armstrong, et. al, 1986; Salter, 1970; Silverstain, et. al, 1986). From now onwards, the author will use CTDs and MSDs interchangeably.

There are many forms of upper extremity musculoskeletal disorders, and different authors have classified them into different categories. Vern Putz-Anderson (1988) identifies three major types of disorders according to an anatomical view: tendon disorders, neurovascular disorders, and nerve disorders. Other authors classify them as alterations of the muscle-tendon unit, the peripheral nerves, or the vascular system (Grieco, et al., 1998). Muggelton, Allen and Chappell categorize upper extremity disorders as falling into one of the following three categories: vibration white finger and related dysfunctions; nerve compression disorders; and tendon and tendon-tendon related disorders (1999). Feurstein, et. al., call them nerve entrapment, tendon, or musculoskeletal-related disorders (1998). Though the terminology differs, the basic classifications are very similar. For the purpose of this report, the author has selected three classification groups: tendon disorders, vascular and neurovascular disorders, and nerve disorders.

Tendon Disorders

Tendons attach muscles to bone and transfer forces and movements from the muscles (Chaffin, Anderson, &Martin, 1999; Putz-Anderson, 1988). Tendons are surrounded sheaths of fibrous tissue in areas where friction could potentially be a problem (Chaffin, et. al., 1999). The sheath has an inner lining, the synovius, which produces synovial fluid, a lubricant that facilitates gliding of the tendon (Chaffin, et. al., 1999). The tendon glides back and forth in the sheath as the muscle contracts and relaxes. With accustomed overuse, the lubricating fluid in the tendon sheath may be lessened causing friction between the tendon and the sheath (Putz-Anderson, 1988). The tendon area then feels warm, tender and painful, signaling the onset of inflammation (Putz-Anderson, 1988). Inflammation is an immune system response by the surrounding tissue and blood vessels designed to limit bacterial invasion and initiate repair (Putz-Anderson, 1988). Swelling and sensation of the warmth occurs in the injured tissue from the inflow of blood (Putz-Anderson, 1988). Tendon disorders can include: tendonitis and tenosynovitis (Atcheson, 1988; Feurstein, et. al., 1988; Gordon, 1995; Greico, et. al., 1998; Fernandez & Marley, 1988, Muggleton, et. al., 1999; Putz-Anderson, 1988), as well as bursitis, and ganglionic cysts (Fernandez & Marley, 1988, Fernandez & Marley, 1988).

Tendinitis and Tenosynovitis. Tendinitis refers to tendon inflammation specifically, whereas tensosynovitis is a general term describing injury involving the tendon sheath (Muggleton, et. al., 1999). Ranney further defines the two as tendonitis being inflammation as a result of microtears, and tenosynovitis as inflammation resulting from friction (1993). These conditions are most commonly found in the flexor and extensor tendons of the wrists and thumbs, the extensor tendons of the elbow, and the rotator cuff and biceps tendons of the shoulders (Herrington & Morse, 1995). It is most likely to occur in areas where the tendon is restricted by anatomical feature (i.e., bony channels and tunnels) (Fernandez & Marley, 1998). This form of tendon inflammation occurs when a muscle/tendon unit is repeatedly tensed, then with further exertion tendon fibers may fray or tear apart (Gordon, 1995; Putz-Anderson, 1988). If this happens, the tendon becomes thickened, bumpy and irregular (Putz-Anderson, 1988). The repetitiveness of the task, the force required, and the position of the joint are all factors in the pathogenesis of this problem (Gordon, 1995). Since tendons have virtually no blood supply, they are not capable to repair themselves, thus damage can become instrumental (Pecina & Bojanic, 1993), and without rest and sufficient time for the tissues to heal, the tendon may be permanently weakened (Putz-Anderson, 1988).

Tenosynovitis is a general term for a repetitive-induced tendon injury, which involves the synovial sheath (Putz-Anderson, 1988). With extreme repetition, the sheath will produce unnecessarily large amounts of synovial fluid that accumulates and causes the sheath to be swollen and painful (Putz-Anderson, 1988), resulting in an inflammatory reaction within the tendon sheath (Fernandez & Marley, 1988).

Stenosing tenosynovitis is another type of tensosynovitis, that may be diagnosed if the tendon becomes irritated and rough, and if the sheath becomes inflamed and presses on the tendon (Muggleton, et. al., 1999; Putz-Anderson, 1988). DeQuervian's disease is the most recognized stenosing tenosynovitis. It is a disorder that affects the tendons on the side of the wrist and at the base of the thumb (Muggleton, et. al., 1999; Putz-Anderson, 1988). These tendons are connected to muscles on the back of the forearm and contract to pull the thumb back and away from the hand (Putz-Anderson, 1988). De Quervian's disease is attributed to excessive friction between two thumb tendons and their common sheath (Putz-Anderson, 1988).

If the tendon sheath of a finger becomes exceedingly swollen, it can cause the tendon to get locked in the sheath, then attempts to move the finger result in snapping and jerking movements, called stenosing tenosynovitis crepitans or "trigger finger" (Muggleton, et. al., 1999; Putz-Anderson, 1988). In later stages of the disease, snapping ceases and the finger remains permanently locked (Muggleton, et. al., 1999). The palm side of the fingers is the usual site for trigger finger. This disorder is often associated with using tools that have handles with hard or sharp edges (Putz-Anderson, 1988).

<u>Bursitis.</u> Bursae are anti-friction devices found throughout the body where bony prominences are close to the skin surface and friction from outside the body or where tendons and ligaments may rub against the prominences (Rowe, 1985). In the presence of high degrees of friction, the bursae will oversecrete lubricating fluids and bursal sacs will become enlarged and distended. If friction persists, the walls of the sac will thicken and become inflamed (Fernandez & Marley, 1998).

<u>Ganglionic Cyst.</u> Caused by the swelling of a tendon sheath with synovial fluid, a ganglionic cyst is common and is generally related to wrist usage (Birnbaum, 1986). Though rarely causing symptoms of nerve compression, such a cyst can often be painful

and is usually treated by aspiration or by surgical removal if the ganglion recurs (Fernandez & Marley, 1998).

Neurovascular Disorders

Neurovascular disorders are those CTDs which involve both the nerve and adjacent blood vessels

Thoracic Outlet Syndrome. Probably the most common form of neurovascular disorder is the thoracic outlet syndrome (Putz-Anderson, 1988). Thoracic outlet syndrome is a general term for compression of the nerves and blood vessels as they pass through the neurovascular bundle between the neck and shoulder.

Also known as cervicobrachial disorder, thoracic outlet syndrome is generally thought to result from heavy workloads combined with repetitive straining or unnatural static positioning of the arms (Sallstorm and Schimdt, 1984). Typical symptoms of thoracic outlet syndrome include numbness and tingling in the fingers and hand as well as a sensation of the arm "going to sleep." The blood pulse at the wrist may also become weakened.

Vibration Syndrome. Sometimes referred to as vibration induced white finger, Raynaud's syndrome, or traumatic vasospastic disease, vibration syndrome is characterized by episodes of blanching (whiteness or paleness) of the fingers due to closure of the digital arteries (Putz-Anderson, 1988). Due to the blockage of circulation in the fingers, coldness and pain is often associated with vibration syndrome (Taylor, 1974). This condition is caused by the transmission of vibration (varying in acceleration, power, and frequency) from a tool to the hand. It is believed to be in part a vascular disturbance due to changes in the blood vessel walls and in part a nervous disturbance caused by reflex contraction of the smooth muscles of the blood vessels.

Nerve Entrapment Disorders

Carpal Tunnel Syndrome. Carpal tunnel syndrome (CTS) is one of the major forms of cumulative trauma disorders of the upper extremities (Putz-Anderson, 1988). Also described as occupational neuritis, partial thenar atrophy and median neuritis, CTS is generally attributed to insult, usually compression, to the median nerve within the wrist as it passes through the carpal tunnel (Armstrong and Chaffin, 1979a). This compression in turn is associated with repeated or sustained activities of the fingers and hands, often combined with the application of force, as well as pressure from hard work surfaces and sharp edges on hand tools (Feldman, et. al, 1983).

Occupational Risk Factors Causing CTDs

CTDs are often considered to be work-related. Majority of the occupational factors causing CTDs can be characterized as involving one or more of the following components: awkward postures of the wrist or shoulders, excessive manual force, and high rates of manual repetition (Putz-Anderson, 1988). It is generally accepted that force, repetition, posture, recovery time and type of grasp are important factors in the causation of distal upper extremity disorders (Moore and Garg, 1995). In addition to these factors, other job factors that combine to increase risk include cold temperature, use of gloves, use of vibrating tools, etc. (Moore and Garg, 1995). Even though not studied in detail with regard to distal upper extremity disorders, duration of exposure, static muscular work, and use of the hand as a tool are also generally accepted as risk factors (Moore and Garg, 1995). Risk factors posture, force and repetition are discussed in detail in the following sections.

Posture

Certain jobs require the worker to assume a variety of awkward postures that pose significant biomechanical stress to the joints of the upper extremity and surrounding soft tissues. Awkward postures include any fixed or constrained body position. Other undesirable postures include those that overload the muscles and tendons, load joints in an uneven or asymmetrical manner, or involve a static load on the musculature (Putz-Anderson, 1988).

Force

The force required to perform various occupational activities is also a critical factor in contributing to the onset of CTDs. As the muscle effort increases in response to high task load, circulation to the muscle decreases causing more rapid muscle fatigue. When force requirements are high, recovery time can exceed actual work time. Deprived of sufficient recovery time, soft tissue injury will occur. Bones will break and skin and muscles will tear if the strain is too great. The mechanical stresses on the tendons and nerves produced by contact with sharp edges of hard objects are not quite obvious (Putz-Anderson, 1988).

Repetition

In general, a job is considered repetitive if the basic (fundamental) cycle time is less than 30 seconds or 50% (or more) of total cycle time performing the same fundamental task element (Konz and Johnson, 2000;Fernandez and Marley, 1998). These are the two generally accepted definitions of repetition.

Jobs that require the worker to perform highly repetitive motions also contribute to the onset of CTDs. The more repetitive the task, the more rapid and frequent are the muscle contractions. Muscles required to contract at a high velocity develop less tension than when contracting at a slower velocity for the same load. Hence, tasks requiring high rates of repetition require more muscle effort, and consequently more time for recovery, than less repetitive tasks. So, tasks with high repetition rates can become sources of trauma even when the required forces are minimal and normally safe (Putz-Anderson, 1988).

CTD Statistics

CTDs have become a prevalent form of injury in modern industry. The Bureau of Labor Statistics (BLS, 2002), US department of Labor, provides the following summaries related to CTD's. When looking specifically at work-related musculoskeletal disorders, BLS reports that in 2000, 66.7% (241,800) of all illness cases were due to disorders associated with repeated trauma. This figure does not include back injuries. BLS also reports that recently the number of cases of repeated trauma has decreased considerably, lowering from 308,200 cases in 1995 to 216,400 cases in 2001—a 2.98% decrease as shown in Figure 2.

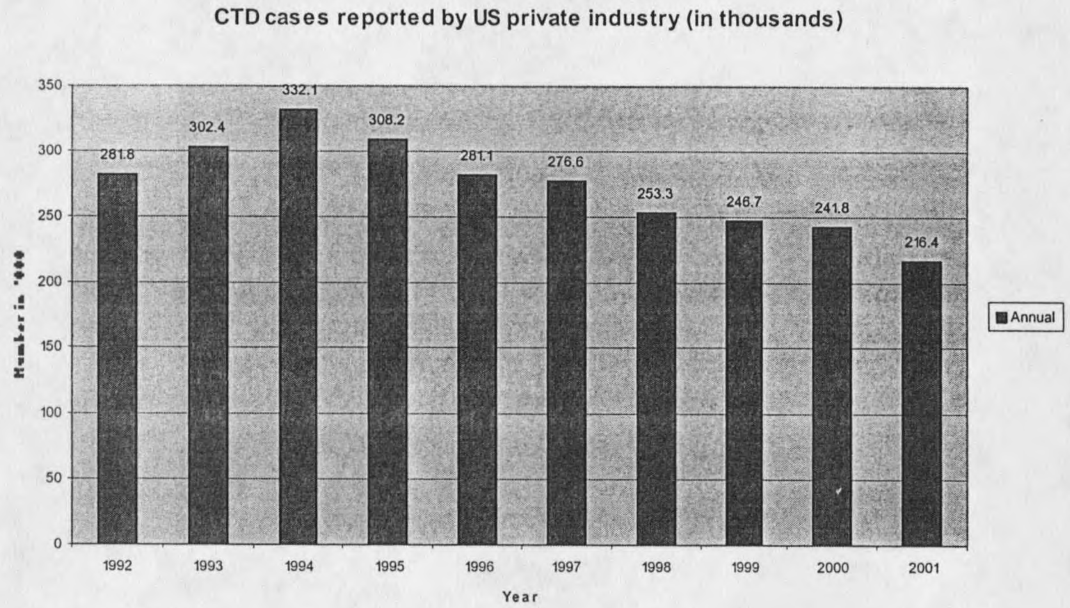


Figure 2 Number of CTDs reported in the US from 1992 to 2001

When looking specifically at cases involving days away from work, for which more detailed information is available, BLS reports that in 2000, approximately 32% or 523,043 cases were the result of overexertion or repetitive motion. This figure includes back injuries. Out of the repetitive trauma cases, 33% of injuries were due to manual lifting primarily affecting the back and over 32% resulted from hand tool use, data entry and repetitive grasping tasks. Cost estimates in the US vary from $13 to $20 billion annually (NIOSH 1996). In 1993, Webster and Snook estimated mean per case cost of compensable low-back pain at $8,321 and mean per case cost of compensable upper-extremity CTDs at $8,070. Recent updates of these costs are currently not available but are believed to have risen substantially since 1993.

CHAPTER 3

OBJECTIVES

It has been shown that the number of musculoskeletal disorders (MSDs) has increased dramatically in recent years and has become a key concern for occupational safety and health care professionals. CTDs is also an ever increasing cost to business and industry in terms of reduced productivity, lost work time, high insurance and disability claims. Several factors such as repetition, forceful exertion and awkward postures have been linked to the development of work-related MSDs as discussed in Chapter 1. Thus, risk of MSDs is a critical concern for ergonomists.

Evaluation of assessment methods plays an important role in strategy to reduce and control MSDs. There are certain techniques to aid the ergonomist in understanding and identifying CTDs problem areas. The ergonomist should have access to certain "trailing indicators", which are defined as measures that document injuries after the fact. Examples include injury rate statistics, lost time statistics, cost data, etc. Trailing indicators should be viewed as benchmark data by which system design will ultimately be judged. Such indicators should also be analyzed thoroughly to look for undesired trends, or (hopefully) to verify that ergonomic changes are having the desired effect. Trailing indicators are not, by definition, predictive. By contrast, "leading indicators" are measures that aid the ergonomist in assessing *potential* ergonomic concern. Leading indicator methodologies are useful for regular monitoring or auditing for CTDs risk. One such methodology is self-report, often used for inter and intra-task comparisons. These data can be correlated with other statistical trend data. One such technique known as the "Body Map" was developed by Marley and Kumar in 1996 and has been shown to be a reliable "leading indicator" of CTD risk for the whole-body. Another well-known technique is Rapid Upper Limb Assessment (RULA), which is a survey method for the investigation of work-related upper limb disorders (McAtamney and Corlett, 1993). Both these methods take repetition into account.

From the above models it can be inferred that, MSD risk is likely due to some combination of force application and awkward postures. Most knowledge has been derived from examination of repetitive manufacturing or office environments and with one variable at a time constraint.

Thus, valid and reliable evaluation techniques for MSD risk are lacking though some reasonable attempts have been made. This is particularly true for jobs in non-manufacturing industries or otherwise classified as non-repetitive. The activities within these jobs are varied with long cycle times. "Field crews" in utilities, for example. Further, many tasks in these activities may contain more than one known risk factor.

A method to examine MSD risk in non-repetitive, whole-body work is needed. Therefore a study was conducted to achieve the following.

1. Find the natural groupings of whole-body related musculoskeletal variables associated with CTDs using cluster analysis and further interpret the clusters.

2. Evaluate and interpret the cluster variable weights using principal component analysis I. This is comparable to the approach of Moore and Garg (1995) but not limited to upper extremity only.

3. Model and validate the clusters using ordinal logistic regression, linear discriminant analysis and nearest neighbor rule and identify the significant clusters.

4. Develop a web application to the ordinal logistic regression model using C, MYSQL, and PHP.

CHAPTER 4

METHODS

This chapter describes the methods adopted by Marley, et al., 1997 for data collection. Marley, et al., 1997 previously examined jobs in the power distribution industry with a goal of better understanding which factors may be associated with MSDs related to outdoor activities that are not repetitive in nature and less frequently performed. Injury data from over 2000 workers in one public utility company in the state of Montana were recorded by job classification (12 categories) from 1990 to 1995 as shown below in Table 1.

Table 1 Injury Data by Job Classification (1990-1995)

Job categories	Number of sub-categories	Total number of MSD's	Total number of injured workers
Line Worker	9	80	308
Mechanic	10	48	145
Office Personal	0	30	1115
Gas Trade	11	23	152
Technician	0	22	162
Operator	0	20	219
Utility Man	0	16	114
Warehouse	3	11	56
Hydro	5	10	47
Janitor/ Janitress	0	8	33
Maintenance	4	8	52
Meter Reader	0	7	79

Injury data details for major and sub-categories of these jobs with respect to the extremity affected are presented in Appendix A.

Data from utility company was examined for MSD injuries such as sprains, strains, low-back, CTS, tendonitis, bursitis, inflammation/irritation of joints, tendons and muscles. The bar chart in Figure 3 is arranged to illustrate 'number of MSDs' versus 'job classification'. The chart shows that the number of MSDs was highest for electric line crews, lowest for meter readers and nearly between was gas trade.

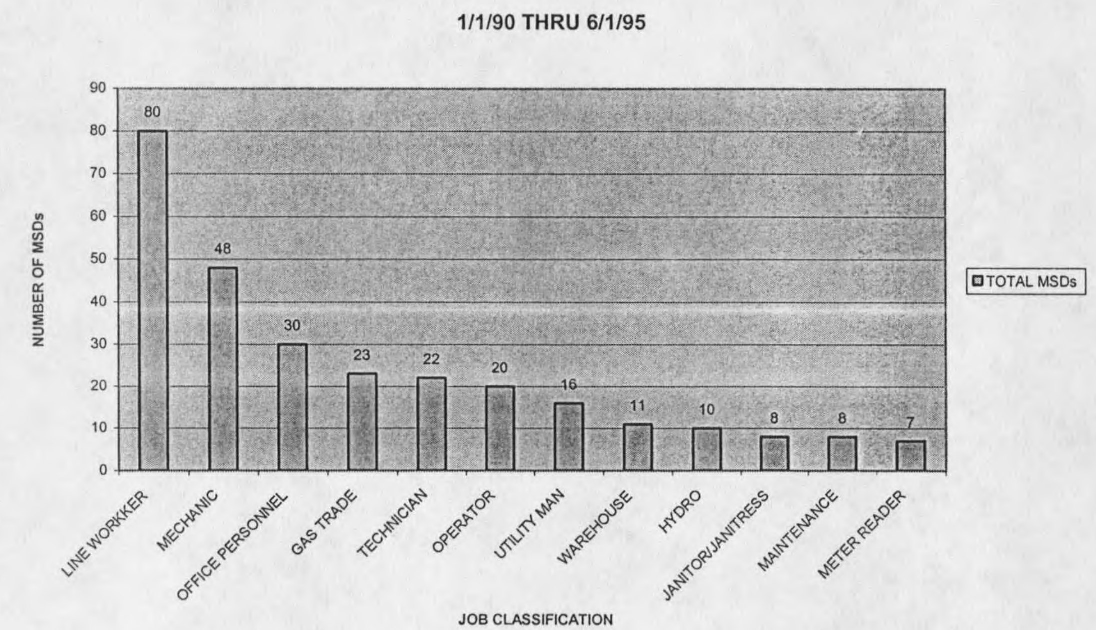


Figure 3 Number of MSD's by job classification

Figure 4 illustrates that the number of MSD's per 100 employees was highest for mechanic and lowest for office personnel and nearly in between was for gas trade. However, line worker and meter reader were chosen in place of mechanic and office personnel for analyzing high and low risk injury category. One of the strategic reason is both line worker and meter reader are the jobs that are performed outdoors. Most of the jobs in a non-repetitive (non-manufacturing) environment are performed outdoors. After observing these jobs it can also be identified that they have whole body related

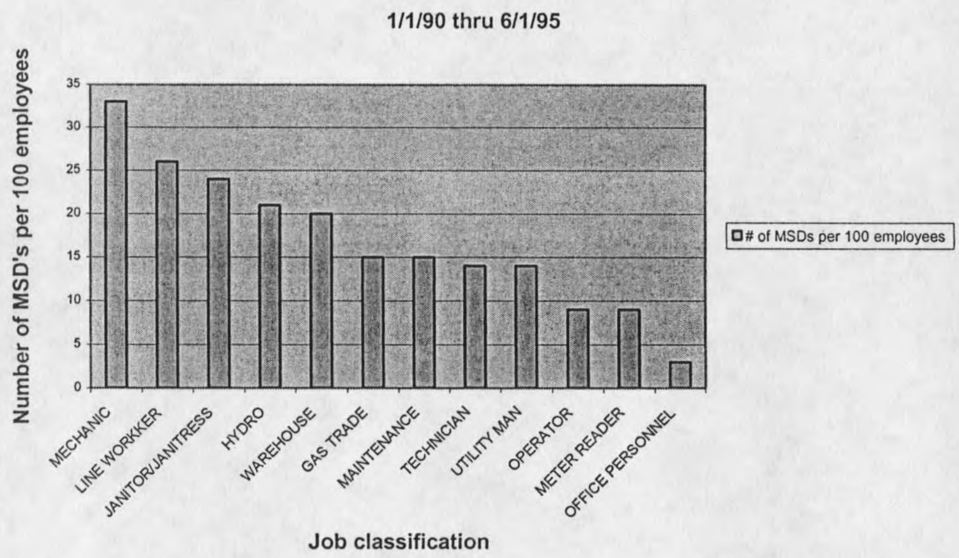


Figure 4 Number of MSD's per 100 employees by job classification

movements in their activities. These two jobs are thus chosen in addition to gas trade for developing a generalized model for whole-body related musculoskeletal disorders.

An ergonomic/work-methods analysis was then performed upon five key activities within these jobs as listed in Table 2. Activities were further broken down into 31 required tasks (e.g., climb pole, make connection, shovel, cut pipe, etc.) as listed in Table 3.

Table 2 Five key activities

Serial Number	Activity
1	Gasline work
2	Setting meters
3	Reading meters
4	Overhead
5	Underground service

All the tasks listed in Table 3 were video taped and analyzed to find the associated fundamental ergonomic variables. For a given task, the video tape was divided into different smaller fragments and analyzed in slow motion. Analysis of the videotape provided detailed information on the body positions required to perform each key activity as well as information on forces, terrain, and exertion duration for a particular task. Fifteen fundamental work elements thus found were quantified into 59 levels that represent the variables for further analysis. The fundamental work elements and their different levels of measurement are documented in Table 4. Data collection sheet used by Marley, et. al., (1997) is attached in Appendix B.

Table 3 Tasks

Serial Number	Task
1	Connect pipe
2	Adjust valves
3	Test flow meters
4	Adjust outside valves
5	Change seals
6	Cut pipe
7	Debur
8	Thread
9	Pipe wrench
10	Set meter
11	Electric meter reading
12	Gas meter reading
13	Work from pole
14	Climbing
15	Work from bucket
16	Hotstick from bucket
17	Operator
18	Saw pole and dispose
19	Open junction can
20	Hotstick junction can
21	Tamping
22	Pulling wire
23	Prep wire
24	Ground wire (hotstick)
25	Shoveling
26	Conduit
27	Lay wire
28	Hook up transformer
29	Hook up meter
30	Ground rod
31	Cover wire

Table 4 Fundamental work elements and their levels

Variable No	Work Element	Levels (measured in degrees) 1	2	3	4	5
1	Upper arms	+/-20	-20	20-45	45-90	>90
2	Lower arms	0-60	60-100	100+		
3	Wrist (Ulnar)	0-20	>20			
4	Wrist (Radial)	0-20	>20			
5	Wrist flexion	0-20	>20			
6	Wrist extension	0-20	>20			
7	Neck	+/-0-10	+/-0-20	+/-20+		
8	Trunk twist	0-30	30-45	45-90	90+	
9	Trunk (Flex/Ext)	0-15	15-30	30-45	45-90	
10	Exertion Duration	In minutes				
11	Legs					
	Standing	Not present	Present			
	Sitting	Not present	Present			
	Kneeling	Not present	Present			
12	Terrain					
	Poor terrain	Not present	Present			
	Fair terrain	Not present	Present			
	Good terrain	Not present	Present			
13	Gloves					
	No glove	Not present	Present			
	Light glove	Not present	Present			
	Heavy glove	Not present	Present			
14	Force					
	Low force	Not present	Present			
	Medium force	Not present	Present			
	High force	Not present	Present			
15	Grip					
	Power grip	Not present	Present			
	Chuck grip	Not present	Present			
	Pencil grip	Not present	Present			
	Key grip	Not present	Present			

Different measurement levels of upper arms, lower arms, wrist, neck, trunk and legs are set based on Rapid Upper Limb Assessment (RULA) tool (McAtamney & Corlett, 1993) as shown in Figure 5 (adapted from (McAtamney & Corlett, 1993) and

Figure 6 (adapted from McAtamney& Corlett, 1993) respectively. However, the metrics were slightly changed as stated in Table 4. For example, the angle for upper arms was measured using a goniometer after pausing the video tape at the point where subject shows maximum upper arm deviation. The same procedure was adopted for other variables also.

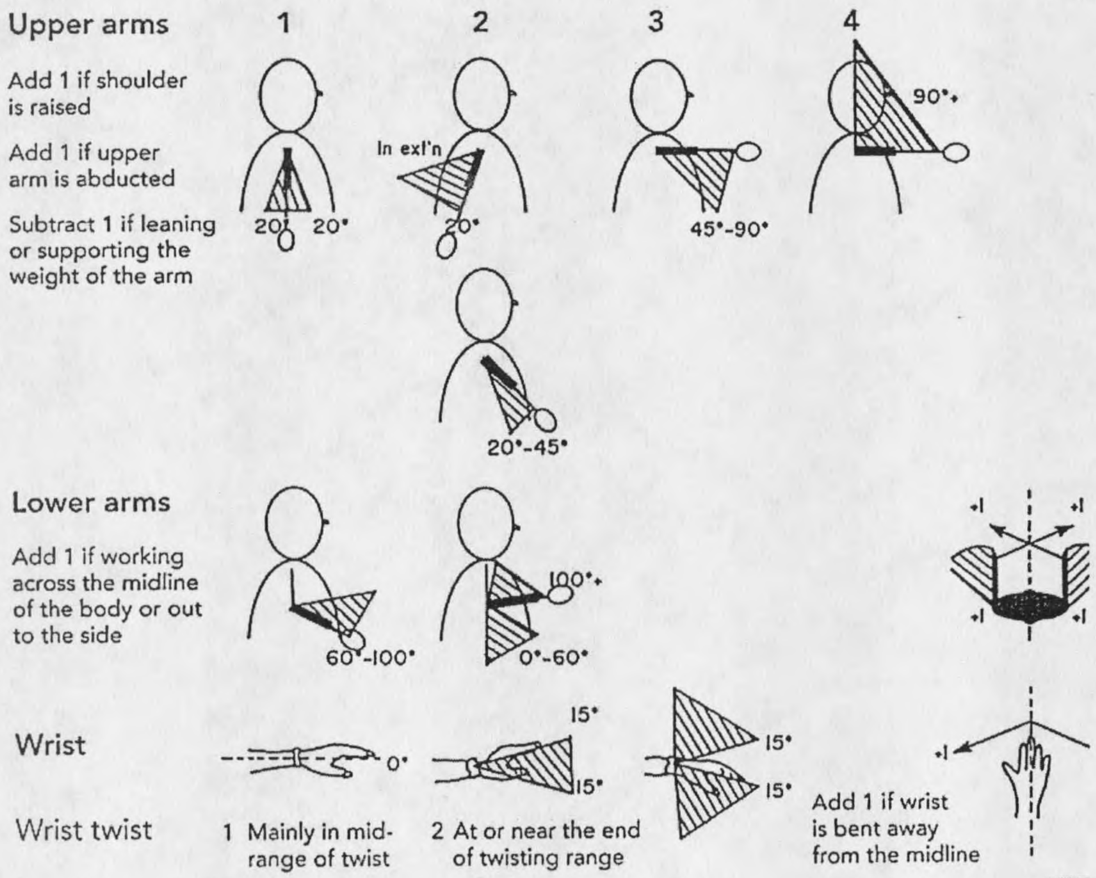


Figure 5 Posture scores for upper arm, lower arm and wrist

Different positions of the wrist and different handgrips are shown in Figures 7 and 8 (adapted from Putz-Anderson, 1988), respectively.

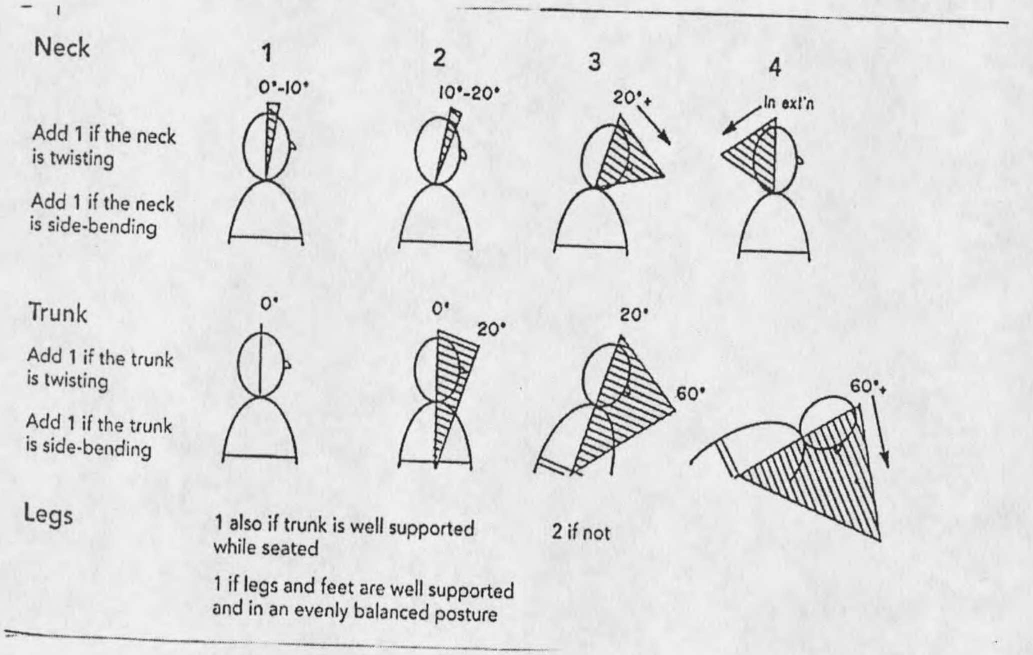


Figure 6 Posture scores for neck and trunk

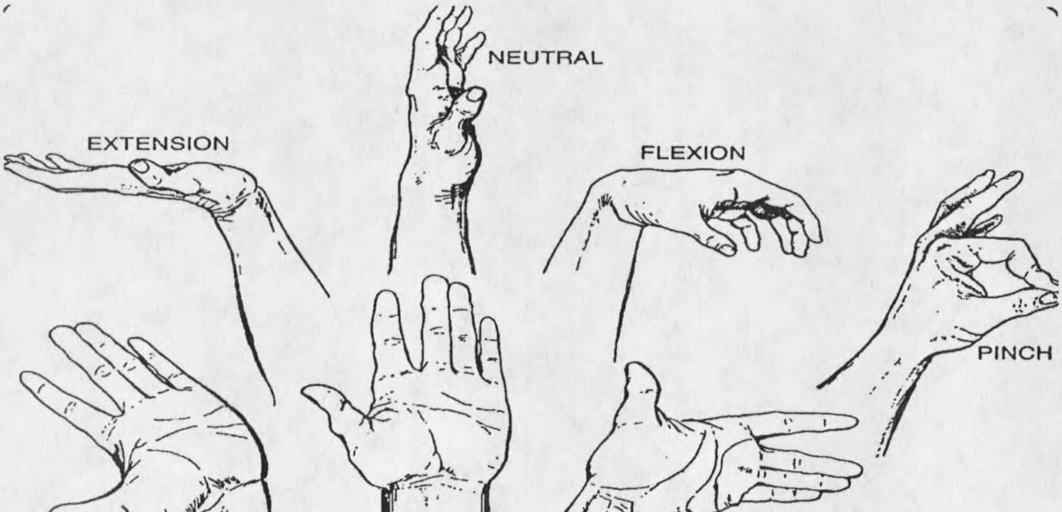


Figure 7 Wrist positions

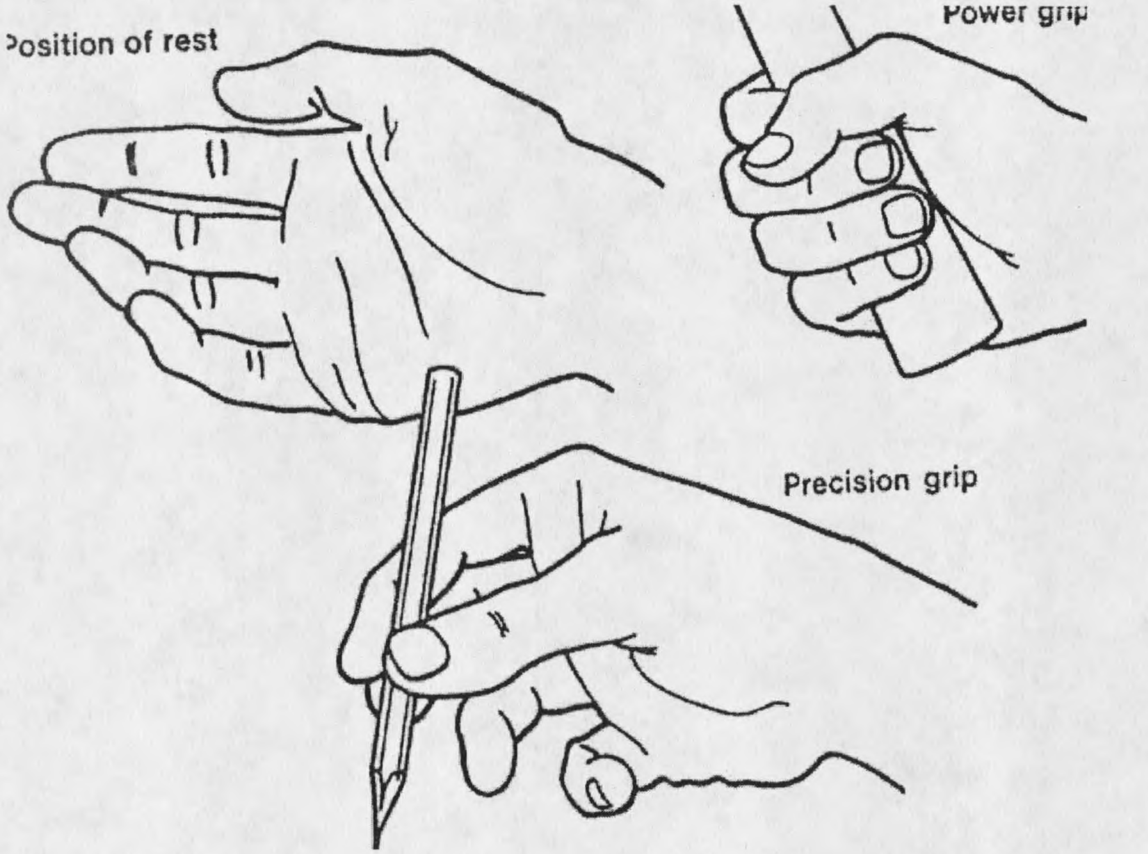


Figure 8 Handgrips

Data were collected on 67 different tasks performed within 5 different activities. The job, activity, and task combination for all 67 observations are shown in Table 5. The job classification numbers 1, 2, and 3 correspond to gas trade, meter reader, and line worker, respectively. Activity numbers are as shown in Table 2. The original data are shown in Appendix B.

Table 5 Utility company data

Observation No	Job class	Activity	Task
1	1	1	1
2	1	1	2
3	1	1	3
4	1	1	4
5	1	1	2
6	1	2	6
7	1	2	7
8	1	2	8
9	1	2	9
10	1	2	10
11	1	2	1
12	1	2	6
13	1	2	7
14	1	2	8
15	1	2	9
16	2	3	11
17	2	3	12
18	2	3	11
19	2	3	12
20	2	3	11
21	2	3	12
22	3	4	13
23	3	4	14
24	3	4	16
25	3	4	16
26	3	4	16
27	3	4	16
28	3	4	13
29	3	4	14
30	3	4	15
31	3	4	14
32	3	4	13
33	3	4	18
34	3	5	19
35	3	5	20
36	3	5	22
37	3	5	17
38	3	5	23
39	3	5	17
40	3	5	25
41	3	5	26

Table 5 Utility company data – Continued

42	3	5	27
43	3	5	28
44	3	5	29
45	3	5	30
46	3	5	17
47	3	5	25
48	3	5	28
49	3	5	30
50	3	5	23
51	3	5	17
52	3	5	25
53	3	5	29
54	3	5	27
55	3	5	23
56	3	4	14
57	3	4	13
58	3	4	31
59	3	4	28
60	1	2	6
61	1	2	7
62	1	2	8
63	1	2	9
64	1	2	6
65	1	2	7
66	1	2	8
67	1	2	9

CHAPTER 5

RESULTS AND DISCUSSION

Variable Reduction and Data Standardization

This section describes the procedure adopted in reducing the number of variables and standardizing the data from Marley, et. al., (1997).

Marley, et. al, (1997) designated each level of variable numbers 1, 2, and 7 to 9 in Table 4 as a separate binary variable with values '0' representing 'No' and '1' representing 'Yes' to injury claim, for the data collection and analysis (See Appendix B for original data). For variable numbers 3 to 6 in Table 4, levels 1 and 2 were designated as '1' and '2' (See Appendix B for original data) also for data collection and analysis purposes. Duration was considered as a continuous variable measured in minutes. Variables 11 to 15 in Table 4 are binary with values 0 and 1, respectively. Thus, the variable classification method resulted in 40 variables for the analysis. Using cluster analysis, Marley, et. al., (1997) obtained ten clusters as documented in Appendix B.

Examination of the original data (Appendix B) revealed that values for variables 1, 2 and 7 to 9 in Table 4 for a given observation fall into only one of their defined levels of measurement. Hence, each level of measurement in these variables is not a different variable but just a level of a main (ordinal) variable. For example, +/-20 is level one of ordinal variable, upper arms. It was not efficient to consider it as a separate binary variable. Similarly, the sub-category values of variables 11 to 15 were found to be mutually exclusive. For example, main variable (terrain) can be classified into only one of three sub-categories (poor terrain, fair terrain, and good terrain) and therefore it was inefficient to consider each sub-category as a separate binary variable. Hence, these sub-categories have been changed to different levels of their main variable. Power grip was found to be present in all 67 observations such as wrist flexion (>20 deg). Hence, these variables are constants and eliminated from cluster analysis. However, they were still accounted as separate variables. Duration is the only continuous variable with larger domain and all other variables are binary in the original data set of Marley, et al., 1997 (See Appendix B). It is easy to obtain a binary variable from interval-scaled measurements by cutting the measurement axis into two (Kaufman & Rousseeuw, 1990). The range of values for 'duration' were cut into two parts, less than or equal to 30 seconds, and greater than 30 seconds, using 30 seconds as threshold following the definition of repetition discussed in Chapter 1. Thus, duration is transformed into a binary variable with the above two levels representing 0 and 1, respectively. Revised work elements and their new levels of measurement are shown in Table 6.

The data set now has a collection of mixed variables. Some are ordinal variables while the others are binary variables. Data with mixed variables can be treated in several ways. It is more practical to process the data together and then perform a single cluster analysis (Kaufman & Rousseeuw, 1990). For instance, one can treat all variables as if they were interval-scaled. This is quite appropriate for symmetric binary variables, for the ranks originating from ordinal variables, and for the logarithms of ratio variables (Kaufman & Rousseeuw, 1990).

From Table 6, it can be observed that ordinal variables (1, 2, 6 7, 8, 13, 14, 15) have a different number of levels ranging from 2 to 5. All binary variables (3, 4, 5, 9, 10-12, 16-18) have 0 and 1 as their level values.

Table 6 Revised work elements and their new levels

Variable No	Work Element	Levels (Measured in degrees) 0	1	2	3	4
1	Upper arms	+/-20	-20	20-45	45-90	> 90
2	Lower arms	0-60	60-100	100+		
3	Wrist (Ulnar)	0-20	>20			
4	Wrist (Radial)	0-20	>20			
5	Wrist extension	0-20	>20			
6	Neck	+/-0-10	+/-0-20	+/-20+		
7	Trunk twist	0-30	30-45	45-90	90+	
8	Trunk (Flex/Ext)	0-15	15-30	30-45	45-90	
9	Exertion Duration	<=30secs	>30secs			
	Legs					
10	Standing	Not present	Present			
11	Sitting	Not present	Present			
12	Kneeling	Not present	Present			
13	Terrain	Good terrain	Fair terrain	Poor terrain		
14	Gloves	No glove	Light glove	Heavy glove		
15	Force	Low force	Med force	High force		
	Grip					
16	Chuck grip	Not present	Present			
17	Pencil grip	Not present	Present			
18	Key grip	Not present	Present			

Thus, the ordinal variables under study possess different levels and it is useful to convert all variables under study to the 0-1 range in order to achieve equal weighting of the variables (Kaufman & Rousseeuw, 1990). The standardized score for a given level r, is determined by the following formula (Kaufman & Rousseeuw, 1990):

Z = (r-1)/(M-1), Where 1,2,3...M are the previous levels of the ordinal variable. 'M', refers to the maximum ordinal level. Z values for different levels of an ordinal variable are shown in Table 7. Modified data are shown in Appendix C. The job classification numbers 1, 2, and 3 correspond to meter readers, gas trade, and line worker respectively.

Table 7 Standardized scores for different levels of an ordinal variable

M (Maximum level)	Levels				
	0	1	2	3	4
5	0	0.25	0.5	0.75	1
4	0	0.33	0.67	1	
3	0	0.5	1		
2	0	1			

Grouping Data

Searching the data for a structure of "natural" groupings is an important exploratory technique. Groupings can provide an informal means for assessing dimensionality, identifying outliers, and suggesting interesting hypotheses concerning relationships (Johnson and Wichern, 2002). This section discusses the first objective of finding natural groupings of whole-body related musculoskeletal variables associated with CTDs from Table 6 using cluster analysis.

Cluster Analysis

Cluster analysis is a more primitive technique in that no assumptions are made concerning the number of groups or group structure. Grouping is done on the basis of similarities or dissimilarities (distances like Euclidean, Minkowski metric, Canberra metric, etc). The inputs required are similarity measures or data from which similarities are computed.

Linkage Method. Hierarchical clustering techniques proceed by either a series of successive mergers or successive divisions of variables. Agglomerative hierarchical methods start with the individual objects. Thus, there are initially as many clusters as objects. The most similar objects are first grouped, and these initial groups are merged according to their similarities. Eventually, as similarity decreases, all subgroups are fused into a single cluster. The results of agglomerative methods are displayed in the form of a two-dimensional diagram known as dendogram. The dendogram illustrates the mergers or divisions made at successive levels. Linkage methods are particular agglomerative hierarchical procedures. They are suitable for clustering items as well as variables. However, this is not true for all hierarchical agglomerative procedures.

Average linkage (average distance) method is used in the analysis. This method treats the distance between two clusters as the average distance between all pairs of items/variables where one member of a pair belongs to each cluster. The input to the average linkage algorithm may be distances or similarities, and the method can be used to group objects or variables. The average linkage algorithm proceeds in the same manner of the general (hierarchical clustering methods) algorithm (Johnson and Wichern, 2002) as documented in Appendix C. Initially, the distance matrix $D = \{d_{ik}\}$ is used to find the nearest (most similar) objects – for example, U and V (where d_{ik} is the distance between object i in the cluster U and object k in the cluster V). The Pearson product-moment correlation coefficient is used as distance metric in the analysis. It is discussed in detail in the next section. These objects are merged to form the cluster (UV). For step 3 of the general agglomerative algorithm, the distances between (UV) and the other cluster W are determined by

$$D_{(UV)w} = \Sigma\Sigma\, d_{ik} / (N_{(UV)} N_W)$$

Where d_{ik} is the distance between object i in the cluster (UV) and object k in the cluster W, and $N_{(UV)}$ and N_W are the number of items in clusters (UV) and W, respectively (Johnson and Wichern, 2002).

Similarity and Distance Measure. There are many ways to measure the similarity between pairs of objects. Most practitioners use distances (Euclidean, Minkowski metric, Canberra metric etc) to cluster items and correlations (Pearson's product moment, Spearman rank order, Kendall's Tau) to cluster variables.

Pearson's product moment correlation coefficient was used as the distance measure. This measure of linear association between two variables does not depend on the units of measurement. The sample correlation coefficient for the i^{th} and k^{th} variables is defined as

$$r_{ik} = s_{ik}/(\sqrt{s_{ii}}\sqrt{s_{kk}}) = \frac{\sum_{j=1}^{n}(x_{ji}-\overline{x_i})(x_{ji}-\overline{x_k})}{\sqrt{\sum_{j=1}^{n}(x_{ji}-\overline{x_i})^2}\sqrt{\sum_{j=1}^{n}(x_{ji}-\overline{x_k})^2}}$$

Clusters. The modified data (67 observation points) was divided randomly into two sets: training (80% of the modified data, 50 observations) and testing (20% of the modified data, 17 observations), as shown in Appendix C. For cluster analysis, all the data were used. The MINITAB statistical package (MINITAB, 2000) was used for cluster analysis. The average linkage method was applied to determine the distance between two clusters. 1- Pearson product moment correlation coefficient was used as the distance measure. The similarity, s(ij), between two clusters i and j is given by s(ij) = 100(1-d(ij))/ d(max) where d(max) = 2 (MINITAB, 2000).

Amalgamation steps and the dendogram obtained from MINITAB are listed in Appendix D. The final grouping of clusters (also called the final partition) identifies groups whose observations share common characteristics. The decision about final grouping is also called cutting the dendogram. The complete dendogram (tree diagram) is a graphical depiction of the amalgamation of observations into one cluster. Cutting the dendogram is akin to drawing a line across the dendogram to specify the final grouping. In order to discover the cutting point of the dendogram, cluster analysis was executed without specifying a final partition. The similarity and distance levels were examined in the MINITAB session window results and in the dendogram. The similarity level at any step was the percent of the minimum distance at that step relative to the maximum inter-observation distance in the data. The pattern of how similarity or distance values change from step to step helps in choosing the final grouping. The step that reveals an abrupt change in values may identify a suitable point for cutting the dendogram, if this makes sense for the data (MINITAB, 2000). Thus, the cutting point was chosen at a similarity level of 54.67, where the difference between the previous similarity value is 4.6 which abruptly changed from 2.0 (previous pair difference) (See Appendix D). Five clusters obtained at this break point are listed in Table 8. Clusters 1 and 2 generally represented the upper and lower part of upper extremities, respectively. Clusters 4 and 5 generally represented lower extremities of the human body. Cluster 3 had miscellaneous ergonomic variables not specific to any one section of the human body.

Table 8 Final clusters

Cluster number	Cluster variables
1	Upper arms, trunk twist, trunk flexion, gloves, terrain
2	Lower arms, wrist ulnar, wrist radial, wrist extension, kneeling, chuck grip, key grip
3	Neck, duration, pencil grip
4	Standing, force
5	Sitting

Cluster Scores. Principal component analysis is widely used for explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables. Its main objectives are data reduction and interpretation (Johnson and Wichern, 2002). This section discusses the second objective of finding and interpreting cluster variable weights of the clusters listed in Table 8 using Principal Component

Analysis I (PCA I). Cluster score is the sum of the product of cluster variable weights and their values.

Principal components are algebraically particular linear combinations of the p random variables X_1, X_2, ...,X_p. These linear combinations geometrically represent the selection of a new coordinate system obtained by rotating the original system with X_1, X_2, ...,X_p as the coordinate axes. The new axes represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance matrix. Principal components solely depend on either the covariance matrix or the correlation matrix of X_1, X_2, ...,X_p and their development does not require a multivariate normal assumption (Johnson and Wichern, 2002). Initially, all the variables are standardized and their principal components are calculated as shown below. The first principal components represent the uncorrelated linear combination with maximum variance.

$$Z_i = (x_i - \mu_i)/\ \sigma_{ii}, \quad \text{where } i = 1,2,\ldots,p.$$

In matrix notation, $\mathbf{Z} = (\sqrt{D})^{-1}(\mathbf{X}\text{-}\mu)$

where $E(\mathbf{Z}) = 0$ and $\mathrm{Cov}(\mathbf{Z}) = \rho$

$$D = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots\ldots, \sigma_n^2)$$

The i^{th} principal component of the standardized variables $\mathbf{Z}' = [Z_1 \ldots.. Z_P]$ with $\mathrm{Cov}(\mathbf{Z}) = \rho$ is given by

$$Y_i = e'_i\mathbf{Z}$$

Moreover, $$\sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} Var(Z_i) = p$$

and $$\rho_{Y_i, Z_k} = \sum_{k=1}^{p} e_{ik} \lambda_k e_{jk} \quad \text{where } i,k = 1,2\ldots\ldots,p$$

In this case, (λ_1, e_1), (λ_2, e_2),…………, (λ_k, e_k) are eigenvalue-eigenvector pairs for ρ, with $\lambda 1 \geq \lambda 2 \geq \ldots\ldots \geq \lambda p \geq 0$.

For the clusters listed in Table 8, Principal component analysis I was performed using correlation matrix in MINITAB 2000. The weights associated obtained for each variable in all of the five clusters are shown in Table 9. MINITAB output is shown in Appendix D. Since all the variables within each cluster have the same sign (see Table 9) for the weights, it can be inferred that increase in the value of any variable in a given cluster increases cluster score and hence the risk level.

Table 9 Cluster Scores

Variables	PCA I scores
Cluster 1	
Upper arms	0.347
Trunk twist	0.400
Trunk flexion	0.363
Gloves	0.525
Terrain	0.559
Cluster 2	
Lower arms	0.226
Wrist ulnar	0.589
Wrist radial	0.575

Table 9 Cluster Scores - Continued

Wrist extension	0.254
Kneeling	0.215
Chuck grip	0.340
Key grip	0.215
Cluster 3	
Neck	0.601
Duration	0.645
Pencil grip	0.472
Cluster 4	
Standing	0.707
Force	0.707
Cluster 5	
Sitting	1

Modeling and Validation

Figures 3 and 4 in Chapter 4 clearly revealed that the rate of MSDs was highest for electric line crews, lowest for meter readers and nearly between were gas line crews. Sixty-seven data points were collected from five different activities within these jobs, as discussed in Chapter 4. Response variable is defined as the 'Risk level'. It is a categorical (ordinal variable) that falls into three categories: low, medium, and high and is assigned values 1, 2, and 3, respectively. Training and testing data sets from Appendix C were used for modeling and validation, respectively. This section focuses on the third objective of modeling and validating data using three different multivariate statistical techniques; ordinal logistic regression, linear discriminant analysis, and nearest neighbour analysis; and further identifying significant clusters.

Ordinal Logistic Regression

Ordinal logistic regression was used to perform logistic regression on an ordinal variable. Ordinal variables are categorical variables that have three or more possible response levels with a natural ordering such as strongly disagree, disagree, neutral, agree, and strongly agree. A model was fit (MINITAB, 2002) using an iterative–reweighted least squares algorithm to obtain maximum likelihood estimates of the parameters (McCullagh and Nelder, 1992). Parallel regression lines were assumed, and therefore a single slope was calculated for each covariate. Logit link function, the inverse of cumulative logistic distribution function (logit) was used in this ordinal response model.

The model was defined as

$$\ln(X_{ij}/(1-X_{ij}))=\theta+X_j'\beta, \qquad i=1,\ldots,k\text{-}1$$

where k = the number of distinct values of the response = 3

$\ln(X_{ij}/(1-X_{ij}))$ = logit function

X_j' = A vector of predictor variables associated with the jth covariate pattern (cluster variable vector)

j = 1, 2,....,67 (in our case)

β = A vector of coefficients associated with the predictors

θi = The constant associated with the ith distinct response

In other terms,

$$X_{ij}=P(R\leq i/X_j')=\mathrm{e}^{\theta i+Xj'\beta}/(1+\mathrm{e}^{\theta i+Xj'\beta})$$

Output shown in Appendix D has all the parameters of the above model. *Xij* for both training and testing data sets are shown in Appendix D as predicted by the model. The results are summarized in Tables 10 and 11 for training and testing data, respectively.

Table 10 Ordinal logistic regression- training data results

Observed Response	Predicted Response			
	1	2	3	Total observations
1	3	2	0	5
2	2	12	0	14
3	0	0	31	31

Table 11 Ordinal logistic regression- testing data results

Observed Response	Predicted Response			
	1	2	3	Total observations
1	0	1	0	1
2	3	3	3	9
3	0	0	7	7

It can be seen from Tables 10 and 11 that ordinal logistic regression predicted 92% and 58.8% of training and testing data sets accurately, respectively. The reasons for the low accurate prediction rate of testing data set may be

1. Small training data set

2. It failed to predict responses 1 and 2 because of fewer number of observations in those categories in the training set

3. It does not use prior probabilities

However, after using prior probabilities ordinal logistic regression predicted 76.5% of the testing data accurately as shown in Table 12. Prior probabilities are calculated from training data set and they were found to be 0.62 (5 out of 50 observations), 0.28 (14 out of 50 observations) and 0.10 (31 out of 50 observations) for responses 1, 2, and 3 respectively. Multiply the probabilities obtained in Appendix D with their respective prior probabilities and find the highest probability and assign the observation to that category. New probabilities of prediction for the testing data after using prior probabilities are shown in Appendix D.

Table 12 Ordinal logistic regression - testing data results using prior probabilities

Observed Response	Predicted Response			
	1	2	3	Total observations
1	0	1	0	1
2	0	6	3	9
3	0	0	7	7

From Table 12, it can be observed that the model failed to predict 4 out of 17 observations in the testing data set. These observations belong to categories 1(1 out of 4) and 2 (3 out of 4) respectively. The model needs to be trained with more data in these categories. Results in Table 12 also reveal that the incorrectly predicted responses are predicted at a higher-level meaning that responses 1 and 2 are predicted as 2 and 3 respectively. From an ergonomic point of view in developing a model for assessing risk, the author believes that it is better to have false positives (predicting higher risk when there is low risk) than false negatives (predicting low risk when there is high risk).

From the logistic regression table of ordinal logistic regression in Appendix D, one cluster representing lower extremities, one cluster representing the upper part of upper extremities, one cluster having miscellaneous ergonomic variables were found to be significant ($p<0.05$). Deviance and Pearson goodness of fit test measures from Appendix D need to be ignored because sample size is not large enough to conduct the test

Fisher's Linear Discriminant Analysis

The main idea behind using this technique is to transform multivariate observations **x** (cluster scores) to univariate observations y such that the y's derived from different populations (clusters) would be separated as much as possible. Y's are a linear function of **x**. This approach does not assume that the populations are normal. It does, however, implicitly assume that the population covariance matrices are equal, because a pooled estimate of the common covariance matrix is used (Johnson and Wichern, 2000).

A fixed linear combination of the x's take the values $y_{11}, y_{12}, \ldots, y_{1n1}$ for the observations from the first population and $y_{21}, y_{22}, \ldots, y_{2n}$ from the second population and so on. The analysis was been done using SAS, 2001. Three linear discriminant functions obtained for the three responses are as follows.

Y1 = -11.43389 + 4.05774*c1+3.11724*c2+9.42431*c3+2.64192*c4+6.62153*c5

Y2 = -11.90690 +4.82151*c1+4.39619*c2+5.80931*c3+7.18168*c4+9.88753*c5

Y3 = -18.27902 +10.99859*c1+4.06439*c2+6.14536*c3+8.02179*c4+15.68814*c5

where c1, c2,….,c5 represent cluster scores.

After computing Y1, Y2, and Y3,the largest value among them is found. If Y1 is the largest value, then that observation has '1(low risk)' as the response. The results are summarized in Tables 13 and 14 for training and testing data, respectively.

Given equation for Y_i :

$$Y_i = \ln \Pi_i + \overline{x_i}\, S_p^{-1} (X - \frac{1}{2}\overline{x_i})$$ where S_p is the pooled covariance matrix

Table 13 Linear discriminant analysis- training data results

Predicted Response	Observed Response			
	1	2	3	Total observations
1	4	1	0	5
2	1	12	1	14
3	0	4	27	31

Table 14 Linear discriminant analysis- testing data results

Predicted Response	Observed Response			
	1	2	3	Total observations
1	1	0	0	1
2	0	6	3	9
3	0	0	7	7

It can be calculated from Tables 13 and 14 that linear discriminant analysis predicted 86% and 76.5% of training and testing data sets accurately, respectively. This technique predicted comparably well with ordinal logistic regression on the testing set but not on the training set. However, assumptions of equal covariance matrices and multivariate normality (because of binary values in *cluster 5*) were not realistic. This technique will not be pursued further for modeling.

Nearest Neighbor Classification Rule

Nearest neighbor classification rule is also know as k nearest neighbor rule, the earliest non-parametric classification method developed by Fix and Hodges in 1951. The method does not assume any distribution with the cluster variables. The procedure is conceptually simple as described below (Rencher, 2002):

The distance from an cluster score $\mathbf{y}_i$ to all other points $\mathbf{y}_j$ using the distance function

$$(y_i - y_j)' S_{pl}^{-1} (y_i - y_j), \qquad j \neq i$$

where S_{pl} is the pooled covariance matrix

To classify $\mathbf{y}_i$ into one of three groups (corresponding to three responses), k points nearest to $\mathbf{y}_i$ are examined, and if the majority of the k points belong to G1 (group 1), assign $\mathbf{y}_i$ to G1; otherwise to G2 or G3. Let number of points in G1, G2, and G3 are k1, k2, and k3 respectively, where k = k1+k2+k3. Let prior probabilities for G1, G2, and G3 be p1, p2, and p3 (0.1, 0.28 and 0.62 from SAS, 2001). Assign $\mathbf{y}_i$ to G1 if

$$\frac{k1/n1}{k2/n2} > \frac{p2}{p1} \, and \, \frac{k1/n1}{k3/n3} > \frac{p3}{p1}.$$

In general, assign $\mathbf{y}_i$ to the group that has the highest proportion *piki/ni,* where *ki* is the number of points in *Gi* among the *k* nearest neighbors of the $\mathbf{y}_i$ in question (Rencher, 2002). The analysis was done using SAS, 2001 with k =5. The results are summarized in Tables 15 and 16 for training and testing data, respectively.

Table 15 Nearest neighbor analysis- training data results

Predicted Response	Observed Response			
	1	2	3	Total observations
1	0	5	0	5
2	2	11	1	14
3	0	5	26	31

Table 16 Nearest neighbor analysis- testing data results

Predicted Response	Observed Response			
	1	2	3	Total observations
1	1	0	0	1
2	0	6	3	9
3	0	0	7	7

It can be calculated from Tables 15 and 16 that nearest neighbor analysis predicted 78% and 64.7% of training and testing data sets accurately, respectively. This technique did not perform better than ordinal logistic regression though it was conceptually expected. This technique does not assume any distribution associated with the data. This technique can still be used for modeling

Web Application

This section discusses the fourth objective of developing a web application of the ordinal logistic regression model discussed in the previous section. The web application will help the user to enter and store the inputs (observations) and get the response (risk level). The four important components of the web application are

1. Database Management System (DBMS)

2. Model-base Management System (MBMS)

3. User interface

4. Mail or Message Management System (MMS)

All the components are discussed in detail in the following sections.

Database Management System (DBMS)

"The DBMS provides access to data as well as all of the control programs necessary to get those data in the form appropriate for the analysis under consideration, without the user programming the effort" (Sauter, 1997). It gives users access to the data even though the users are usually unaware of data's physical location. It also facilitates the merger of data from various sources without explicit instructions from the user regarding the accomplishment of the task.

The database is not relational; it has only a single entity (table) called 'observation' with 19 attributes and serial number as the primary key. The database schema is shown in Figure 9. MYSQL source file is shown in Appendix E.

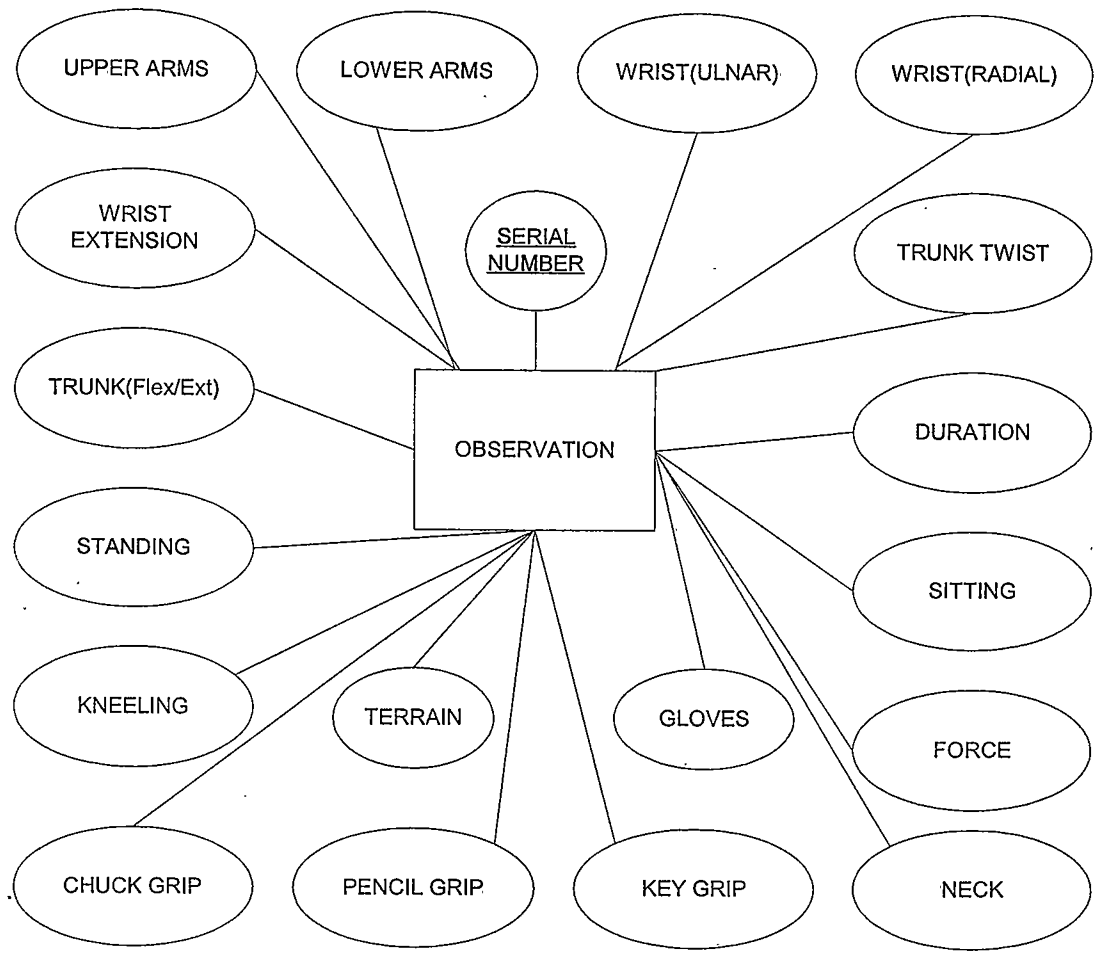


Figure 9 Database schema diagram

The entity type (entity) describes the schema or intension for a set of entities sharing the same structure. The collection of entities of a particular entity type is grouped into an entity set, also called the extension of the entity type (Elmasri and Navathe, 1999). The intension and extension of observation entity is shown in Figure 10.

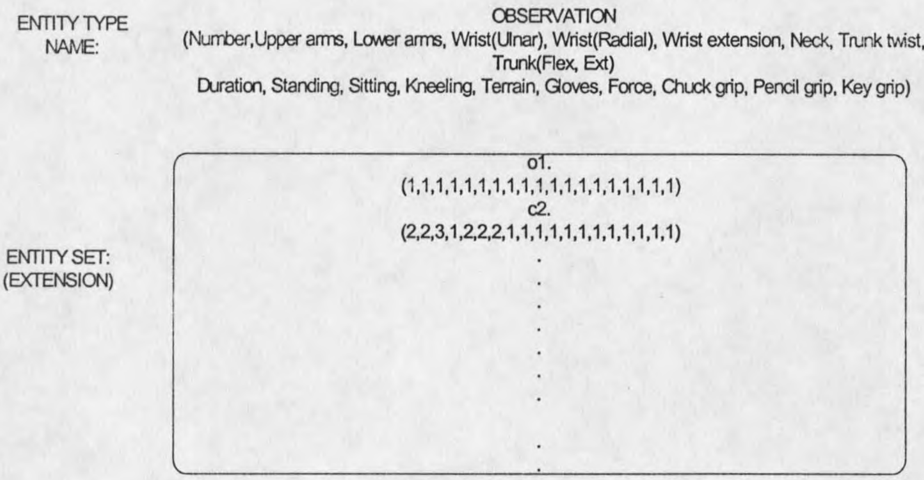


Figure 10 Intension and extension of the Observation entity

The table 'Observation' satisfies all the four normal forms and the reasons are explained in Table 17.

Table 17 Database normalization

Normal Form	Satisfied (Yes/No)	Reason
First form	Yes	All attributes have simple atomic values.
Second form	Yes	None of the primary key's are multiple keys
Third form	Yes	No non-key attribute is functionally determined by another non-key attribute
Fourth form	Yes	No non trivial multi-valued dependencies

The flow of data in the web application is shown in Figure 11. The data entered by the user using forms will be read into the database using MYSQL PHPAPI and then be sent through MYSQL CAPI to the C program for processing using the ordinal logistic procedure. The C program reads the data from the database and writes it back to the database after processing using MYSQL CAPI. Once the database is updated after processing, the user can view the results on the screen that come through MYSQL PHPAPI.

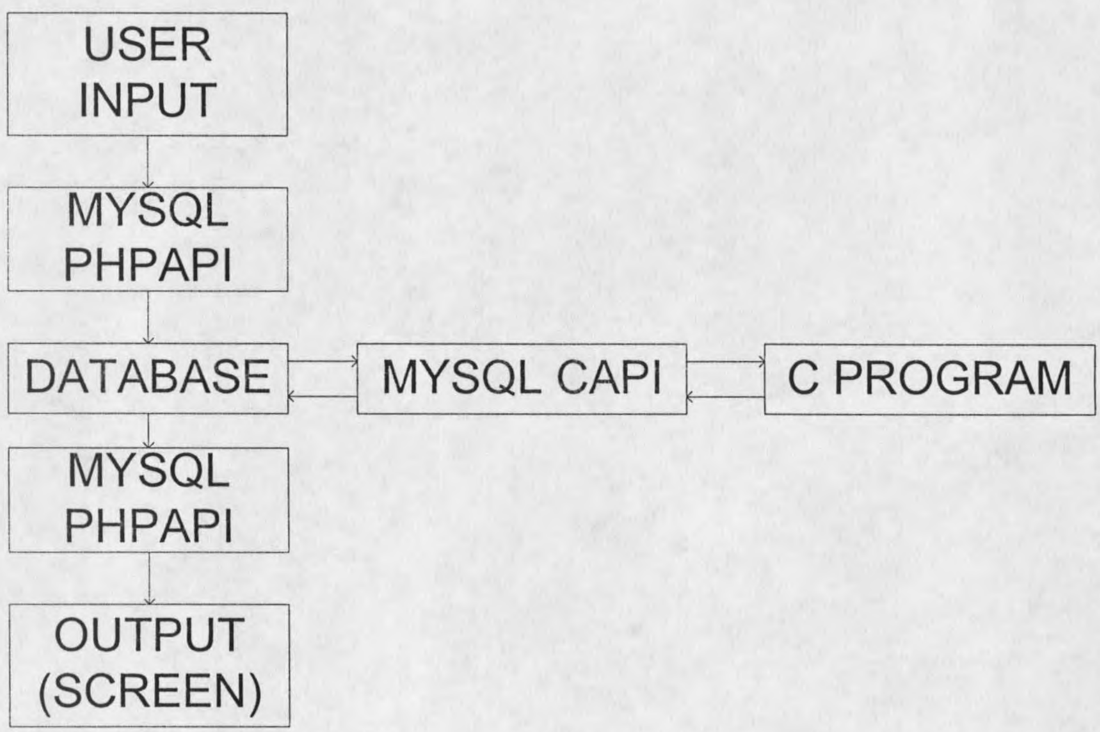


Figure 11 Data flow diagram

In order to manage the database, the user should be able to insert, update and delete the records and view the table. After these operations are performed, the updated table is displayed. A screen capture from the *database and model* link is shown in Figure 12 for the *display table* and *insert record* options.

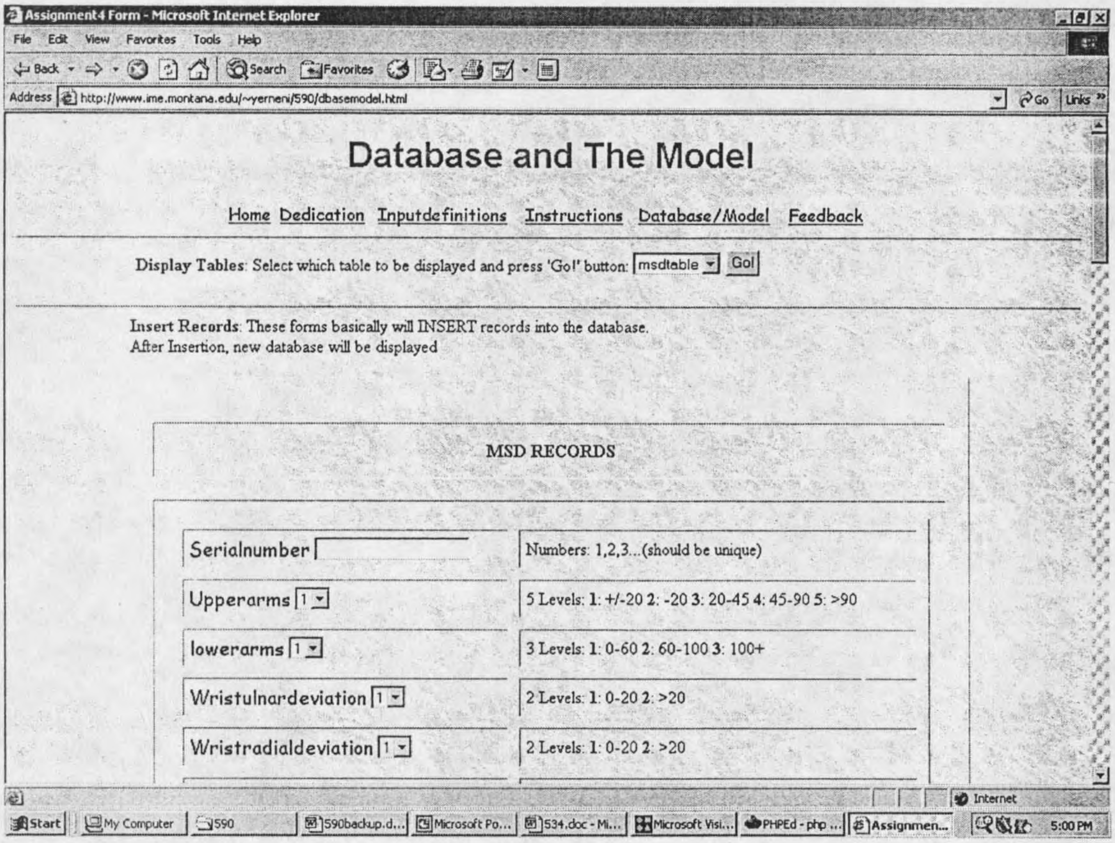


Figure 12 Database management: View table and insert record

Furthermore, the input, serial number entered by the user is validated. For example, if a user tries to enter a character instead of an integer in the serial number, an error message pops up as shown in Figure 13. The error check will make sure the user enters input correctly.

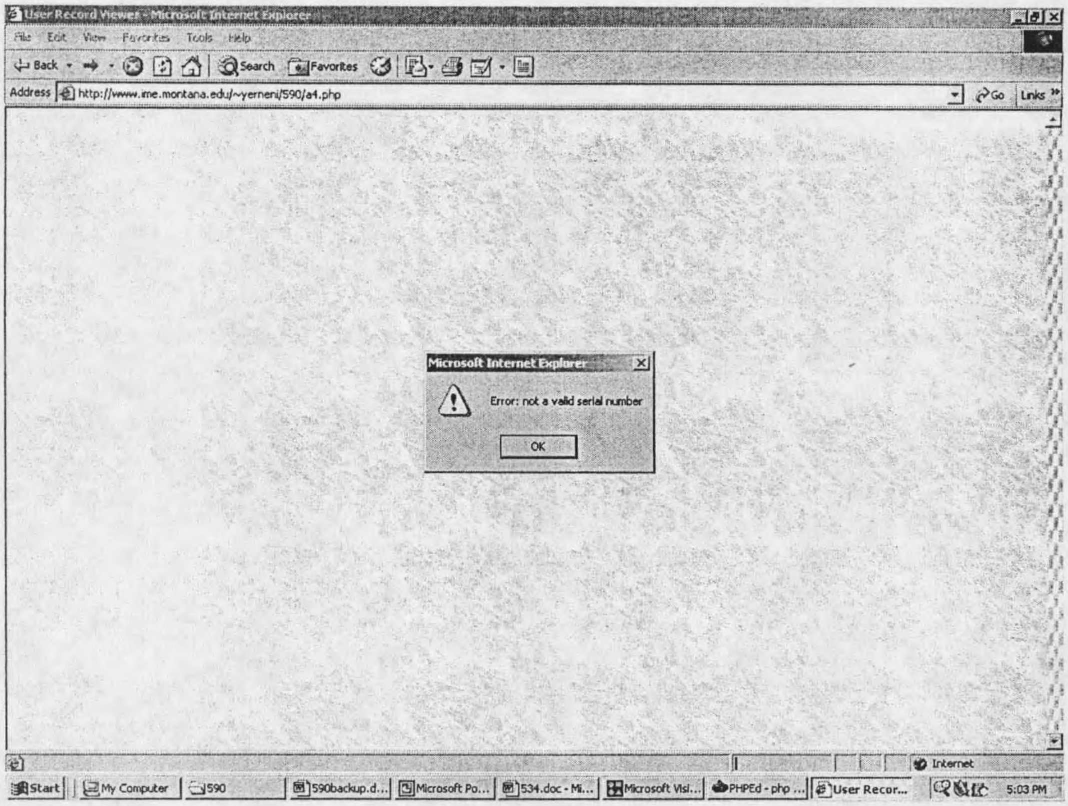


Figure 13 Database management: Input validation

Model-base Management System (MBMS)

A model is an abstraction or simplified representation of a system. Model-base management systems consist of models, easy access to models, and help in using those models and presenting understandable results.

A model is a generalized description of a decision environment. Models are created to simplify a phenomenon in order to understand its behavior. There are several types of models such as statistical models, accounting models, marketing models, personnel models, etc. Sauter (1997) uses three different dimensions to describe models. They are

1. The representation

2. The time dimension

3. The process or methodology

The representation of the data in the current model is considered "objective" because of the way they are specified, constant, and independent of the specific decision maker's experiences. There is no room for subjectivity in the analysis of the data. The time dimension identifies whether the model is static or dynamic. It is a static model because it represents a snapshot in time of all factors affecting the decision environment. The third dimension, methodology, addresses how the data will be collected and processed. Ordinal logistic regression is an analytical model. The model is described in the Modeling and Validation section of this chapter. C program code using MYSQL CAPI was written for the model and is shown in Appendix E. The flow chart for the C program is shown in Figure 14. The user can easily access the model and the database in *database and the model* link. The user is given instructions on how to operate the database and the model as shown in Figure 15. The web application was aimed at the novice user. The output (risk level) along with the inputs is presented in tabular form as shown in Figure 16.

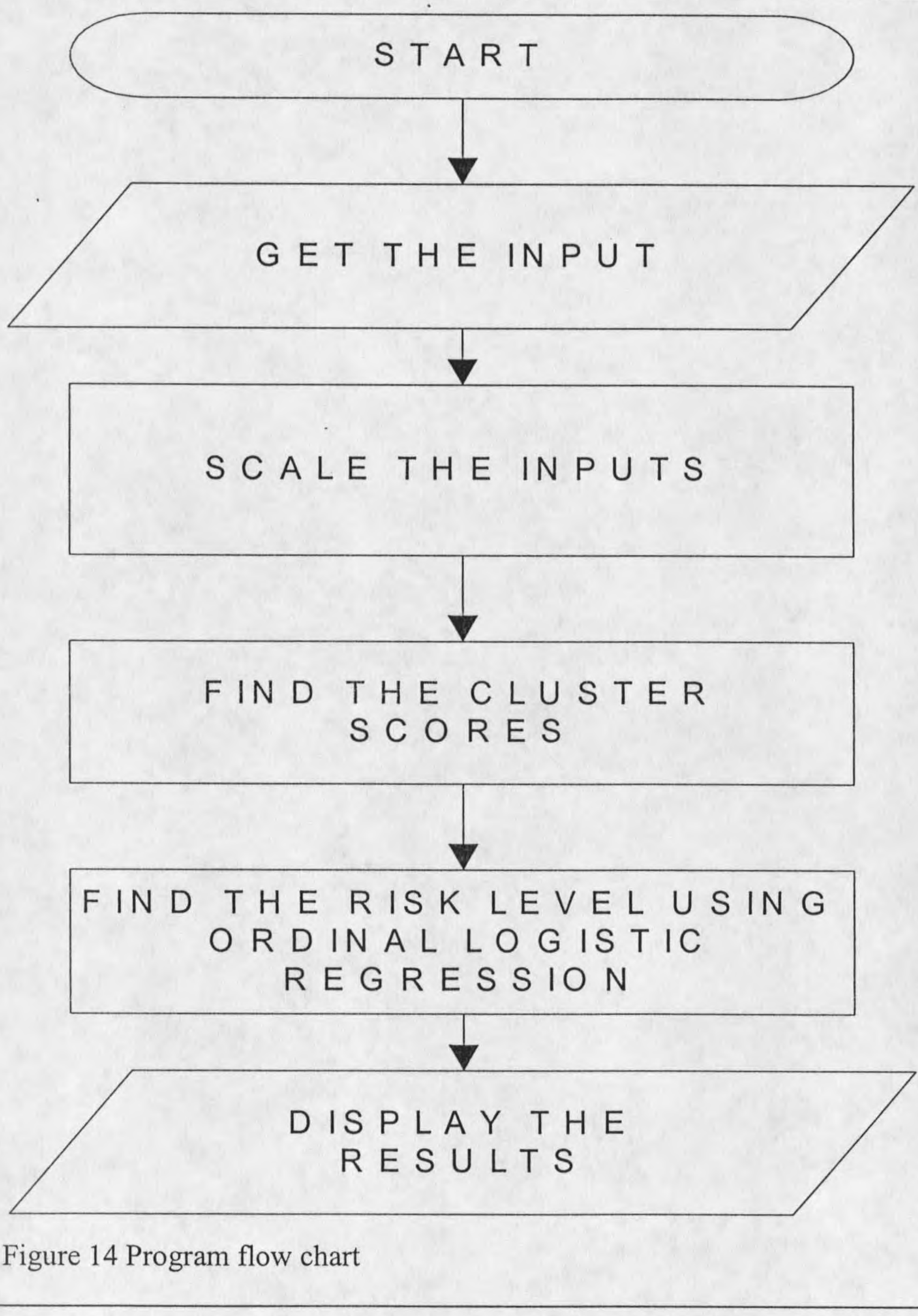
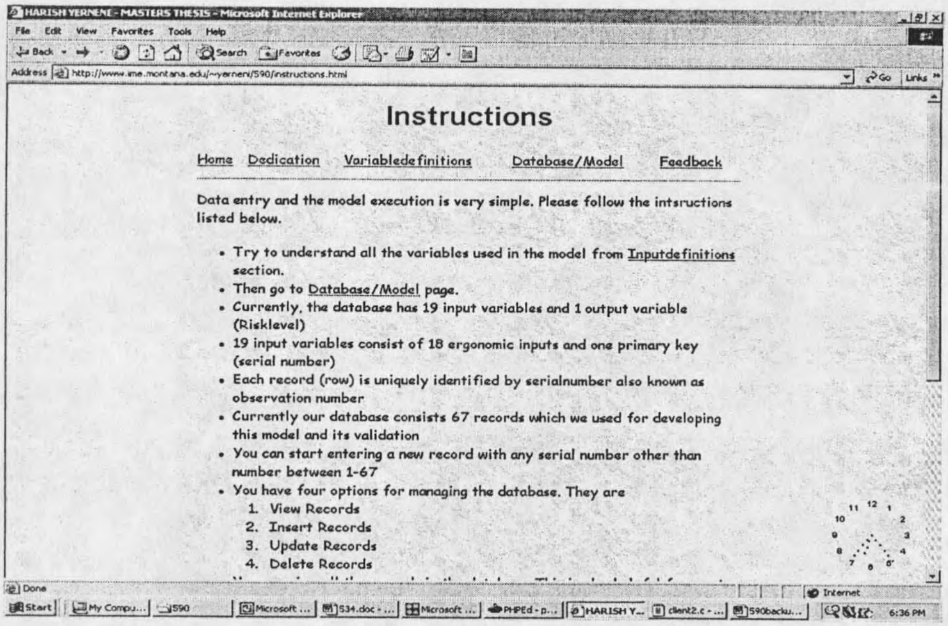


Figure 14 Program flow chart



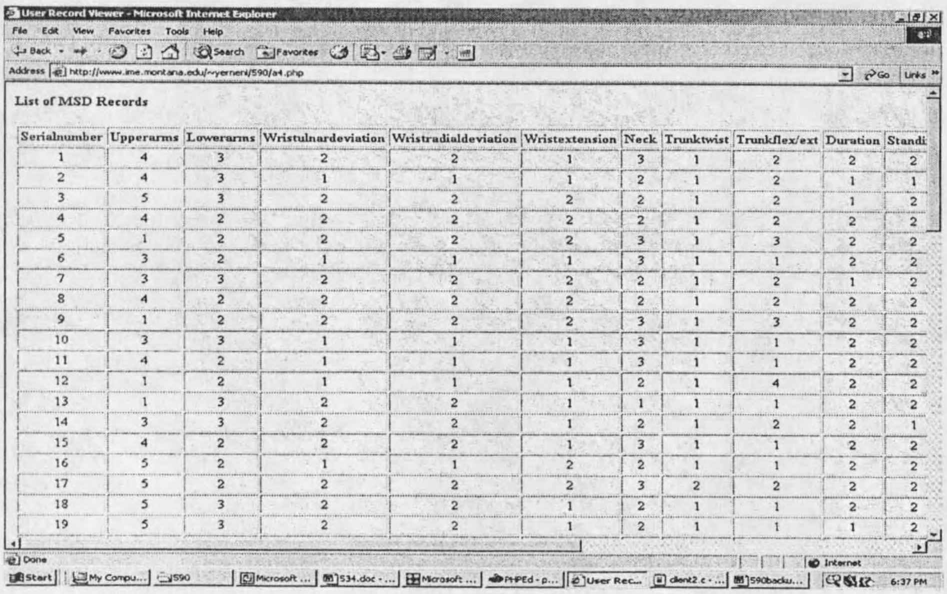
Instructions

Home Dedication Variabledefinitions Database/Model Feedback

Data entry and the model execution is very simple. Please follow the intsructions listed below.

- Try to understand all the variables used in the model from Inputdefinitions section.
- Then go to Database/Model page.
- Currently, the database has 19 input variables and 1 output variable (Risklevel)
- 19 input variables consist of 18 ergonomic inputs and one primary key (serial number)
- Each record (row) is uniquely identified by serialnumber also known as observation number
- Currently our database consists 67 records which we used for developing this model and its validation
- You can start entering a new record with any serial number other than number between 1-67
- You have four options for managing the database. They are
    1. View Records
    2. Insert Records
    3. Update Records
    4. Delete Records

Figure 15 Database and model instructions



List of MSD Records

Serialnumber	Upperarms	Lowerarms	Wristulnardeviation	Wristradialdeviation	Wristextension	Neck	Trunktwist	Trunkflex/ext	Duration	Standi
1	4	3	2	2	1	3	1	2	2	2
2	4	3	1	1	1	2	1	2	1	1
3	5	3	2	2	2	2	1	2	1	2
4	4	2	2	2	2	2	1	2	2	2
5	1	2	2	2	2	3	1	3	2	2
6	3	2	1	1	1	3	1	1	2	2
7	3	3	2	2	2	2	1	2	1	2
8	4	2	2	2	2	2	1	2	2	2
9	1	2	2	2	2	3	1	3	2	2
10	3	3	1	1	1	3	1	1	2	2
11	4	2	1	1	1	3	1	1	2	2
12	1	2	1	1	1	2	1	4	2	2
13	1	3	2	2	1	1	1	1	2	2
14	3	3	2	2	1	2	1	2	2	1
15	4	2	2	2	1	3	1	1	2	2
16	5	2	1	1	2	2	1	1	2	2
17	5	2	2	2	2	3	2	2	2	2
18	5	3	2	2	1	2	1	1	2	2
19	5	3	2	2	1	2	1	1	1	2

Figure 16 Results

User Interface

The user interface includes all the mechanisms by which commands, requests, and data are entered into the web application, as well as all the methods by which results and information are output by the system. The user interface is usually described in terms of its components as well as its mode of communication.

The three important components of user interface are

1. Action language
2. Display or presentation language
3. Knowledge base

The action language identifies the form of input used by the user to enter requests to the system. In the present application, the user enters the input through forms having text boxes and drop-down menus. Essentially, the input entry is in menu format. The application put forth the results to the user in tabular, graphical and text formats. Also, during input validation phase, the user gets feedback through error message pop ups. The application is designed for a novice user. The user is provided with input definitions in an easy to understand style as shown in Figure 17.

From the discussion in the previous section, the user communicates mainly through drop-down menus, text boxes and the application communicates through tables, graphs, and text formats. It is easy to navigate through the application because hyperlinks are provided at each and every page.

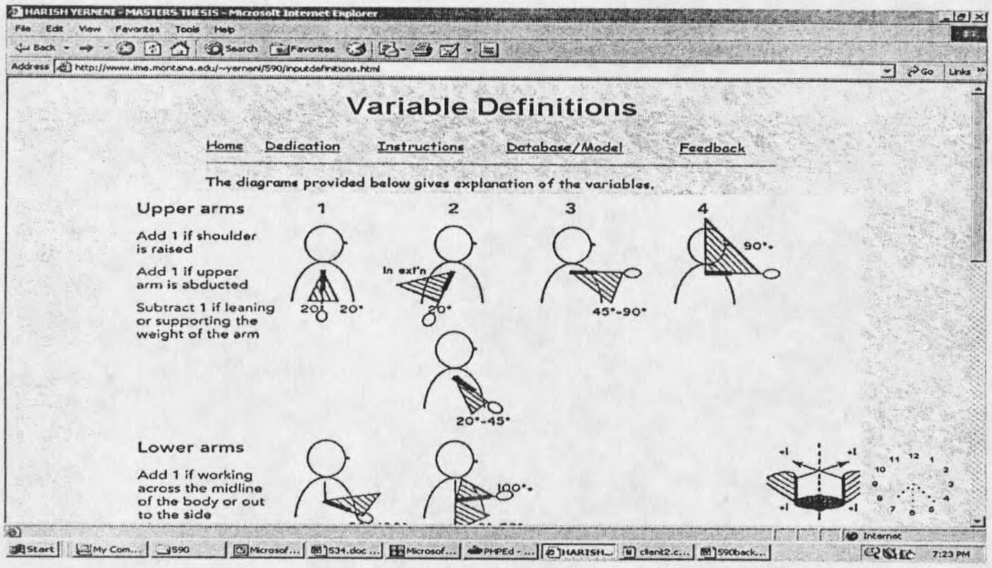


Figure 17 Knowledge base: Input definitions

Message Management System (MMS)

The user can express any concerns and questions to the administrator of the application, using the feedback form shown in Figure 18. Once the user sends the feedback form, the program generates an acknowledgement window immediately as shown in Figure 19. The feedback sent by the user is written into a text file in append mode and the administrator of the application can gets back to the user as soon as possible. All the php files used for the application are listed in Appendix E and html files can be viewed using view source option of the browser at www.ime.montana.edu/~yerneni/590/index.html.

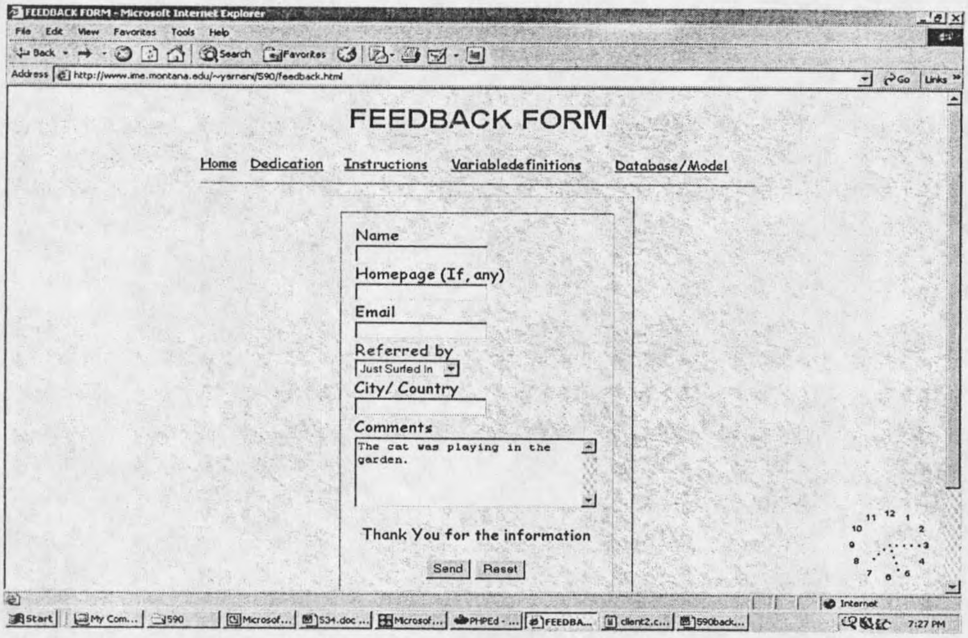


Figure 18 Feedback form

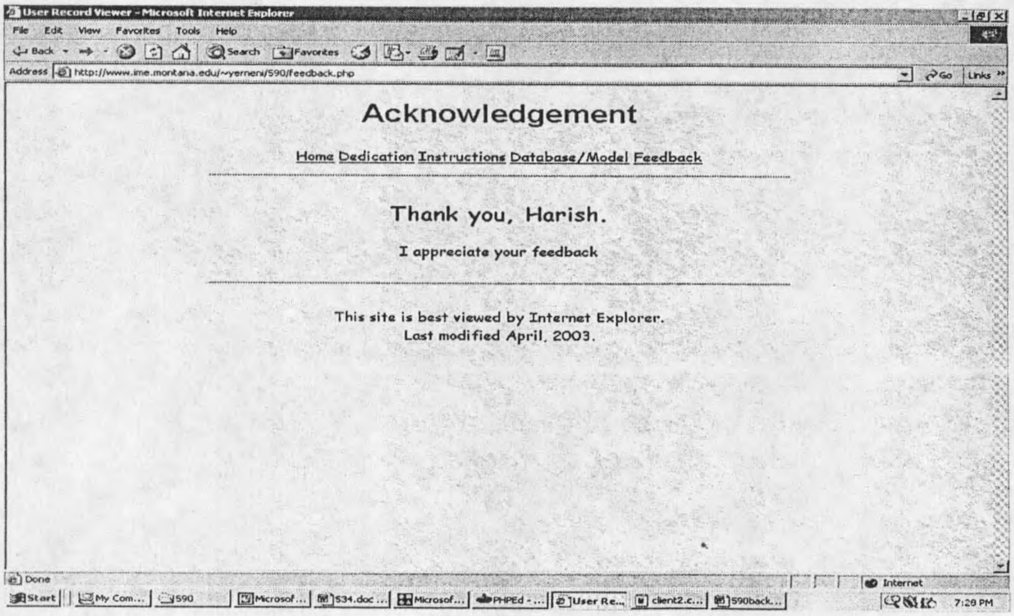


Figure 19 Feedback acknowledgement

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

Following are the conclusions drawn by the author from the multivariate statistical analysis of Marley, et. al., (1997) study.

1. Forty-one mixed ergonomic variables (binary, ordinal and continuous) used by Marley, et. al., (1997) were reduced to 18 mixed variables (binary, ordinal). Standardized scores in the range [0,1] were assigned to different levels of the mixed variables.

2. 18 variables were grouped into 5 clusters using cluster analysis. The cluster analysis used average linkage method; and, Pearson product moment correlation coefficient was used as the distance measure.

3. Two clusters represented upper and lower parts of upper extremities. Two clusters represented lower extremities, and one cluster represented miscellaneous parts of the human body.

4. The clusters weights (scores) were determined using principal component analysis (PCA) I. PCA I used correlation matrix of the cluster variables for determining their weights.

5. All the variable weights within each of the five clusters have the same sign, implying that an increase in the value of any variable in a given cluster will increase cluster score and hence the risk level.

6. The modified data (67 observation points) were divided randomly into two sets training (80% of the modified data, 50 observations) and testing (20% of the modified data, 17 observations).

7. The data were modeled and validated using ordinal logistic regression, Fisher's linear discriminant analysis, and nearest neighbor rule.

8. Using ordinal logistic regression, one cluster representing lower extremities, one cluster representing the upper part of upper extremities, and one cluster having miscellaneous ergonomic variables were found to be significant ($p<0.05$). Deviance and Pearson goodness of fit test measures were ignored because sample size was not large enough to conduct the test. The technique with prior probabilities predicted 92% and 76.5% of training and testing data sets accurately, respectively. The reasons for the low accuracy in prediction rate of testing data set may be due to:

    a. Small training data set

    b. Failure to predict responses 1 and 2 because of too few observations in those categories of the training set

9. Linear discriminant analysis predicted 86% and 76.5% of training and testing data sets accurately, respectively. This technique predicted comparably well with ordinal logistic regression on the testing set but not on training set. However, assumptions of

equal covariance matrices and multivariate normality (because of binary values in *cluster 5*) were not realistic. This technique will not be pursued further for modeling.

10. Nearest neighbor analysis predicted 78% and 64.7% of training and testing data sets accurately, respectively. This technique did not perform better than ordinal logistic regression though it was conceptually expected. The technique does not assume any distribution associated with the data. This technique can still be used for modeling.

11. A user-friendly web application (www.ime.montana.edu/~yerneni/590/index.html) targeting the novice user was developed for the ordinal logistic regression model. In the web application, the user can enter 18 input variables and obtain the risk level for a given task.

Recommendations for the future work are

1. Once MSD risk is evaluated based on cluster scores, the correlation between cluster scores and the extremity affected (arms, back, legs and other) will help in determining specifically which extremity of the body is affected.

2. The model should be trained with more external data for better prediction.

3. Nine other jobs can also be analyzed using the revised work elements and their new levels developed in the current work.

REFERENCES CITED

REFERNCES CITED

Armstrong, T.J., & Chaffin, D.B. 1979a. Some biomechanical aspects of the carpal tunnel. *Journal of Biomechanics*, 12(7): 567-570.

Armstrong, T.J., Radwin, R.G., Hansen, D.J., & Kennedy, K.W. 1986. Repetitive trauma disorders: job evaluation and design. *Human Factors*, 28(3): 325-226.

Atcheson, S.G., Ward, J.R., & Lowe, W. 1998. Concurrent medical disease in work-related carpal tunnel syndrome. *Archives of Internal Medicine*, 158(14): 1506.

Bureau of Labor Statistics. 2002. Occupational injuries and illnesses: industry data (1989-most current). *United States Department of Labor*. http://data.bls.gov/cgi-bin/surveymost?sh

Birnbaum, J.S. 1986. *The Musculoskeletal Manual* (2nd Ed.). Philadelphia: W.B. Saunders Company.

Chaffin, D.B., Anderson, G.B.J., & Martin, B.J. 1999. *Occupational Biomechanics* (3rd Ed.). New York: John Wiley & Sons, Inc.

Choi, W., Kent, A., Lea, C., Prasad, G., Ullman, C., 2000. *Beginning PHP 4* (1st Ed.). Birmingham: Wrox Press Ltd.

DuBois, P., 2000. *MySQL*. Indianapolis: New Riders Publishing Company.

Elmasri, R., Navathe, S.B., 1999. *Fundamentals of Database Systems* (3rd Ed.). Massachusetts: Addison-Wesley.

Feldman, R.G., Goldman, R., & Keyserling, W.M. 1983. Peripheral nerve entrapment syndromes and ergonomic factors. American Journal of Industrial Medicine, 4: 661-681.

Fernandez, J.E., & Marley, R.J. 1990. Monitoring and screening tests for carpal tunnel syndrome. In, Lovesey, E.J.(Ed.). *Contemporary Ergonomics,* London: Taylor & Francis, 63-68.

Fernandez, J.E., & Marley, R.J. 1998. *Applied Occupational Ergonomics*, Dubuque, Iowa: Kendall/Hunt Publishing Company.

Feurstein, M., Miller, V.L., Burrell, L.M., & Berger, R. 1998. Occupational upper extremity disorders in the federal workplace. *Journal of Occupational and Environmental Medicine*, 40(6): 546.

Gordon, L. 1995. Hand and wrist disorders. In: Herington, T.N., Morse, L.H (Eds.), *Occupational injuries: evaluation, management and prevention*. St Louis: Mosby, p. 115-124.

Herington, T.N & Morse, L.H. 1995. In: Herington, T.N., Morse, L.H (Eds.), *Occupational injuries: evaluation, management and prevention*. St Louis: Mosby, p. 333-345.

Greico, A., Molteni, G., DeVito, G. & Sias, N. 1998. Epidemiology of musculoskeletal disorders due to biomechanical overload. *Ergonomics*, 41(9): 1253-1260.

Johnson, R.W., & Wichern, D.W. 2002. *Applied Multivariate Statistical Analysis* (5th Ed.). New Jersey: Prentice Hall.

Kaufman. L., & Rousseeuw. P.J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons, Inc.

Kernighan, B.W., & Ritche, D.M. 1999. *The C Programming Language* (2nd Ed.). New Delhi: Prentice-Hall of India Private Limited.

Kroemer, K.H.E., 1989. Cumulative trauma disorders: Their recognition and ergonomic measures to avoid them. *Applied Ergonomics,* 20(4): 274-280.

Marley, R.J., & Kumar, N. 1996. An improved musculoskeletal discomfort assessment tool. *International Journal of Industrial Ergonomics* , 17: 21-27.

Marley, R.J., Gebhardt, A.J., & Lavier, L. 1997. Work factors associated with musculoskeletal disorders in power distribution jobs. *American Industrial Hygiene Conference & Exposition*, May 17-23, 1997

McAtammey, L., & Corlett, E.L. 1993. RULA: A Survey Method for the Investigation of Work-Related Upper Limb Disorders. *Applied Ergonomics*, 24: 91-99.

McCullagh, P., & Nelder, J.A., 1992. Generalized Linear Models. Chapman & Hall.

Meloni, J.C., 2000. *PHP Essentials* (1st Ed.). California: Prima Publishing

MINITAB (2000): *Electronic documentation of Stat Guide*, Version 13. State College, PA: Minitab Inc.

Mooney, E.L., 2003. *IME 534 class notes.*

Mooney, E.L., 2003. *IME 534 class website* (http://www.ime.montana.edu/~emooney/ime534/).

Moore, J.S., & Garg, A. 1995. The Strain Index: A proposed method to evaluate jobs for risk of distal upper extremity morbidity. *American Industrial Hygiene Association Journal,* 56(5): 443-458.

Muggleton, J.M., Allen, R., & Chappell, P.H. 1999. Hand and arm injuries associated with repetitive manual work in industry: a review of disorders, risk factors and preventive measures. *Ergonomics*, 42(5): 714-739.

Naderi, B., & Ayoub, M.M. 1989. *Cumulative hand trauma disorders*. Lubbock, TX: Institute for Ergonomics Research, Texas Tech University.

NIOSH (1997). Musculoskeletal disorders (MSDs) and workplace factors: A critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back. *United States Department of Health and Human Sciences*. http://www.cdc.gov/niosh/ergosci1.html

Pecina, M.M., & Bojanic, I. 1993. *Overuse injuries of the musculoskeletal system*. CRC Press, Inc. Boca Raton, Florida.

Putz-Anderson, V. 1988. *Cumulative Trauma Disorders: A Manual for Musculoskeletal Disorders of the Upper Limbs*. London: Taylor & Francis.

Ranney, D. 1993. Work-related chronic injuries of the forearm and hand: their specific diagnosis and management. *Ergonomics*, 36(8): 871-880.

Rowe, M.L 1995. *Orthopaedic problems at work*. Fairport, NY: Perinton press.

Sallstrom, J. & Schmidt, H. 1984. Cervicobrachial disorders in certain occupations, with special reference to compression in thoracic outlet. *American Journal of Industrial Medicine*, 6: 45-52

Salter, R.B. 1970. *Textbook of disorders and injuries of the musculoskeletal system.* Baltimore: Williams & Wilkins.

SAS (2000): SAS User's Guide, Version 8. Cary, NC: SAS Institute, Inc.

Schildt, H. 1999. *C++: The Complete Reference* (3rd Ed.). New Delhi: Tata McGraw-Hill Publishing Company Limited.

Silverstein, B.A., Fine, L.J., & Armstrong, T.J. 1986. Hand wrist cumulative trauma disorders in industry. *British Journal of Industrial Medicine*, 43: 779-784.

Sauter, V. 1997. *Decision Support Systems* (1st Ed.). New York: John Wiley and Sons, Inc.

Taylor, W. 1974. The vibration syndrome: Introduction. In Taylor, W. (Ed.), *The Vibration Syndrome.* London: Academic Press.

APPENDICES

APPENDIX A

INJURY DATA DETAILS

INJURY DATA DETAILS

Job classification	*Extremity affected				
	Total	Arms	Back	Legs	Other
Line worker	80	18	49	5	8
Mechanic	48	16	23	8	1
Office personal	30	11	16	1	2
Gas trade	23	8	15	0	0
Technician	22	5	12	3	2
Operator	20	5	15	0	0
Utility man	16	2	10	2	2
Warehouse	11	0	8	1	2
Hydro	10	1	8	1	0
Janitor/Janitress	8	4	3	1	0
Maintenance	8	1	5	1	1
Meter reader	7	3	2	2	0
Total	283	74	166	25	18

*Some injuries affected multiple parts of the body and were counted in each category

Job classification (Line worker)	Extremity affected				
Sub-classification	Total	Arms	Back	Legs	Other
Lineman	51	12	30	4	5
Sub foreman	10	2	6	0	2
Foreman-Working electric	9	3	5	1	0
Patrolman	4	1	3	0	0
Manager-Town	2	0	2	0	0
Electric foreman	1	0	1	0	0
Groundman	1	0	0	0	1
Lineman-Apprentice	1	0	1	0	0
Substation Journeyman	1	0	1	0	0

Job classification (Gas trade)	Extremity affected				
Sub-classification	Total	Arms	Back	Legs	Other
Serviceman	6	2	4	0	0
Craftsman	3	1	2	0	0
Foreman-Working gas	3	2	1	0	0
Gas serviceman	2	1	1	0	0
Operator-field	2	0	2	0	0
Pipefitter	2	2	0	0	0
Gas controller	1	0	1	0	0
Gas laborer	1	0	1	0	0
Journeyman-Excavator	1	0	1	0	0
Working foreman	1	0	1	0	0
Welder	1	0	1	0	0

Job classification (Maintenance)	Extremity affected				
Sub-classification	Total	Arms	Back	Legs	Other
Electrician	5	1	3	0	1
Const. Supervisor	1	0	1	0	0
Maintenance Leadman	1	0	0	1	0
Maintenance supervisor	1	0	1	0	0

Job classification (Mechanics)	Extremity affected				
Sub-classification	Total	Arms	Back	Legs	Other
Mechanic	19	6	10	3	0
Journeyman mechanic	11	2	8	1	0
Mechanic/ Welder/ Machinist	6	3	2	1	0
Machinist	4	2	0	2	0
Mechanic/Welder	3	0	1	2	0
Foreman-Mechanic	1	1	0	0	0

Fuel crew mechanic	1	0	0	0	0
Machinist-Sr	1	0	1	0	0
Mechanic-Apprentice	1	0	1	0	0
Welder	1	0	1	0	0

Job classification (Hydro)	Extremity affected				
Sub-classification	Total	Arms	Back	Legs	Other
Operator-Maintenance	5	1	3	1	0
Hydro-Maintenance	2	0	2	0	0
Maintenance man	1	0	1	0	0
Maintenance man-Electric	1	0	1	0	0
Maintenance Operator	1	0	1	0	0

Job classification (Warehouse)	Extremity affected				
Sub-classification	Total	Arms	Back	Legs	Other
Warehouse person	9	0	7	1	1
Storekeeper	1	0	1	0	0
Toolroom attendant	1	0	0	0	1

APPENDIX B

DATA COLLECTION FORM, INITIAL CLUSTER ANALYSIS, ORGINAL DATA

Job Class________

Activity__________

Task____________

Recorded By:______________

Date:___________________

Posture										
Element	Upper Arms	Lower Arms	Wrist (Ulnar/Radial Dev.)	Wrist (Flex./Ext.)	Neck	Trunk (Twist)	Trunk (Flex./Ext.)	Legs	Nature of Injury	Body Area Affected
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

Time Variables						Other				
Element	Exertion Duration	Recovery Duration	Duration Per day	# of Exertions Per Day	Gloves	Force	Grasp Type	Terrain	MSD's/FTE	Lost Days/FTE
					1 = None 2 = Light 3 = Heavy	1 = Low 2 = Medium 3 = High	1 = Power 2 = Chuck 3 = Pencil 4 = Key	1 = Good 2 = Fair 3 = Poor		
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

DATA SHEET COLLECTION FORM

INITIAL CLUSTER ANALYSIS OF TASK FACTORS

Cluster 1

wrist extension (20+ deg.)
wrist ulnar deviation (20+ deg)
wrist radial deviation (20+ deg)

Cluster 2

large lower arm deviation (elbow flexion 100+ deg)
medium lower arm deviation (elbow flexion 60-100 deg)
normal upper arm deviation (0-20 deg)
forward reach (upper arm deviation 45-90 deg)
wrist flexion (20+ deg)
neck flexion (forward 20+ deg)
very large trunk twist (90+ deg)
sitting (present)

Cluster 3

kneeling (present)
trunk twist (15-30 deg)
neutral neck flexion (<10 deg)
large trunk flexion (30-45 deg)
small lower arm deviation (0-60 deg)
large trunk twist (45-90 deg)
overhead reach (upper arm deviation 90+ deg)
heavy gloves (present)

Cluster 4

straight leg (present)
neutral trunk twist (0-15 deg)
neutral trunk flexion (0-15 deg)

Cluster 5

medium neck flexion (10-20 deg)
medium upper arm deviation (20-45 deg)
medium trunk flexion (15-30 deg)
medium force application (10-35 lbs)
good terrain (solid, level surface present)

Cluster 6

fair terrain (uneven surface present, but reasonable footing)
pencil type grip present
light gloves present
low force application (< 10 lbs)

no glove present

Cluster 7

chuck grip present
key grip present

Cluster 8

high force application (35+ lbs)
very large trunk flexion (45+ deg)
poor terrain (uneven surface with soft or unstable footing present)

Cluster 9

neutral wrist flexion (0-20 deg)
neutral wrist radial deviation (0-20 deg)
neutral wrist ulnar deviation (0-20 deg)
neutral wrist extension (0-20 deg)

Cluster 10

activity duration (minutes)

ORIGINAL DATA

CASE	JOBCLAS	ACTIVITY	TASK	UARM1	UARM2	UARM3	UARM4	UARM5	LARM1	LARM2	LARM3	WULN1
1	1	1	1	0	0	1	0	0	0	0	1	2
2	1	1	2	0	0	0	1	0	0	0	1	2
3	1	1	3	0	0	0	1	0	0	0	1	1
4	1	1	4	0	0	0	0	1	0	0	1	2
5	1	1	2	0	0	0	0	1	0	1	0	1
6	1	2	6	0	0	0	1	0	0	1	0	2
7	1	2	7	1	0	0	0	0	0	1	0	2
8	1	2	8	0	0	1	0	0	0	1	0	1
9	1	2	9	0	0	1	0	0	0	0	1	2
10	1	2	10	0	0	0	1	0	0	0	1	2
11	1	2	1	0	0	0	0	1	0	0	1	2
12	1	2	6	0	0	0	1	0	0	1	0	2
13	1	2	7	1	0	0	0	0	0	1	0	2
14	1	2	8	0	0	1	0	0	0	1	0	1
15	1	2	9	0	0	1	0	0	0	0	1	2
16	2	3	11	0	0	1	0	0	0	0	1	1
17	2	3	12	1	0	0	0	0	0	0	1	1
18	2	3	11	0	0	0	1	0	0	1	0	1
19	2	3	12	1	0	0	0	0	0	1	0	1
20	2	3	11	1	0	0	0	0	0	0	1	2
21	2	3	12	0	0	1	0	0	0	0	1	2
22	3	4	13	0	0	0	0	1	0	1	0	1
23	3	4	14	0	0	0	1	0	0	1	0	2
24	3	4	16	0	0	0	0	1	0	1	0	1
25	3	4	16	0	0	0	0	1	0	1	0	1
26	3	4	16	0	0	0	0	1	0	1	0	2
27	3	4	16	0	0	0	0	1	0	0	1	2
28	3	4	13	0	0	0	0	1	0	0	1	2
29	3	4	14	0	0	0	1	0	0	1	0	1
30	3	4	17	0	0	0	1	0	0	1	0	1
31	3	4	14	0	0	0	0	1	0	1	0	1
32	3	4	13	0	0	0	0	1	0	0	1	2
33	3	4	18	0	0	0	1	0	0	0	1	2

ORIGINAL DATA - CONTINUED

34	3	5	19	0	0	0	1	0	0	1	0	2
35	3	5	20	0	0	0	1	0	0	1	0	2
36	3	5	22	0	0	0	1	0	0	0	1	2
37	3	5	17	0	0	0	1	0	0	1	0	2
38	3	5	23	0	0	0	0	1	0	0	1	2
39	3	5	17	0	0	0	1	0	0	1	0	2
40	3	5	25	0	0	0	1	0	0	0	1	1
41	3	5	26	0	0	1	0	0	0	1	0	2
42	3	5	27	0	0		1	0	0	0	1	1
43	3	5	28	0	0	0	0	1	0	0	1	2
44	3	5	29	0	0	1	0	0	0	1	0	2
45	3	5	30	0	0	0	0	1	0	0	1	1
46	3	5	17	0	0	1	0	0	0	1	0	1
47	3	5	25	0	0	0	1	0	0	1	0	1
48	3	5	28	0	0	0	0	1	0	1	0	2
49	3	5	30	0	0	0	0	1	0	0	1	2
50	3	5	23	0	0	0	1	0	0	0	1	2
51	3	5	17	0	0	1	0	0	0	1	0	1
52	3	5	25	0	0	0	1	0	0	1	0	1
53	3	5	29	0	0	0	1	0	0	1	0	2
54	3	5	27	1	0	0	0	0	0	0	1	2
55	3	5	23	0	0	1	0	0	0	1	0	2
56	3	4	14	0	0	0	1	0	1	0	0	1
57	3	4	13	0	0	0	0	1	0	0	1	2
58	3	4	31	0	0	0	0	1	0	0	1	1
59	3	4	28	0	0	0	0	1	0	0	1	1
60	1	2	6	0	0	0	1	0	0	1	0	2
61	1	2	7	1	0	0	0	0	0	1	0	2
62	1	2	8	0	0	1	0	0	0	1	0	1
63	1	2	9	0	0	1	0	0	0	0	1	2
64	1	2	6	0	0	0	1	0	0	1	0	2
65	1	2	7	1	0	0	0	0	0	1	0	2
66	1	2	8	0	0	1	0	0	0	1	0	1
67	1	2	9	0	0	1	0	0	0	0	1	2

ORIGINAL DATA - CONTINUED

WULN2	WRIFLEX	WREXT	NECK1	NECK2	NECK3	TTWIS1	TTWIST2	TTWIS3	TTWIST4	TFLX/EX1	TFLX/EX2	TFLX/EX3
2	2	2	0	1	0	1	0	0	0	0	1	0
2	2	1	0	0	1	1	0	0	0	0	1	0
1	2	1	0	1	0	1	0	0	0	0	1	0
2	2	2	0	1	0	1	0	0	0	0	1	0
1	2	1	1	0	0	1	0	0	0	0	1	0
2	2	2	0	1	0	1	0	0	0	0	1	0
2	2	2	0	0	1	1	0	0	0	0	0	1
1	2	1	0	0	1	1	0	0	0	1	0	0
2	2	2	0	1	0	1	0	0	0	0	1	0
2	2	2	1	0	0	1	0	0	0	0	1	0
2	2	2	0	1	0	1	0	0	0	1	0	0
2	2	2	0	1	0	1	0	0	0	0	1	0
2	2	2	0	0	1	1	0	0	0	0	0	1
1	2	1	0	0	1	1	0	0	0	1	0	0
2	2	2	0	1	0	1	0	0	0	0	1	0
1	2	1	0	0	1	1	0	0	0	1	0	0
1	2	1	0	1	0	1	0	0	0	0	0	0
1	2	1	0	0	1	1	0	0	0	1	0	0
1	2	1	0	1	0	1	0	0	0	0	0	0
2	2	1	1	0	0	1	0	0	0	1	0	0
2	2	1	0	1	0	1	0	0	0	0	1	0
1	2	1	1	0	0	0	0	1	0	0	1	0
2	2	1	0	0	1	1	0	0	0	1	0	0
1	2	2	0	1	0	1	0	0	0	1	0	0
1	2	2	0	1	0	1	0	0	0	1	0	0
2	2	2	0	0	1	0	1	0	0	0	1	0
2	2	1	0	1	0	1	0	0	0	1	0	0
2	2	1	0	1	0	1	0	0	0	1	0	0
1	2	1	1	0	0	1	0	0	0	0	1	0
1	2	1	0	0	1	0	1	0	0	0	1	0
1	2	1	0	0	1	1	0	0	0	1	0	0
2	2	1	0	0	1	1	0	0	0	0	0	1
2	2	2	0	0	1	1	0	0	0	0	0	0

ORIGINAL DATA - CONTINUED

2	2	1	0	0	1	1	0	0	0	0	0	0
2	2	1	0	1	0	1	0	0	0	0	1	0
2	2	1	0	0	1	1	0	0	0	0	0	0
2	2	2	0	0	1	1	0	0	0	0	1	0
2	2	2	0	1	0	1	0	0	0	0	0	0
2	2	2	0	0	1	0	0	0	1	0	0	1
1	2	2	0	0	1	0	0	1	0	0	0	1
2	1	1	0	1	0	1	0	0	0	0	1	0
1	2	2	0	0	1	1	0	0	0	0	0	0
2	2	2	1	0	0	0	1	0	0	0	0	1
2	2	2	1	0	0	0	1	0	0	0	0	1
1	2	1	0	0	1	1	0	0	0	0	0	0
1	2	2	0	0	1	1	0	0	0	0	1	0
1	2	1	0	1	0	0	0	1	0	0	0	0
2	2	1	0	0	1	1	0	0	0	0	0	1
2	2	1	0	1	0	1	0	0	0	0	0	0
2	2	1	0	0	1	1	0	0	0	0	0	0
1	2	2	0	0	1	1	0	0	0	0	1	0
1	2	1	0	1	0	0	0	1	0	0	0	0
2	2	2	0	0	1	1	0	0	0	0	0	0
2	2	1	0	0	1	1	0	0	0	0	1	0
2	2	1	0	1	0	1	0	0	0	0	1	0
2	2	1	0	1	0	1	0	0	0	1	0	0
2	2	1	0	1	0	1	0	0	0	0	1	0
1	2	1	0	1	0	1	0	0	0	0	1	0
1	2	2	0	0	1	1	0	0	0	0	1	0
2	2	2	0	1	0	1	0	0	0	0	1	0
2	2	2	0	0	1	1	0	0	0	0	0	1
1	2	1	0	0	1	1	0	0	0	1	0	0
2	2	2	0	1	0	1	0	0	0	0	1	0
2	2	2	0	1	0	1	0	0	0	0	1	0
2	2	2	0	0	1	1	0	0	0	0	0	1
1	2	1	0	0	1	1	0	0	0	1	0	0
2	2	2	0	1	0	1	0	0	0	0	1	0

ORIGINAL DATA - CONTINUED

TFLX/EX4	STRAITLE	SITLEG	KNEELLE(	DURATIOI	NOGLOVE	LITEGLOV	HEAVGLO	LOWFOR(	MEDFOR(	HIGHFOR(	POWGRIP	CHUCKGF
0	0	0	1	20	1	0	1	0	1	0	1	1
0	1	0	0	30	1	0	0	1	0	0	1	1
0	0	0	1	120	1	0	0	1	0	0	1	0
0	1	0	1	120	1	0	1	0	1	0	1	0
0	1	0	0	240	0	0	1	0	0	1	1	0
0	1	0	0	30	1	0	0	0	1	0	1	1
0	1	0	0	2	0	0	1	0	1	0	1	0
0	1	0	0	30	1	0	0	0	1	0	1	0
0	1	0	0	180	1	0	0	0	1	0	1	0
0	0	1	0	120	1	0	0	0	1	0	1	0
0	1	0	0	30	0	1	0	0	1	0	1	0
0	1	0	0	30	1	0	0	0	1	0	1	1
0	1	0	0	2	0	0	1	0	1	0	1	0
0	1	0	0	30	1	0	0	0	1	0	1	0
0	1	0	0	180	1	0	0	0	1	0	1	0
0	1	0	0	0.25	1	1	0	1	0	0	1	0
1	1	0	0	0.33	1	1	0	1	0	0	1	0
0	1	0	0	0.25	1	1	0	1	0	0	1	0
1	1	0	0	0.33	1	1	0	1	0	0	1	0
0	1	0	0	0.5	1	0	0	1	0	0	1	0
0	0	0	1	0.5	1	0	0	1	0	0	1	0
0	1	0	0	60	0	0	1	0	0	1	1	0
0	1	0	0	2	0	0	1	0	1	0	1	0
0	1	0	0	5	0	0	1	0	1	0	1	0
0	1	0	0	10	0	0	1	0	1	0	1	0
0	1	0	0	15	0	0	1	0	1	0	1	0
0	1	0	0	15	0	0	1	0	1	0	1	0
0	1	0	0	120	0	0	1	0	1	0	1	0
0	1	0	0	2	0	0	1	0	1	0	1	0
0	0	1	0	20	1	0	0	1	0	0	1	0
0	1	0	0	2	0	0	1	0	1	0	1	0
0	1	0	0	30	0	0	1	1	0	0	1	1
1	1	0	0	10	0	0	1	0	0	1	1	0

ORIGINAL DATA - CONTINUED

1	0	1	0	5	1	0	0	1	0	0	1	1
0	1	0	0	5	1	0	0	0	0	1	1	0
1	1	0	0	15	0	0	1	0	0	1	1	0
0	0	1	0	60	0	0	1	1	0	0	1	0
1	1	1	0	75	1	0	0	0	1	0	1	1
0	0	1	0	90	0	0	1	1	0	0	1	0
0	1	0	1	30	0	0	1	0	1	0	1	0
0	0	0	1	10	0	0	1	0	1	0	1	0
1	1	0	0	5	0	0	1	0	1	0	1	0
0	0	0	1	20	0	0	1	0	1	0	1	0
0	1	0	1	30	0	0	1	0	1	0	1	1
1	1	0	0	5	0	0	1	0	0	1	1	0
0	0	1	0	150	1	0	0	1	0	0	1	0
1	1	0	0	15	0	0	1	0	0	1	1	0
0	1	1	1	30	1	0	0	1	0	0	1	1
1	1	0	0	10	0	0	1	0	0	1	1	0
1	1	0	0	20	0	0	1	0	0	1	1	0
0	0	1	0	60	1	0	0	1	0	0	1	0
1	1	0	0	15	0	0	1	0	0	1	1	0
1	1	0	1	30	0	0	1	1	0	0	1	0
0	1	0	0	5	1	0	0	0	0	1	1	1
0	1	0	0	5	1	0	0	0	1	0	1	1
0	1	0	0	2	0	0	1	1	0	0	1	0
0	1	0	0	90	0	0	1	0	1	0	1	0
0	1	0	0	2	0	0	1	0	1	0	1	0
0	1	0	0	20	0	0	1	0	1	0	1	1
0	1	0	0	30	1	0	0	0	1	0	1	1
0	1	0	0	2	0	0	1	0	1	0	1	0
0	1	0	0	30	1	0	0	0	1	0	1	0
0	1	0	0	180	1	0	0	0	1	0	1	0
0	1	0	0	30	1	0	0	0	1	0	1	1
0	1	0	0	2	0	0	1	0	1	0	1	0
0	1	0	0	30	1	0	0	0	1	0	1	0
0	1	0	0	180	1	0	0	0	1	0	1	0

ORIGINAL DATA - CONTINUED

PENCILGR	KEYGRIP	GOODTER	FAIRTERR	POORTERR	
0	0	0	0	1	
0	1	0	1	0	
0	1	0	1	0	
0	0	0	0	1	
0	0	0	0	1	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
1	0	0	1	0	
1	0	0	1	0	
1	0	0	1	0	
1	0	0	1	0	
0	1	1	0	0	
0	1	1	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	1	0	

ORIGINAL DATA - CONTINUED

0	0	0	1	0	
0	0	0	1	0	
0	0	0	0	1	
0	0	0	0	0	
0	0	1	0	1	
0	0	0	0	0	
0	0	0	0	1	
0	0	0	0	1	
0	0	0	0	1	
0	0	0	0	1	
0	1	0	0	1	
0	0	0	0	1	
0	0	0	0	0	
0	0	0	0	1	
0	0	0	1	0	
0	0	0	0	1	
1	1	0	1	0	
0	0	0	0	0	
0	0	0	0	1	
0	0	0	0	1	
0	1	0	0	1	
0	1	0	0	1	
0	0	0	0	0	
0	0	0	0	0	
0	0	0	0	0	
0	1	0	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	
0	0	1	0	0	

APPENDIX C

GENERAL CLUSTERING ALGORITHM, MODIFIED DATA-TRAINING AND VALIDATION SETS

GENERAL CLUSTERING ALGORITHM

"The following are the steps in the agglomerative hierarchical clustering algorithm for grouping N objects (items or variables):

1) Start with N clusters, each containing a single entity and an NxN symmetric matrix of distances (or similarities) $\mathbf{D} = \{d_{ik}\}$.

2) Search the distance matrix for the nearest (most similar) pair of clusters. Let the distance between "most similar" clusters U and V be d_{UV}.

3) Merge clusters U and V. Label the newly formed cluster (UV). Update the entries in the distance matrix by (a) deleting the rows and columns corresponding to clusters U and V and (b) adding a row and column giving the distances between cluster (UV) and the remaining clusters.

Repeat steps 2 and 3 a total of N-1 times. (All objects will be in a single cluster after the algorithm terminates.) Record the identity of clusters that are merged and the levels (distance or similarities) at which merger takes place." (Johnson and Wichern, 2002).