A MULTIVARIATE STATISTICAL MODEL FOR WHOLE-BODY RELATED MUSCULOSKELETAL DISORDERS

Harish Yerneni, Montana State University
Robert J. Marley, Montana State University
Robert J. Boik, Montana State University
Edward L. Mooney, Montana State University

yerneni@trex2.oscs.montana.edu

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 how non-repetitive 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 both upper and lower part of the arms, 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 should be trained with larger data sets for better prediction and more robust applications. However, the current model may be useful for predicting 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.

.

INTRODUCTION

Cumulative Trauma Disorders (CTDs) are 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. 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. CTDs is an ever increasing cost to business and industry in terms of reduced productivity, lost work time, high insurance and disability claims.

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.

METHODS

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. 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.

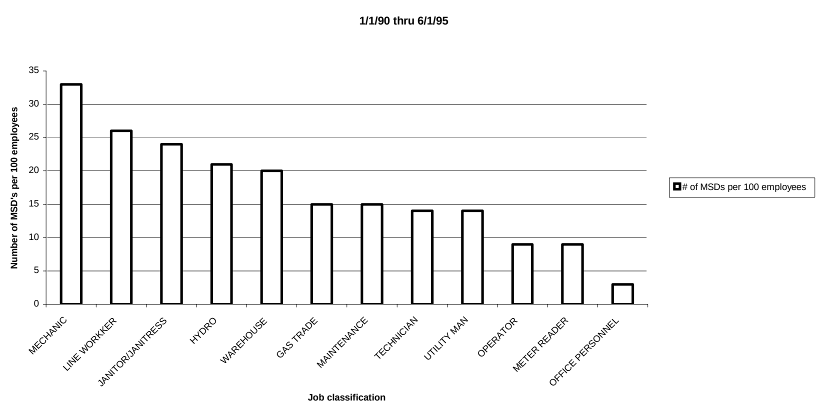

Figure 1. Number of MSDs per 100 employees by job classification

Figure 1 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 reasons is both line worker and meter reader the jobs 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 movements in their activities. Gas trade job category is chosen for analyzing medium risk. These three categories are thus chosen for developing a generalized model for whole-body related MSDs in a non-repetitive environment.

An ergonomic/work-methods analysis was then performed upon five key activities within these jobs as listed in Table 1. Activities were further broken down into 31 required tasks (e.g., climb pole, make connection, shovel, cut pipe, etc.) as listed in Table 2.

Table 1. Five key activities

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

Table 2. 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

All the tasks listed in Table 2 were video taped and analyzed to find the associated fundamental ergonomic variables. For a given task, the videotape 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. Eighteen fundamental work elements thus found were quantified into 48 levels that represent the variables for further analysis. The fundamental work elements and their different levels of measurement are documented in Table 3.

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). However, the metrics were slightly changed as stated in Table 3. Data were collected on 67 different tasks performed within 5 different activities. For example, the angle for upper arms was measured using a goniometer after pausing the videotape at the point where subject shows maximum upper arm deviation. The same procedure was adopted for other variables also. Power grip and wrist flexion (>20 deg) were found to be present in all 67 observations. Hence, these variables are considered as constants and eliminated from further analysis.

Table 3. Fundamental work elements and their 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			

RESULTS AND DISCUSSION

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). 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 3 to 5. All binary variables (3, 4, 5, 9, 10-12, 16-18) have 0 and 1 as their level values. 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):

$$Z = (r-1)/(M-1), \qquad \text{(Equation 1)}$$

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 4.

Table 4. 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). Cluster analysis is used to find natural groupings of whole-body related musculoskeletal variables associated with CTDs from Table 3.

3.1.1.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.

The data (67 observation) was divided randomly into two sets: training (80% of the data set, 50 observations) and testing (20% of the data set, 17 observations). 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 and d(ij) is the distance between two clusters i and j (MINITAB, 2000)..

The dendogram obtained from MINITAB is shown in Figure 2. In order to discover the cutting point of the dendogram, cluster analysis was executed without specifying a final partition. The step that reveals an abrupt change in similarity 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 Figure 2).

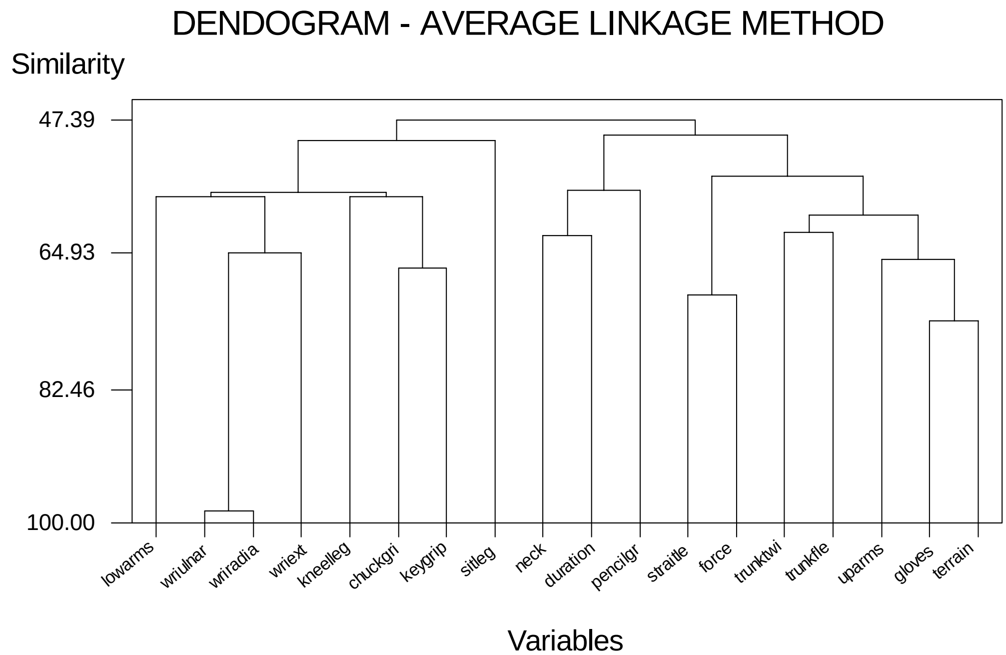

Figure 2. Dendogram

Five clusters obtained at this break point are listed in Table 5. 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 5. 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

3.1.2.Principal component analysis. 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). 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. Cluster score is the sum of the product of cluster variable weights and their standardized values as shown in Equation 2. The values of mean and variance for Z_i are obtained from the training set.

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

In matrix notation, $\mathbf{Z} = (D)^{-1/2}(\mathbf{X}\text{-}\mu)$

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

$$\mathbf{D} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots\ldots, \sigma_P^2)$$

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

$$Y_i = \mathbf{e}_i'\,\mathbf{Z} \qquad \text{(Equation 2)}$$

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

and $\rho_{Y_i, Z_k} = \sqrt{\lambda i}\, \mathbf{e}_{ik}$ where $i,k = 1,2\ldots\ldots,p$

In this case, $(\lambda_1, \mathbf{e}_1)$, $(\lambda_2, \mathbf{e}_2)$,…………., $(\lambda_k, \mathbf{e}_k)$ are eigenvalue-eigenvector pairs for ρ, with $\lambda_1 \geq \lambda_2 \geq \ldots.. \geq \lambda_p \geq 0$.

For the clusters listed in Table 5, 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 6. Since all the variables within each cluster have the same sign (see Table 6) 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 6. Cluster Scores

Cluster number	Variables	Weights (e_1')
1	Upper arms, Trunk twist, Trunk flexion, Gloves, Terrain	0.347, 0.400, 0.363, 0.525, 0.559
2	Lower arms, Wrist ulnar, Wrist radial, Wrist extension, Kneeling, Chuck grip, Key grip	0.226, 0.589, 0.575, 0.254, 0.215, 0.340, 0.215
3	Neck, Duration, Pencil grip	0.601, 0.645, 0.472
4	Standing, Force	0.707, 0.707
5	Sitting	1

Modeling and Validation

Ordinal logistic regression is 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. Response variable (R) 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, corresponding to meter reader, gas trade and line worker job categories, respectively. Training and testing data sets were used for modeling and validation, respectively. A model was fit (MINITAB, 2000) 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

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

where

k	=	the number of distinct values of the response = 3
$\ln(X_{ij}/(1-Xi_j))$	=	logit function
$X_j{}'$	=	A vector of predictor variables associated with the jth covariate pattern (cluster variable vector)
j	=	1, 2,….,67
β	=	A vector of coefficients associated with the predictors
θi	=	The constant associated with the ith distinct response

In other terms,

$$X_{ij} = P(R \le i / X_j{}') = \mathrm{e}^{\theta i + Xj'\beta} / (1 + \mathrm{e}^{\theta i + Xj'\beta}) \qquad \text{(Equation 3)}$$

Using prior probabilities, the model predicted accurately 92% (46 out of 50 observations) and 76.5% (13 out of 17 observations) of training and testing data sets respectively. From Table 7, three incorrectly predicted 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 7 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 authors believe 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). Also, 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 ($p<0.05$). These significant clusters are useful in the understanding of physical job stress in these environments

.

Table 7. Ordinal logistic regression - testing data results

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

The reasons for the low accuracy in prediction rate of testing data set may be due to smaller training data set. More importantly, it failed to predict responses 1 and 2 (See Table 7) because of too few observations in those categories in the training set.

Web Application

A web application has been developed to the ordinal logistic regression model discussed in the section 3.2. The web application will allow the user to enter the inputs and generate the response (risk level) along with the inputs. The database is not relational; it has only a single entity (table) called ‘Observation’ with 19 attributes; 18 input variables and a serial number, the primary key. The table ‘Observation’ satisfies first four normal forms. Hence, The user can view, insert, update and delete the records (observations) without any anomalies. All the inputs were validated. The flow of data in the web application is shown in Figure 3. 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.

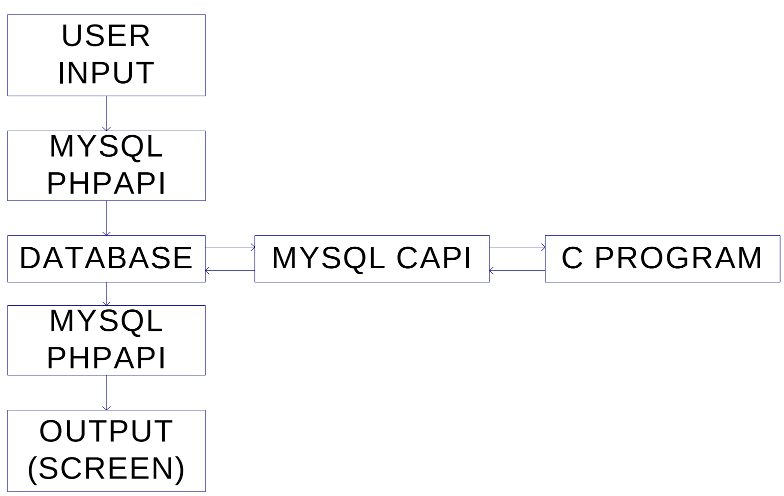

Figure 3. Data flow diagram

The web application was aimed at the novice user and the user is given instructions on how to operate the database and the model. The user communicates to the application mainly through drop-down menus, text boxes and the application communicates back to the user through tables, graphs, and text formats. It is easy to navigate through the application because hyperlinks are provided at each and every page. The user can express any concerns and questions to the administrator of the application, using the feedback form.

CONCLUSIONS

The model 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. The web application to the ordinal logistic model will be very useful to nurses, doctors etc.

REFERENCES

1. 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.
2. 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
3. Johnson, R.W., & Wichern, D.W. 2002. *Applied Multivariate Statistical Analysis* (5th Ed.). New Jersey: Prentice Hall.
4. Kaufman. L., & Rousseeuw. P.J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley & Sons, Inc.
5. 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
6. Marley, R.J., & Kumar, N. 1996. An improved musculoskeletal discomfort assessment tool. *International Journal of Industrial Ergonomics* , 17: 21-27.
7. McAtammey, L., & Corlett, E.L. 1993. RULA: A Survey Method for the Investigation of Work-Related Upper Limb Disorders. *Applied Ergonomics*, 24: 91-99.
8. McCullagh, P., & Nelder, J.A., 1992. Generalized Linear Models. Chapman & Hall.
9. MINITAB (2000): *Electronic documentation of Stat Guide*, Version 13. State College, PA: Minitab Inc.
10. 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.
11. Naderi, B., & Ayoub, M.M. 1989. *Cumulative hand trauma disorders*. Lubbock, TX: Institute for Ergonomics Research, Texas Tech University.
12. Salter, R.B. 1970. *Textbook of disorders and injuries of the musculoskeletal system*. Baltimore: Williams & Wilkins.

13. 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.

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