Proceedings of Jaffna University International Research Conference (JUICE-2012), pp. 252-257, published: March 2014, Sri Lanka

Statistical Modelling for Injuries among the Soccer Players in Jaffna

S. Gobinath† and S. Arivalzahan
Department of Mathematics and Statistics, University of Jaffna, Sri Lanka
†sigo1985@gmail.com

Abstract: Many studies have reported the frequency and types of injuries in soccer players. However, a few have assessed the relationship of playing position, climate, psychological effects and infra structure facilities with injury. The objective of the study was to develop a statistical model for injuries among the soccer players in Jaffna. The observations on the soccer injury-related variables, age, Body Mass Index (BMI), playing position, years of experience, training method, equipment, ground facilities, climate and psychological effect were collected from a simple random sample of 125 individual soccer players from Jaffna district. These nine variables were grouped into three factors using the factor analysis techniques. The first factor (TIF) consists of Training method and Infra structure Facilities; the second factor (AE) consists of Age and years of Experience and the third factor (BMP) consists of BMI and playing Position. It was interestingly noted that three variables in the first factor were common for a soccer team and the variables in the next two factors were associated with an individual player. Significant associations were found between injuries and standardized BMI groups as well as playing positions. The odds of getting injury was significantly increased from back to forward direction in soccer field. Logistic regression analysis was used to fit a model for soccer injury in a team by considering the factor TIF and another logistic regression model was fitted for soccer injury for an individual player considering the rest of the two factors AE and BMP. Further, a Sample Maximum Likelihood Discriminant Function (SMLDF) was developed to classify a soccer player as injured or not. Using the SMLDF and based on an individual soccer player's observations on the above nine variables, a soccer player can be advised about the risk of getting injury in future.

Keywords: Discriminant Analysis, Factor Analysis, Odds Ratio, Principal Component Analysis, Soccer Injury.

I. Introduction

Soccer is one of the most popular team sports in the world and continues to provide many young people an opportunity for healthy exercise. The total number of participants, approximately three million is registered in high school or youth soccer associations [18]. Moreover, soccer is one of the growing sports with reported increases in participation ranging from 11.4% to 21.8% annually [18]. Injury rates in youth soccer, known as football outside the United States, are higher than in many other contact/collision sports [6, 10, 13, 16]. The US Consumer Product Safety Commission estimated that there were 186,544 soccer related injuries in 2006 [5, 19]. Most of the soccer injuries are found to be in ankle, thigh and knee, and they mainly occur during training [4, 8, 11, 17]. Many studies have reported the frequency and types of injuries in soccer players. However, a few have assessed the relationship of playing position, climate, psychological effects and infra structure facilities with injury.

In Jaffna district, there are a lot of sports teams; even though soccer is one of the major and popular games in Jaffna district. Jaffna district soccer team leads in the national level and a reasonable number of players are participating in soccer with their maximum involvement. Every year most of the soccer teams are participating in number of tournaments and friendly and practice matches at various levels. Among the soccer players in Jaffna district, the knowledge about sports injuries is very low; knowledge about preventive measures and first aid of common sports injuries is also inadequate to maintain themselves in good health conditions and avoid injuries.

II. Objectives

The main objective of this study was to develop a statistical model for the injuries among the soccer players in Jaffna. The high risk position in soccer injuries could be found from the playing positions goalkeeper, back, half and forward. Furthermore, low risk injury group in soccer could also be identified from the internationally standardized body mass index groups–underweight, normal weight, overweight and obesity. Another objective was developed to advice the players individually in future whether they will injure or not based on the above mentioned nine variables by constructing a sample maximum likelihood discriminant function.

III. Literature Review

Giannotti *et al.* [9], developed a logistic regression model for head injuries in soccer in Canada using the variables of age and sex and he concluded that the highest proportion of head injuries was amongst males, especially children. Giannotti *et al.* [9], used logistic regression model and odds ratio techniques for acute soccer injuries in Canadian children and youth and concluded that head, face and neck were the injury risk parts and occurring by the contact with the surface. Turbeville *et al.* [20], fitted a logistic regression model for injuries in high school football players considering the variables of a player's characteristics and concluded that body mass index and strength were not associated with risk of injury, but playing experience and a history of injury were associated. Kucera *et al.* [14], undertook a multivariate generalized Poisson regression model for the history of injuries among youth soccer and found that injury history was high risk. A prospective study of injuries encountered during participation in a summer soccer camp for youths aged 6 through 17 years revealed an injury incidence of 10.6 per 1000 hours for girls and 7.3 per 1000 hours for boys [3]. Gabbe *et al.* [7], undertook a logistic regression analysis and concluded that there were significant differences between the age groups with respect to body weight, body mass index. Mechelen *et al.* [15], reported that previous injury with an odds ratio of 9.41 was one of the strongest independent predictors for sports injury. Powell and Barber-Foss [18] reported in a population based study that in high school soccer 8.4% of injuries to boys and 10.4% of injuries to girls were re-injuries. Arnason *et al.* [2], fitted a multivariate model and concluded that age, previous injury were the main factors. Willer *et al.* [21], investigated on head injuries by logistic regression and concluded that young children were more likely to have a head injury than older children.

IV. Data Collection

None of the soccer teams in Sri Lanka maintains their records especially about the soccer injuries. Since the researcher could not get any professionally soccer injury related recorded data from any associations or federation allied to soccer, the observations on the soccer injury-related variables, Age (A), Body mass index-BMI (B), playing Position (P), years of Experience (E), Training method (T), Equipment (Eq.) Ground facilities (G), Climate (C) and Psychological effect (Ps) were collected from a simple random sample of 125 individual soccer players from Jaffna district via self-administrated questionnaire which covered all specific objectives in this research. The limitation in the above self-administrative questionnaire was that a soccer player might not correctly identify his injury type as well as his injured body part. However, injury type and injured part of the body were not considered as response variables. In this research the response variable is binary, which is whether a specific player is injured or not.

V. METHODOLOGY

(i). Factor analysis has been performed to identify the important latent factors those influencing in soccer injuries.
(ii). Categorical data analysis techniques such as odds, odds ratios, confidence intervals and contingency table were performed [1].
(iii). Logistic regression analysis was undertaken to model for soccer injury for a team as well as to an individual soccer player [1].
(iv). Discriminant analysis was performed to classify the soccer players as injured or not-injured.

VI. DATA ANALYSIS

Factor analysis was conducted to identify the important factors influencing in soccer injuries. Principal component analysis method was used to identify the factors and varimax rotation was used for the purpose of interpretation [12]. KMO (Kaiser-Myer-Olkin), multicollinearity, singularity, and Bartlett's test of sphericity were conducted to examine whether the data were appropriate for factor analysis [2, 12]. Results indicated that KMO analysis was greater than 0.5 and Bartlett's test of sphericity was highly significant $(p < 0.0001)$ which say that the sample was adequate to perform a factor analysis.

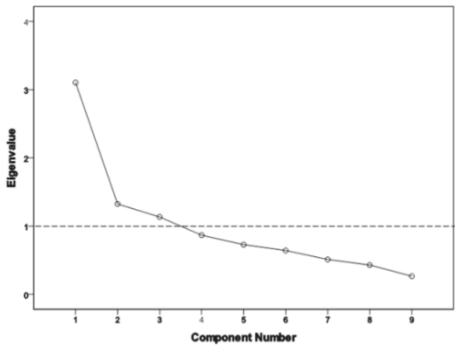

Figure 1: Scree plot

An elbow is obtained in the plot in Figure 1 at about three which indicates that only three eigenvalues are greater than one. Others are relatively small and about same size. The first three principal components effectively summarize the total sample variance. Therefore, it was decided to retain the first three factors.

Table 1: Rotated component matrix

Variables	Component		
	1	2	3
Age	- 0.136	0.911	0.015
BMI	- 0.287	0.149	0.669
Experience	- 0.174	0.897	0.002
Position	0.114	- 0.115	0.841
Equipment	0.813	- 0.204	- 0.019
Ground	0.765	- 0.167	- 0.153
Training	0.838	- 0.041	0.000

Table 1 gives the simpler structure and the highlighted values are higher values for each variable. Both Table 1 and Figure 2 show that the first factor (TIF) consists of Training method and Infra structure Facilities; the second factor (AE) consists of Age and years of Experience and the third factor (BMP) consists of BMI and playing Position. It was interestingly noted that the three variables in the first factor were common for a soccer team and the variables in the rest of other two factors are associated with individual players.

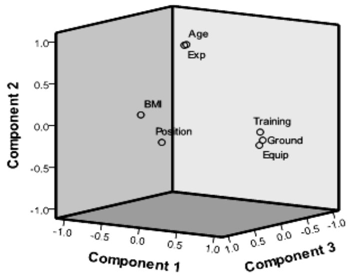

Figure 2: Component plot in rotated space

The above Figure 2 indicates that age and years of experience are coincided. It interprets that age and years of experience have high correlation [12].

Table 2: Odds for BMI

BMI	Injury		Odds of Injury
	Yes	No	
Underweight	48	09	5.33
Normal weight	62	26	2.38
Overweight	14	01	14.00

Table 2 shows the estimated sample odds for the standardized BMI groups. Pearson Chi-squared statistic between BMI and injury was $\chi^2 = 6.1330$ and $p - value = 0.0466 (< 0.05)$. Hence, there exists a valid association between BMI and injury.

Table 3: Odds for playing position

Position	Injury		Odds of Injury
	Yes	No	
Goalkeeper	10	09	1.11
Back	28	22	1.27
Half	24	02	12.00
Forward	62	03	20.67

From Table 3, it could be noticed that the odds of getting injury was significantly increased from the back to forward position. Since Pearson Chi-squared statistic $\chi^2 = 35.1854$ and $p - \text{value} = 0.0001\ (< 0.05)$, there exists a valid association between playing position and injury.

Table 4: Odds ratios and 95% confidence intervals for playing positions

Playing Position	Odds Ratio	95% C.I for Odds Ratio
Goalkeeper to Back	0.87	(0.3025, 2.5189)
Goalkeeper to Half	0.09	(0.0169, 0.5071)
Goalkeeper to Forward	0.05	(0.0123, 0.2332)
Back to Half	0.11	(0.0225, 0.4981)
Back to Forward	0.06	(0.0170, 0.2228)
Half to Forward	0.58	(0.0912, 3.6936)

Table 4 shows the estimated sample odds ratios and the corresponding 95% confidence intervals for the playing positions. There was no significant association between the odds of getting injury and goalkeeper to back position. Moreover, there exists significant association between other playing positions and injury.

Logistic regression analysis was used to fit a model for soccer injury for a team by considering the factor TIF and another logistic regression model was fitted for soccer injury for individual player considering other two factors AE and BMP. Likelihood Ratio test was used for the model selection and forward selection method was used for the variable selection.

The logistic regression model for individual soccer player is,

$$\text{logit(p)} = -8.7455 - 3.8396\times\text{B} + 6.6076\times\text{E} + 7.5079\times\text{P} + 1.2394\times\text{BE} - 3.4389\times\text{PE}$$

and a logistic regression model for soccer team is,

$$\text{logit(p)} = -0.0749 + 8.4609\times\text{Eq} - 9.6920\times\text{G} + 0.8312\times\text{T} - 2.5835\times\text{TEq} + 2.1573\times\text{GT}$$

The sample maximum likelihood discriminant function (SMLDF) which was obtained using all the nine variables is given below.

$$\hat{y} = 0.1333\times\text{A} - 1.5178\times\text{B} + 1.8012\times\text{E} + 1.6892\times\text{P} - 2.7874\times\text{Eq} + 3.8508\times\text{G} + 0.0858\times\text{T} - 0.1455\times\text{C} + 0.4126\times\text{Ps}$$

where the midpoint between the means of the two groups is $\hat{m} = 4.1526$. Hence, if the discriminant rule, $h(x) = \hat{y} - \hat{m} \geq 0$, then the soccer player will be allocated to the injury group and if $h(x) < 0$, then he will be allocated to the non-injury group.

The above SMLDF's efficiency was tested using our training sample of 125 soccer players. The confusion matrix of the classification results is given in Table 5. The results show that 27 individuals were misclassified which means that the overall error rate is 21.6%.

Table 5: Confusion matrix

		Classified memberships	
		Injury	Non-injury
True group memberships	Injury	73	21
	Non-injury	6	25

VII. CONCLUSION

We have identified the Latent factors that influence the soccer injuries. The first factor was related to team's performance which consists of training and infrastructure facilities-ground and equipment, other two factors were related to an individual's performance which consists of age, BMI, years of experience and playing position. Injury risk for the soccer players increased from back to forward direction in the soccer field. That is in the order of goalkeeper, back, half and forward. To find the probability of injury risk two separate models have been constructed for a team and for an individual player. Further, using our SMLDF based on an individual soccer player's observations on the all nine variables, a soccer player will be able to get an attention about the risk of getting injury in future.

REFERENCES

[1]. Agresti, A., "An Introduction to categorical data analysis", 2nd ed., Wiley, pp. 99–120, 2007.

[2]. Arnason, A., Sigurdsson, S.B., Gudmundsson, A., Holme, I., Engebretsen, L. and Bahr, R., "Injuries in Football; Risk factors, injury mechanisms, team performance and prevention", *Oslo Sports Trauma Research Center*, 2004.

[3]. Backous, D.D., Fried, K.E., Smith, N.J., Parr, T.J. and Carpine W.D., "Soccer Injuries and their Relation to Physical Maturity", *Am J Dis Child*, pp. 839–842, 1988.

[4]. Barker, H. B., Beynnon, B.D. and Renstrom, P.A., "Ankle injury risk factors in sports", *Sports Med.*, pp. 69–74, 1997.

[5]. Bill, H., "Analysis Finds Injury Rates are Highest among Kids under 15", *WebMD Health News*, Jan. 25, 2010.

[6]. Emery, C. A., Meeuwisse, W.H. and Hartmann, S.E., "Evaluation of risk factors for injury in youth soccer", *Am J Sports Med.*, pp. 1882–1891, 2005.

[7]. Gabbea, B.J., Bennell, K.L. and Finch, C.F., "Why are older Australian football players at greater risk of hamstring injury", *Journal of Science and Medicine in Sport*, Vol. 9, Issue. 4, Pub Med. pp. 327–333, 2006.

[8]. Le Gall, F., Carling, C. and Reilly, T., "Injuries in young elite female soccer players: An 8-season prospective study," *Am J Sports Med.*, pp. 276–284, 2008.

[9]. Giannotti, M., Al-Sahab, B., McFaull, S. and Tamim, H., "Epidemiology of acute soccer injuries in Canadian children and youth", *Pediatr Emerg Care*, 2011.

[10]. Guskiewicz, K.M., Marshall, S.W., Broglio, S.P., Cantu, R.C. and Kirkendall, D.T., "No evidence of impaired neurocognitive performance in collegiate soccer players", *Am J Sports Med.*, pp. 157–162, 2002.

[11]. Hockenbury, R.T., "Forefoot problems in athletes", *Med. Sci. Sports Exerc.*, pp. 448–458, 1999.

[12]. Johnson, R.A., "Applied Multivariate Statistical Analysis," 4th ed., USA: Prentice-Hall. Inc., 1998.

[13]. Junge, D.J.A., "Football injuries and physical symptoms: A review of the literature", *Am J Sports Med.*, 2000.

[14]. Kucera K.L., Marshall, S.W., Kirkendall, D.T., Marchak, P.M and Garrett, W.E., "Injuries in youth female football," *Oslo Sports Trauma research center*, 2005.

[15]. Mechelen, V., Twisk, W., Molendijk, J., Blom, A., Snel, B., Kemper, J. and Han, C.G., "Injury history as a risk factor for incident injury in youth soccer," *Br J Sports Med.*, 2005.

[16]. Nielsen, A. B., "Sports injuries in adolescents' ball games: soccer, handball and basketball," *Br J Sports Med.*, 1990.

[17]. Orchard, J. and Seward, H., "Epidemiology of injuries in the Australian Football League, seasons 1997–2000," *Br J Sports Med.*, pp. 39–44, 2002.

[18]. Powell, J. and Barber-Foss, K., "Sex-related injury patterns among selected high school sports", *Am J Sports Med.*, pp. 385–391, 2000.

[19]. *Soccer Industry Council of America*, North Palm Beach, FL33408, Sep. 1995.

[20]. Turbeville, S.D., Cowan, L.D., Owen, W.L., Asal, N.R. and Anderson, M.A.., "Risk Factors for Injury in High School Football Players," *The American journal of sports medicine*, 2003.

[21]. Willer, B., Dumas, J., Hutson, A. and Leddy, J., "A population based investigation of head injuries and symptoms of concussion of children and adolescents in schools," *Inj Prev.*, pp. 144–148, 2004.