Predicting Sports Readiness and Injury Risk of UCONN Athletes Using Novel Statistical Metrics

Kristin D. Morgan[1], Catie Dann[2], Yannis K. Halkiadakis[3]

[1]Biomedical Engineering University of Connecticut-Storrs

[2] Sports Medicine, University of Connecticut-Storrs

[3]Biomedical Engineering, University of Connecticut-Storrs

Abstract

The University of Connecticut-Storrs (UCONN) athletes as part of a comprehensive strength and conditioning program are run through a battery of physical test to assess their readiness for sport participation. The present study addresses athletes from four varsity level programs (Division 1- Men & Women Basketball and Women's Lacrosse and Soccer) who range in age from 17-24 years old. Prior to commencing team play these athletes undergo a sport specific pre-season training regimen. After pre-season conditioning programs, the aforementioned test battery is performed with the intent of developing quantitative metrics for assessing athletes readiness for immediate sport participation.

Here our overarching goal is to identify a class of metrics that can enhance current in-house evaluative tools for assessing an injured athlete's readiness for a return to sport activity. Additionally, it is our desire to develop robust statistical regression-based models that characterize sport readiness and isolates useful classification metrics that predicts the potential injury risk an athlete may face if certain quantitative performance levels are not met. This work is significant because it exploits the use of low-cost exercise equipment available to most college and university athletic programs for this purpose.

Key Words: Sports, regression, classification, logistic, metrics, athletes

1. Introduction

At UCONN a myriad of physical assessment data is taken on athletes to determine their readiness for sport participation at the beginning of a season, monitor in season conditioning and recommend when an athlete should return to sport activity after an injury. The goal of the present inquiry is to identify an appropriate set of metrics- associated with extensor and flexor muscle group[14] functions- that can improve our ability to both predict injury risk[5] and substantiate an athlete's return to sports recommendation.

A commonly used ratio for such purposes is the limb symmetry index [9-11] that employs normalized percent differences to assess between limb function. However, this metric has met with limited success because both limbs are often observed to atrophy as a result of injuries. One finds that simply matching limb function does not necessarily lead to optimum outcomes. Additionally, many of the constructed limb symmetry indices are based on quantities that do not capture dynamic responsiveness or are not highly correlated with the specific muscle function required of a given sport.

Here we are proposing the use of novel engineering-based metrics to capture dynamical behavior and enhances our ability to predict injury risk. The significance of this work is the recognition that relatively low-cost Biodex[13] exercise equipment available to most college and university athletic programs may be used to perform such evaluations.

2. Methods

Varsity athletes from four major Division I sports programs at UCONN (Women & Men Basketball, Women's Lacrosse and Soccer) were participants in this study. Their ages ranged from 17-24 years and prior to completing this test protocol had all undergone a sport specific preseason conditioning program. The body mass index (BMI) for this test population varied between 18-30. Each subject was run through a specific battery of exercises on a Biodex[2,15] machine (see Fig.1) that measured the core strength of extensors, flexors, abductors, and adductors large muscle groups. Identical measurements were taken on both preferred and non-preferred limbs.

Flexors and extensors[13] are a class of muscle groups found throughout the body that control joint movement which play a major role in accelerating large powerful muscles like the quads (knee extensors) vs hamstrings (knee flexors), or biceps (elbow flexors) vs triceps (elbow extensors). Specifically, flexor muscles work to decrease the angle between bones on two sides of a joint while extensors increase the angles between limb members. An extensor movement is primarily a backward motion, with the notable exception of the knee joint. Abductor and adductor[13] muscles reside within the hips and thighs and work in conjunction with each other to promote sideways leg movement. They either direct motion away (abductors) from the body's midline or toward (adductors) it.

With the Biodex system respective core measurements are obtained with the subject in a sitting position. Each subject's measurements are captured over time and provide a dynamic assessment of an individual muscle group strength and explosiveness [1,4] (power). Abbreviated Biodex subjects test data results are described in Table 1 and summarizes both individual physical features and extensor data captured on a participant's preferred (P) limb. The non-preferred limb's injury status is also provided in Table 1 and was only used to assess injury frequencies. We anticipated that Biodex[3,15] generated data when combined with subject specific physical features could yield robust metrics that would enhance our ability to predict and classify injury risk. Furthermore, it was also an expressed desire of ours to devise scaling metrics that had a physic or engineering underpinning.

3. Dimensionless Regression Models

Clearly, the test subjects involved in these studies are from an engineering perspective generating rotational torque[12,13], rotational kinetic energy, and potential energy. Hence our initial modeling objective focused on developing normalized expressions for these respective quantities. Our initial step involved dividing the experimental quantities for torques and associated energies by appropriate normalizing values to yield the scaled/dimensionless equations presented below:

<u>Scaled Torque Ratio</u>: $Y = \text{Torque}/2(w\, h_b)^2 m_b$ Eqn. 1

Scaled Energy Ratio: $X=h_b\, w^2/2g$ Eqn. 2

Observe that both scale ratios are normalized using an appropriate combination of the athlete's height (h_b), mass (m_b) and generated angular speed (w). The final analytical step required constructing models based on these derived scaled equations. The present effort discusses results based only on extensor muscle measurements (60% extension). Other muscle groups will be addressed in subsequent follow-on studies. Scatterplots (see Fig. 2a) of selected scaled ratios hinted at a potential model structure and related parametric dependency. It was apparent from that review that a family of regular hyperbolas (see Eqn. 3) were sufficient for data fitting. Such hyperbolas have asymptotes that are perpendicular to each other and align with the respective axes.

$$Y= C/[X+B], \text{ where } B=C-0.05 \qquad \text{Eqn. 3}$$

Eqn. 4 a logarithmic transform of Eqn. 3 produces a consolidated view of the data as depicted in Figure 3. A close inspection of that data reveals there are a family of such curves that are parametrically linked through the variable, C or HCONN which spans from 0.05 to 0.25. Figure 2b highlights this fundamental dependency. The largest deviations from this model occurs when the energy ratios are quite small and where negligible fluctuations in energy values can dramatically alter the torque ratio.

$$\ln Y= -\ln (1+ X_1), \text{ where } X_1 = (X-0.05)/C \qquad \text{Eqn. 4}$$

It is remarkable that such a simple model captures the complex relationship between the energy and torque ratios. The model's R-squared of 0.94 is an excellent goodness of fit value for a relationship with relatively few predictor variables.

4. Classification Models

The above modeling effort suggest that the follow variables: Y, X, C, and the BMI may be useful for classification purposes. We included the body mass index (BMI) along with this group since it is a familiar and commonly known metric that the athletic community often employs to assess overall fitness levels. The current classification approach involves using a subset of the three derived dimensionless variables to characterize injury data. This injury dataset consists of three types: un-injured individuals, single- injured individuals and dual-injured participants whose injuries occurred over the course of a sport season. Several standard classification techniques are available within the MATLAB's Statistic and Machine Learning Toolbox[8] for such analyses and will be used to formulate both supervised and semi-supervised machine learning routines that will be employed for future classification screenings involving all muscle groups. In the present context our classification studies were limited to an examination of the extensor muscle groups.

4.1 Dual Variable Classification Models (Graphical Approach)

The results summarized in this section highlight classification criteria derived using the regression predictors identified in section 3. Our best model was a 2-parameter classification scheme employing the C and BMI parameters as screening variables. It was observed for individuals where $C>0.08$ and $BMI>24$ a very high injury frequency was noted. For this 2-parameter model the classification performance metrics were accuracy=0.800, sensitivity=0.580, and specificity=0.844. Higher order models incorporating additional classification factors did not lead to any improvement in model efficacy. The confusion matrix for this 2-parameter case is depicted in Table 2. For the 17-24 age group a $BMI > 24$ was selected as a classification cut-off value because it represents the onset level for obesity for this age group population. If one removes the C parameter from the 2-parameter model significant degradation occurs. Obviously, one may conclude that the BMI index does not address critical dynamical factors associated with the rapid functioning of the extensor muscle groups; this singular variable along is not adequate for screening purposes. These results suggest that when the energy and torque production rates are insufficient at a given BMI level, the likelihood of injury will increase. Hence, one does not have the ability to control the movement at a joint or stabilize the weight bearing load it is experiencing under dynamics conditions.

4.2 Single Variable Classification & Gender Identification (Graphical Approach)

The upper whisker (UW) obtained from a boxplot (Fig. 4) of respective sport group's energy ratio (ER) was found to be an effective metric for limited classification purposes. This particular metric captured group differences that existed across various sport groups. The bar graphs of Figure 5 also revealed that ER values greater than 1.17 were rarely encountered with female athletes and suggest a major gender difference with regard to energy utilization. Employing this metric (ER=1.17) as a gender classifier yielded an accuracy of 91%. The related injury frequency data of Table 1 suggest a possible strong link between small ER values and high injury rates. Although larger ER values may generate less frequency injuries[11] they may be of a more severe nature. All non-injured subjects had an Energy Ratio<0.70. For all subjects with an Energy Ratio > 1.0, single injuries occurred more frequently than multiple injuries (93%). The respective histograms of Figure 6 summarizing this pattern and estimated injury probability distributions are given by the following respective expressions: (Exponential, $f(ER)= \exp{-\lambda\, ER}$; $\lambda(0) =2.78$, $\lambda(1)= 0.525$ and $\lambda(2)=1.18$.

5. Logistic Model for Injury Prediction Risk

A logistic model was devised to provide the odd for predicting a sport related injury[11] (see Table 3). This binary logistic regression model was generated using the predictor variables (TR, ER,C and BMI) defined in section 3 to determine the probability of individuals residing in an injured or un-injured state. A stepwise routine (MATLAB[8]'s 'stepwiseglm'), was used to isolate the final model structure and estimate the parameters of Equation 5. The outputs of this logistic regression model are binary responses variable 1 or 0 where 0 represents the un-injured individuals and 1 denotes the injured subjects. Note that the odds of an injury rise quickly with an increase in the energy ratio (ER). The lowest value for the odds quantity is 2.33. The generated model highlights that there is a high occurrence of injuries amongst this group of athletes and suggest that lowering energy ratios may reduce injury risk.

$$Probability\ (Y_i = 1 \text{ or } 0 \mid X = x_i) = \frac{\exp(\beta_0 + \beta_1 * x_i)}{1 + \exp(\beta_0 + \beta_1 * x_i)} \qquad \text{Eqn. 5}$$

where,

$Y_i = 1$ is the injured group
$Y_i = 0$ is the un-injured group

$X = x_i$ is the energy ratio, ER defined by Eqn. 2.

$\beta_0 = 0.55$ and $\beta_1 = 3.05$ are the estimated model parameters.

6. Conclusions

A dimensionless scaling of the Biodex system variables identified three critical variable that characterizes the energy production of extensor muscle groups. A simple regression between the energy and torque ratios was adequate for describing the dynamical behavior of the four UCONN sports teams.

The energy ratio when also coupled with the C or HCONN metric performed well as classifiers for isolating injury groups. There is a clear indication that high torque and energy ratios are associated with an increase in the odds of injury. Furthermore, it was interesting to note that very few female athletes had an ER value greater than 1.17.

Several critical questions arose as an outgrow of the current study and are briefly summarized below:

- Is the observed ER gender bound value 1.17 an inherent physiological limit?
- Are higher ERs a requirement for more explosives sports that involve significant jumping- landing?
- Is there a correlation between 'ACL' injury frequencies and low ER levels?
- Can specialized training protocols improve an athlete's ER value?

This above set of questions will be used to frame and augment the next phase of this ongoing effort which will incorporate data from other muscle groups. Bootstrapping will also be used to enhance data efficacy by generating ensemble replicates of the original dataset that will allow for non-parametric estimations of the data's distributional properties i.e. mean and standard deviation and aid in improving classification effectiveness.

The authors would also like to acknowledge the UCONN Athletic Department's support of this ongoing research collaboration.

7. References

1. Andersen LL, Aagaard P. 2006. *Influence of maximal muscle strength and intrinsic muscle contractile properties on contractile rate of force development*. European Journal of Applied Physiology 96(1):46–52. pmid:16249918.

2. Sung E-S and Kim J-H. 2018.*: Relationship between ankle range of motion and Biodex Balance System in females and males***,** Journal of Exercise Rehabilitation 14(1): 133-137.

3. Flansbjer UB and Lexell J. 2010. *Reliability of Knee Extensor and Flexor Muscle Strength Measurements in Persons with Late Effects of Polio*. Journal of Rehabilitation Medicine. 42(6):588–92. pmid:20549165.

4. Folland JP, Buckthorpe MW and Hannah R. 2014. *Human capacity for explosive force production: Neural and contractile determinants.* Scandinavian Journal of Medicine & Science in Sports.24(6):894–906. pmid:25754620.

5. Hamill J, van Emmerik RE and Heiderscheit BC, Li L. 1999 *. A dynamical systems approach to lower extremity running injuries*. Clin Biomech.14(5):297–308. doi:10.1016/S0268-0033(98)90092-4.

6. Kline PW, Morgan KD, Johnson DL, Ireland ML and Noehren B. *2015. Impaired quadriceps rate of torque development and knee mechanics after anterior cruciate ligament reconstruction with patellar 1tendon autograft.* Am J Sports Med. 43(10):2553–2558.

7. Kuhn S, Gallagher A and Malone T.1991. *Comparison of peak torque and hamstring/quadriceps femoris ratios during high-velocity isokinetic exercise in sprinters, cross-country runners, and normal males. Isokinetics Exerc Sci.* 1:138145.

8. MATLAB, version 9.10.0 R(2021a).Natick, Massachusetts: The MathWork, Inc.

9. Morgan KD, Zheng Y, Bush H and Noehren B. 2016. *Nyquist and Bode stability criteria to assess changes in dynamic knee stability in healthy and anterior cruciate ligament reconstructed individuals during walking.* J Biomech. 49(9):1686–1691.

10. Morgan KD. 2019. *Autoregressive Modeling as Diagnostic Tool to Identify Postanterior Cruciate Ligament Reconstruction Limb Asymmetry*. Journal of Applied Biomechanics. 35,388-392.

11. Morgan KD, Donnelly CJ and Reinbolt JA. 2014 *Elevated gastrocnemius forces compensate for decreased hamstrings forces during the weight acceptance phase of single-leg jump landing: implications for anterior cruciate ligament injury risk*. J Biomech. 47(13):3295–3302.

12. Perrin DH, Robertson RJ and Ray RL. 1987. *Bilateral isokinetic peak torque, torque acceleration energy, power, and work relationships in athletes and nonathletes*. *J Orthop Sports Phys Ther.* 9(5):184–189.

13. www.biodex.com/s4.

14. www.differencebetween.com/difference-between-flexor-and-extensor-muscles.

15. Zawadzki J, Boberi T and Siemienski A. 2010. *Validity analysis of the Biodex System 3 dynamometer under static and isokinetic conditions*. Acta of Bioengineering and Biomechanics 12(4).

Figure 1. Biodex System 4[13] Set-up

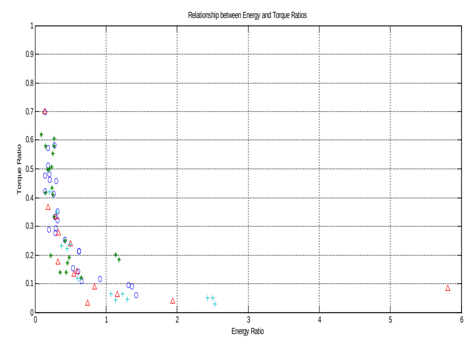

Figure 2a. Data Scatterplot

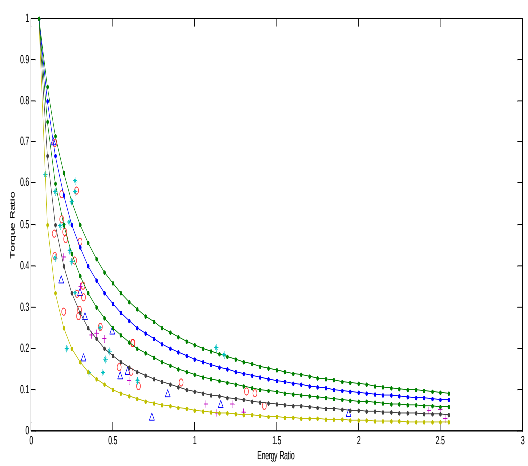

Figure 2b. Parametric Scatterplot

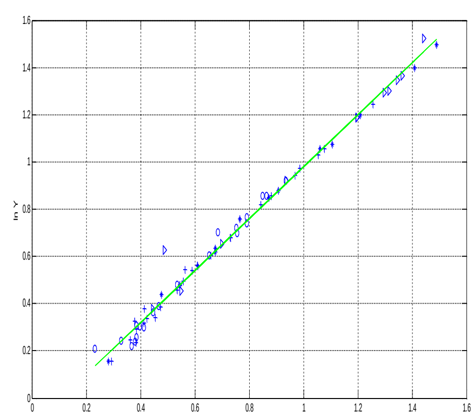

Figure 3. Linear Torque vs Energy

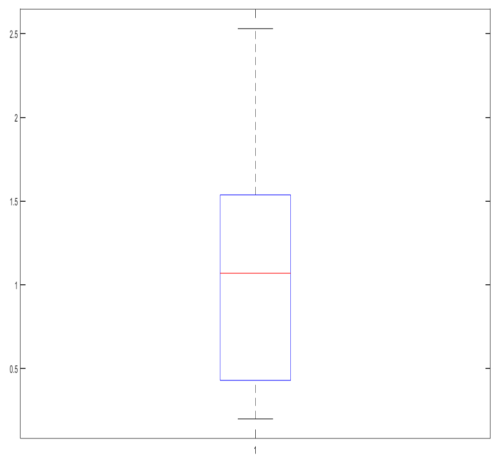

Figure 4. Boxplot of Male Athletes

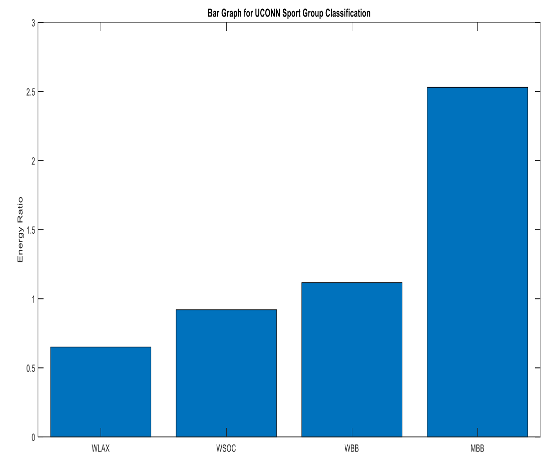

Figure 5. Bar Plot Whisker Groups

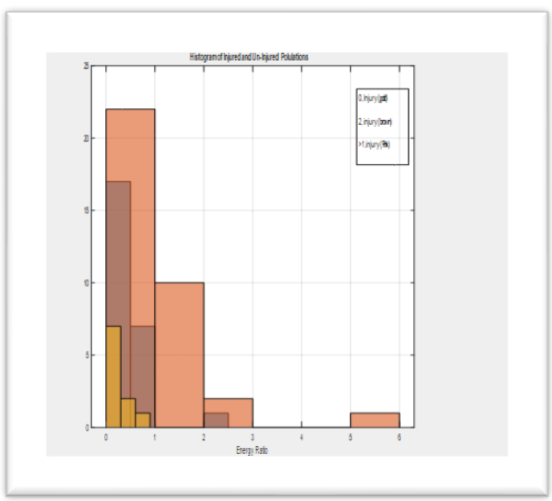

Figure 6. Histogram of Injured Groups

Table 1. Abbreviated Biodex[13] Dataset for Extensor Muscle Group: Preferred Limb

Subject	Height (cms)	Mass (kgs)	Time (s)	Θ (degrees)	Torque (N-m)	P-Injured	N-Injured	P+N Total
1	177.8	79.4	0.56	88	168	1	0	1
2	176.5	69.3	0.53	57	150	0	1	1
3	157.5	60.0	0.75	52	119	1	1	2
4	177.8	80.2	0.85	50	142	0	1	1
5	170.2	67.0	0.56	40	124	1	0	1
6	172.7	84.4	0.82	60	136	1	0	1
7	176.5	74.8	0.92	47	89	0	0	0
8	162.6	55.8	0.57	59	77	1	0	1
9	160.0	59.0	0.59	46	81	1	1	2
10	163.8	67.6	0.75	57	88	0	1	1

Table 2. Injury Confusion Matrix 2-Parameter Model

	Actual Un-injured	Actual Injured	Total
Predicted Un-injured	TP=7	FP=9	16
Predicted Injured	FN=5	TN=49	54
Total	12	58	70

Table 3. Logistic Model Probability and Odd Predictions

Energy Ratio (ER)	Injury Probability	Odds
0.25	0.787	3.60
0.50	0.888	7.93
0.75	0.944	16.87
1.00	0.973	36.04
1.25	0.987	75.92
1.50	0.994	165.67
2.00	0.998	499.00
2.50	1.000	>>1,000,000