Detroit Tigers Performance Science Questionnaire

Please answer the below questions to the best of your ability. Please adhere to the word limit of 300 words per question and save the file with your last name in the title.

1. **From a performance science perspective, describe a project you believe would add substantial value to the Detroit Tigers. Please describe the project and provide an overview of how you would complete it.**
    a. Leveraging Biomechanical Data for Pitching Analysis: Our objective is to harness the biomechanical data, as detailed in question no 5, to create inferential models that dissect the intricacies of pitching. This will encompass several key sub-projects:
        i. **Fastball Velocity Analysis:** The goal here is to construct models that illuminate how pitchers generate their fastball velocity. We aim to identify areas where velocity may be lost and offer strategies for enhancement. This is particularly valuable for nurturing young pitchers, especially those grappling with low fastball speeds. We would also like to converse with the pitching coaches to understand the variables that are easy to change vs the variables that inherently can't be changed.
        ii. **Comprehensive Velocity and Movement Modeling:** We plan to extend our analysis to all types of pitches while predicting velocity and movement. By doing so, we can estimate how a pitcher might perform with pitches they currently do not utilize, thereby optimizing their pitching repertoire. This approach could be instrumental in identifying undervalued pitchers who may benefit from altering their pitch selection.
        iii. **Fatigue Metrics Development:** Another vital aspect is the establishment of a fatigue metric for pitchers. We intend to correlate this with biomechanical data to identify factors contributing to fatigue. This analysis will extend to individual pitch types, aiming to pinpoint specific biomechanical movement or pitches that lead to increased fatigue.
        iv. **Injury Prevention Model:** Finally, we aim to develop a model focused on understanding and preventing injuries, particularly those affecting the elbow and shoulder. By identifying key factors, we can implement strategies to reduce the risk of injuries among pitchers.

    These projects collectively aim to enhance our understanding of pitching mechanics and contribute to the development and wellbeing of pitchers at various levels.

2. **Workload monitoring is a popular topic in Performance Science. With the introduction of in-game-tracking data in addition to technology used in practice, the available data for modeling workload and injury prediction is immense.**

    a. The medical director has asked you to track workload for a shortstop who is particularly prone to injury. Outline the steps you would take, including metrics of interest to generate an overall workload metric for this player.

        i. **Combining Internal, External and Player Tracking Metrics** – We'll look at Internal metrics like heart rate and rate of perceived exertion during practice and External metrics like number of plays, sprints, throws, etc. We'll also use player tracking data like first-step quickness, distance traveled, sudden turns, direction changes, and acceleration.

        ii. **Using Inertial Measurement Units for Detailed Analysis** – We'll gather biomechanical data, with or without markers, to track joint workloads during play or throwing.

        iii. **Player Surveys for Personal Insights** – Simple methods like checking how sore players feel will give us an idea of their fatigue levels from their perspective.

        iv. **Overall Workload Metric** – We'll combine all these metrics above, working with sports scientists, and coaches to figure out how each metric contributes towards overall workload (weights). We'll assign these weights to each metric. A short example is given below :

            1. Let's consider three key metrics:
                a. Number of Plays (NP): Number of defensive plays during games and practices.
                b. Throwing Intensity (TI): Intensity of throws made during practices and games.
                c. Distance Covered (DC): Distance covered by the shortstop during sprints and defensive plays.
            2. Now, assign weights to each metric based on its perceived impact on injury risk. Let's set these weights as W_np, W_ti, and W_dc, respectively.
            3. The overall workload metric (OWM) is then a weighted sum: **OWM=(Wnp×NP)+(Wti×TI)+(Wdc×DC)**
            4. These weights could be based on empirical data, expert opinions, or a combination of both.

        v. **Focus on Acute and Chronic Workload** – Using the overall workload metric, we use the **Acute:Chronic Workload Ratio (ACWR)** to compare the player’s workload over a short-period with their long-term workload average. This can help identify if there’s an increase in acute workload which can increase injury risk.

vi. **Incorporating Recovery Strategies** – We'll also factor in any recovery methods our players are using. This helps us get a complete view of their physical condition and workload management.

b. Using this workload metric, and any other data/information of value, what modeling technique(s) would you employ to assess the possibility of future injury?
After going through references listed below, I understood that with the workload metric identified above and any biomechanical or performance data collected, there are multiple modeling techniques that we can employ. **Tree-based methods** like Random Forest and XGBoost can help us understand what features affect the workload metric identified from part a (by using SHAP values.) We can also use **zero inflated binomial regression** [3] to predict risk and severity of arm injuries based on historical injury data. Even though this method in the reference focuses on MLB pitchers we can use it for position players as well. The paper also focuses on identifying modifiable (physical conditioning) and non-modifiable factors (injury history) that influence the likelihood of arm injuries.

References,

1. *Freeston J, Soloff L, Schickendantz M, Genin J, Frangiamore S, Whiteley R. In-Game Workload Demands of Position Players in Major League Baseball. Sports Health. 2023;0(0). doi:10.1177/19417381231179970*
2. *Dowling B, McNally MP, Chaudhari AMW, Oñate JA. A Review of Workload-Monitoring Considerations for Baseball Pitchers. J Athl Train. 2020 Sep 1;55(9):911-917. doi: 10.4085/1062-6050-0511-19. PMID: 32991703; PMCID: PMC7534929.*
3. *Bullock G, Thigpen C, Collins G, et al. Development of an Injury Burden Prediction Model in Professional Baseball Pitchers. IJSPT. 2022;17(7):1358-1371. doi:10.26603/001c.39741*
4. *https://medium.com/data-science-at-microsoft/data-driven-injury-prevention-in-baseball-maximizing-player-performance-and-longevity-b9fade8992ed*

3. **Players A, B, and C are available to acquire (for this exercise assume positions are inconsequential, they are all the same handedness, that they are the same age and of similar cost). Please rank them from the player you are most interested in, to least interested in. Explain your reasoning.**

Player	Bat Speed 90th %ile (mph)	EV	LA	Squared Up Percentage	Vertical Attack Angle	On Plane Efficiency	Contact Depth (ft)
Player A	86.4	89.9	23.4°	63.8%	16.0°	62.1%	1.21
Player B	82.1	90.8	17.2°	68.4%	8.7°	61.3%	0.01
Player C	71.3	89.9	10.8°	74.5%	5.9°	66.1%	-0.18

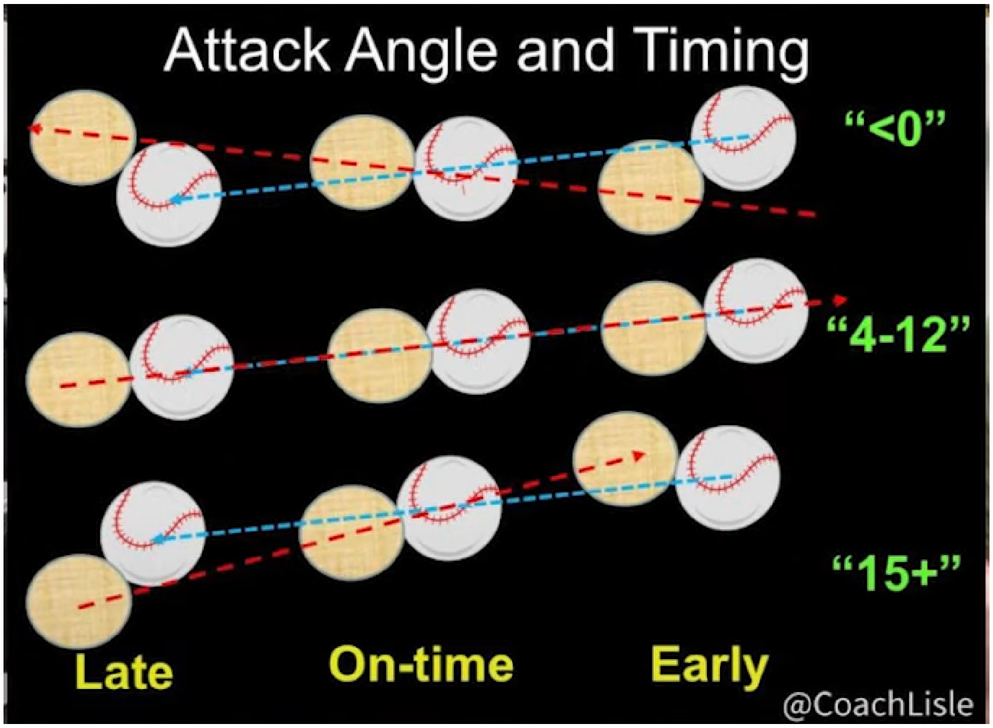
Note: Contact depth is relative to front of home plate.

Ranking all three players,

1. Player B – Player B has a great combination of exit velocity, squared up percentage and low launch angle, which suggest a good balance of power and consistency. His contact depth being 0.01 means that he is hitting the ball right at the plate and hence has good timing. By working on Player B's on-plane efficiency, we can further try to generate more bat speed from him.
2. Player A – Player A has great Bat Speed, but due to his contact depth being 1.21ft (i.e. him hitting the ball in front of the home plate) and vertical attack angle being 16 degrees, could result into rolling the ball while hitting it (which makes sense given his squared up percentage.) Therefore, based on if it's easier to change Player A's timing, we can try to increase his squared up percentage and vertical attack angle.
3. Player C – Player C has terrible Bat Speed when compared to Player A and B. Even though he is able to produce good EV due to his squared up percentage, his high on plane efficiency denotes that there is not a lot of growth we can achieve with his Bat Speed. High on plane efficiency basically denotes that he is transferring his hand speed efficiently into bat speed. Though his timing and squared up percentage being high means that if we can increase his Bat Speed, he could be an effective player.

Here are some of the references that I have used to understand the relationship between Bat Speed, Vertical Attack Angle, On Plane Efficiency and Timing (contact depth).

> On-Plane Efficiency speaks directly to the launch angle, but there are a couple of different factors involved. You've got the pitch plane coming in on a slightly downward trajectory. You hopefully have an attack angle that is positive so that they match the plane. The On-Plane Efficiency is how consistently a player is getting the sweet spot of the bat to the ball. The higher the On-Plane Efficiency, the less variability you'll see in launch angle.



4. A scout has sent you the motion tracking data (biomechanical data) from a college prospect the team is strongly considering drafting:

    a. Outline the process you would use to assess this data and provide a brief overview of what you would report back to the scout.

        i. **Data Review** – Carefully examine the motion tracking data for key biomechanical parameters, such as joint angles, acceleration, and rotation throughout the player's movements.
        ii. **Comparison to Norms** – Compare the prospect's data to established biomechanical norms for baseball players, considering factors like pitching mechanics, pitch velocity, pitch types, or bat speed, and batting mechanics.
        iii. **Identify Strengths and Weaknesses** – Pinpoint areas where the prospect excels in biomechanics, showcasing strengths in mechanics that contribute positively to performance. Identify any potential weaknesses or areas for improvement, such as inconsistencies in movement patterns or deviations from optimal biomechanical standards.
        iv. **Injury Risk Assessment** – Evaluate the data for any indicators of biomechanical issues that could increase the risk of injury. Look for irregularities that might lead to stress on specific joints or muscles.
        v. **Performance Implications** – Relate the biomechanical findings to on-field performance. For instance, how the prospect's mechanics may influence

pitch velocity, bat speed, or defensive agility. Here is where I would like to build inferential models correlating the biomechanical variables with different target variables like pitch velocity, bat speed, launch angle, etc.

vi. **Long-Term Development Considerations** – Provide insights into how the prospect's biomechanics may evolve with further physical development, considering factors like age, muscle growth, and coordination improvements. Here is where aging curves would become important as the next step in my machine learning/statistical analysis.

vii. **Recommendations** – Offer actionable recommendations for improvement, which could include targeted training regimens, drills, or adjustments to specific aspects of the prospect's mechanics.

viii. **Overall Assessment** – Summarize the overall assessment, highlighting the prospect's potential based on biomechanical data and outlining areas that might require attention.

b. Long term, what methods/models would you employ to determine if a high school prospect has the potential to generate a high velocity fastball? Assuming you have access to biomechanical data, statistics etc.

   i. We can apply the method described earlier to pinpoint the factors that contribute to a high velocity fastball. This involves identifying the specific variables that lead to increased velocity as well as any aging curve how they change with age (discussed below).

   ii. We can also look at how certain aspects change as a player gets older by using aging curves and include these in the velocity model to properly forecast the prospects potential velocity in the next 4 years. For example, in the reference below, it was identified that hip abduction and extension tend to improve with age which leads to an increase in pitching velocity and a reduction in the risk of an injury. Thus any prospect who might have a potential of growth in the hip abduction and extension areas, might show better fastball velocity in the future.

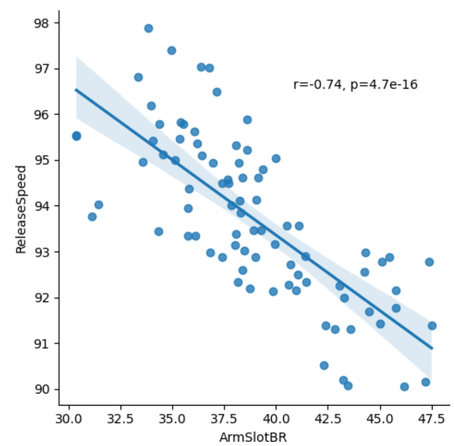
   iii. To better understand how aging affects these biomechanical measurements, we can use two-tailed Pearson correlation coefficients. This approach is similar to what was used in the study, but we'll apply it to all the variables we're examining.

   iv. By gathering biomechanical data from players of various ages, we can leverage machine learning techniques to make predictions. This could help us estimate a player's potential to throw faster fastballs, increase the break on their pitches, or even add new types of pitches to their repertoire.

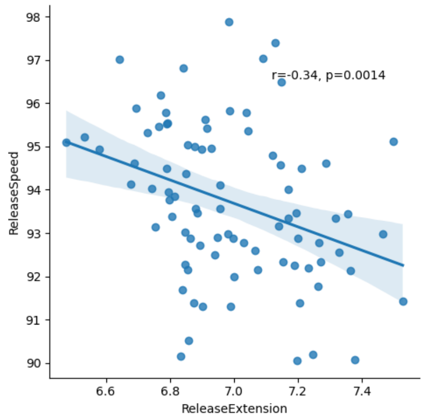
References,

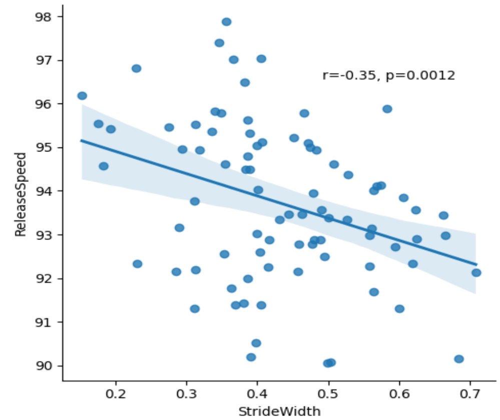
1. *Maxwell L. Albiero, Wesley Kokott, Cody Dziuk, Janelle A. Cross; Hip Strength and Pitching Biomechanics in Adolescent Baseball Pitchers. J Athl Train 1 March 2023; 58 (3): 271–278. doi: https://doi.org/10.4085/1062-6050-0074.22*

5. A coach asks you to investigate a pitcher with high variance in their FB velocities. Please produce a short self-explanatory report to answer the coach's question. Include all code/workbooks. "PitchingData.csv" contains biomechanical pitching data for that pitcher across 7 outings on a play level. "Glossary.pdf" is a description of each biomechanical metric.
    a. **Data Exploration Summary –** We analyzed 84 fastballs out of an initial set of 302 pitches, excluding some fastballs above 105 mph. These fastballs varied from 90.04 to 97.87 mph. Key findings are categorized into different aspects of the pitcher's mechanics:
        i. **Pitcher Release and Arm Data –**
            1. **ArmSlotBR** (arm slot at ball release) has a significant correlation with our pitcher's fastball (FB) velocity. A higher arm slot angle leads to lower FB velocity. [ArmAngleBR is also affecting FB velocity, but we account that in the ArmSlotBR variable]



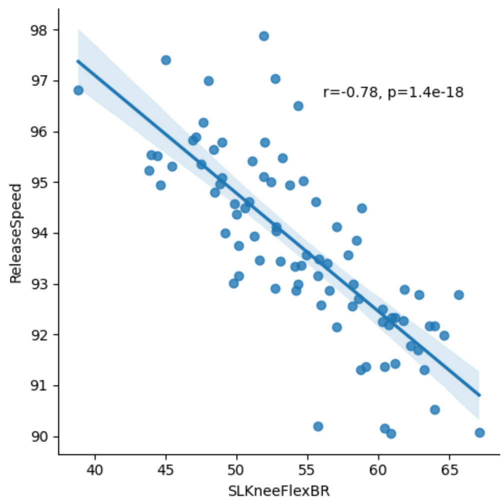
            2. There's a mild correlation between longer **ReleaseExtension**, wider **StrideWidth**, and lower FB velocity.





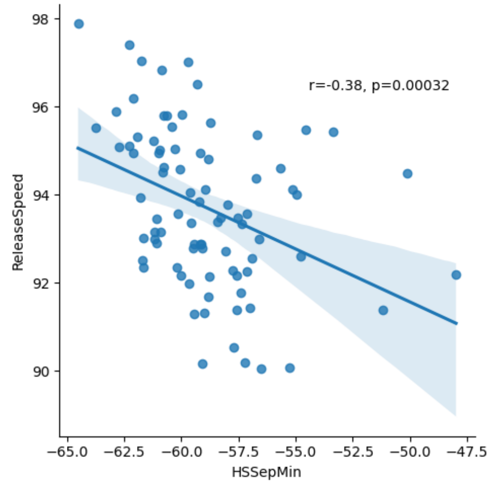
ii. **Kinematic data –**

1. **SLKneeFlexBR** (stride leg knee flexion angle at ball release) is highly correlated with FB velocity. Less flexion at ball release (more extended knee) leads to higher velocity. Our pitcher needs to be more extended with his stride leg knee at ball release.



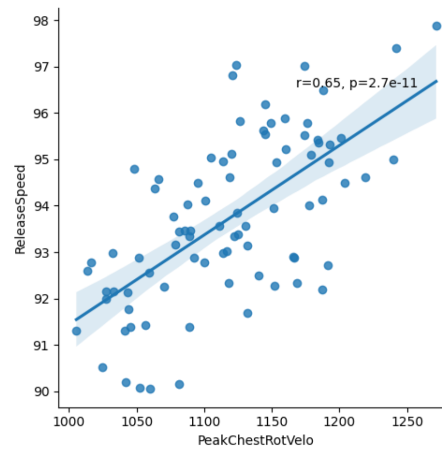
2. **HSSepMin** (hip-shoulder separation minimum angle) also shows a slight correlation. Greater counter-rotation angle correlates with

higher FB velocity.



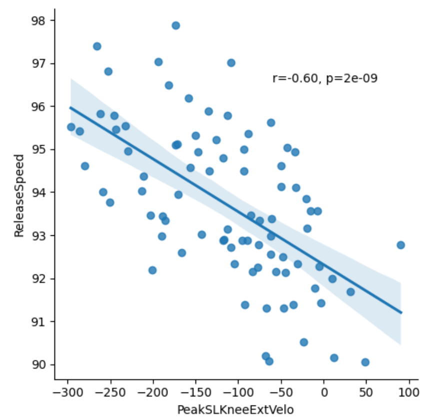
iii. **Kinetic Data** –

1. **PeakChestRotVelo** (peak chest rotational velocity) strongly correlates with FB velocity; higher chest rotation speed leads to faster pitches.

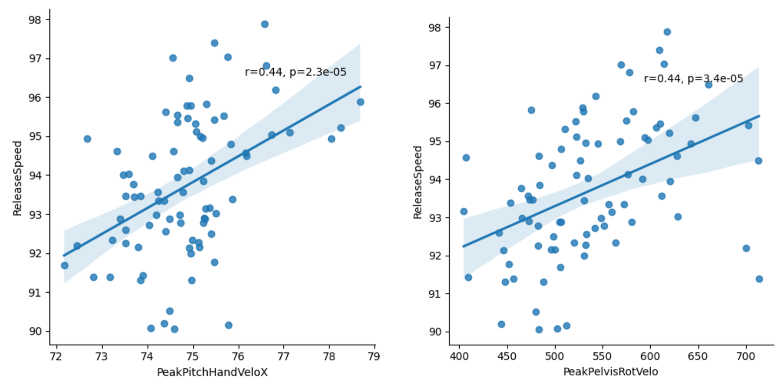


2. **PeakSLKneeExtVelo** (stride leg knee extension velocity after foot plant) is crucial. Slower knee extension after foot plant correlates with higher fastball speed. However, more knee extension is needed at ball release for optimal velocity as seen in the previous section. We want our pitcher to sink after foot plant before

eventually extending at ball release.

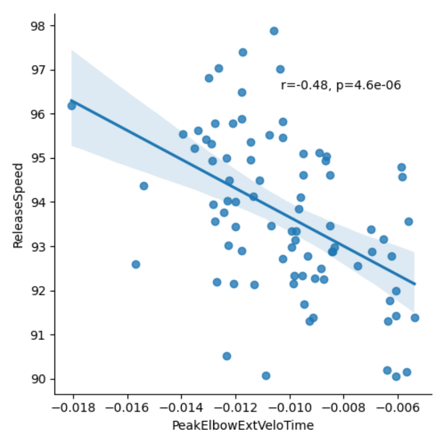


3. **PeakPitchHandVeloX** and **PeakPelvisRotVelo** are directly correlated with high fastball velocity for our pitcher.

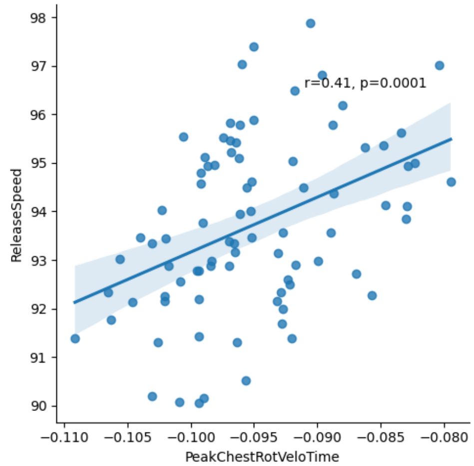


iv. **Timing Data –**

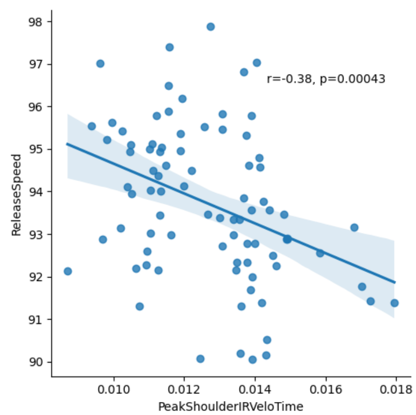
1. **PeakElbowExtVeloTime** (time of peak elbow extension velocity) shows that achieving this peak closer to ball release reduces FB velocity.



2. **PeakChestRotVeloTime** (time of peak chest rotation velocity) indicates that achieving peak chest rotation closer to ball release increases FB velocity.



3. PeakShoulderIRVeloTime (time of peak shoulder internal rotation velocity) suggests that achieving this peak closer to post ball release boosts FB velocity.



b. Modeling Outcome – We modeled an ordinary least squares model using the variables identified from our analysis, we get a good R-squared value of 0.843 with the model summary given below.

OLS Regression Results

Dep. Variable:	ReleaseSpeed	**R-squared:**	0.843
Model:	OLS	**Adj. R-squared:**	0.816
Method:	Least Squares	**F-statistic:**	31.71
Date:	Tue, 28 Nov 2023	**Prob (F-statistic):**	9.21e-24
Time:	22:49:01	**Log-Likelihood:**	-89.907
No. Observations:	84	**AIC:**	205.8
Df Residuals:	71	**BIC:**	237.4
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	69.2808	8.978	7.717	0.000	51.380	87.181
ArmSlotBR	-0.0793	0.046	-1.732	0.088	-0.171	0.012
ReleaseExtension	1.2140	0.481	2.522	0.014	0.254	2.174
StrideWidth	-2.2487	0.817	-2.752	0.008	-3.878	-0.619
SLKneeFlexBR	-0.0616	0.027	-2.311	0.024	-0.115	-0.008
HSSepMin	-0.1288	0.042	-3.103	0.003	-0.212	-0.046
PeakChestRotVelo	0.0032	0.002	1.462	0.148	-0.001	0.008
PeakSLKneeExtVelo	-0.0040	0.002	-2.025	0.047	-0.008	-6.25e-05
PeakPitchHandVeloX	0.2380	0.098	2.435	0.017	0.043	0.433
PeakPelvisRotVelo	0.0040	0.002	2.318	0.023	0.001	0.007
PeakElbowExtVeloTime	-19.0750	44.396	-0.430	0.669	-107.597	69.447
PeakChestRotVeloTime	91.3253	18.488	4.940	0.000	54.461	128.189
PeakShoulderIRVeloTime	20.3676	59.604	0.342	0.734	-98.479	139.214