Quantifying Wellness and its Relationship to Performance and Training Load

An Analysis of Player Data From Canada's National Women's Rugby Team

Sarasadat Golestaneh & Joseph Roussy

December 10th 2019

Introduction

In this project we analyzed the data from Candian National Women’s Rugby team, to quantify wellness, training load and performance for each player into summary measures. After that we investigated the relationships between the three, using both exploratory analysis (such as PCA) and regression models, such as random mixed models and general estimating equations.

Methods

We have included all of our notebooks used for analysis along with a README file that explains the order they should be ran in. Running these notebooks reproduces all of our results. We explain the methods used in this section at a high level along with our goal in using each method. As we explain a method, we point to the locations in our notebooks where it is used to ensure the reader fully understands our methodology. The notebooks also include explanations to move the reader through the narrative of our analysis.

Factor Analysis

Factor analysis is a method for automatically finding latent variables in a data set. It assumes that observed variables are actually a linear combination of a number of latent factors. Often, we are aiming to find a smaller number of factors that explain trends in our data.

This is how factor analysis differs from principal component analysis. In PCA, the components are not interpretable, however, in factor analysis, we can see which observed variables are highly correlated with which factors and group observed variables according to which factors they are correlated with by looking at the loading matrices. We can also see which factors are most important by their eigenvalues. Another technical difference between PCA and factor analysis is that in PCA, all the components are orthogonal but in factor analysis this does not need to be the case.

Thus, we use factor analysis to point us to relevant features correlated to the latent factors in the wellness data for use later in our analysis. We will also aim to combine these relevant features of wellness into a single wellness score which will let us find the high level relationships between wellness, load, and performance. A single summary score for wellness also lets us use wellness as the response in a regression model. Finding a single summary measure for wellness is the general goal of notebook #1: Wellness Exploratory and Factor Analysis. Specifically, this process unfolds in the last few sections: “Factor Analysis” and “Summarize Wellness”.

We also used factor analysis to show that the training load data can be summarised in part by the acute load and chronic loads. This led us to use acute chronic ratio, a combination of the two, as a predictor for performance as this is known to also be the case in the field of sports science. This analysis is done in the “Factor analysis” heading of notebook #2: RPE Exploratory and Factor Analysis.

To ensure factor analysis was appropriate for our data, we used the Bartlett Sphericity and Kaiser-Meyer-Olkin (KMO) adequacy tests. The Bartlett test computes the probability that the correlation matrix for the data is the identity matrix. A data like this is not appropriate for factor analysis because the data is assumed to be correlated as it represents a linear combination of latent factors. The KMO test is a measure of the proportion of variance among the variables that might be common variance, which suggest the data is due to some latent factors. These tests are performed in each of the “Factor Analysis” sections the Wellness Exploratory and Factor Analysis and RPE Exploratory and Factor Analysis notebooks.

Data Imputation

In some cases, we had too much missing data for a given measure to include in our analysis. For example, of 5011 measures of wellness, we had 4370 that were missing data for USG. This does not leave us with much to work with and we were forced to exclude USG from our analysis.

On the other hand, if there is a relatively small amount of data missing for a given measurement, we can impute it without having to disregard the entire observation, as there might be valuable data in the other measurements of that observation. A good example of this is the nutrition column of the wellness data. In this case, we had only 832 wellness observations that were missing nutrition data. It would be a shame to disregard all those wellness observations just because they are missing nutrition information. At the same time, it would be a shame to ignore nutrition in our analysis because of these missing values. Thus, we imputed the missing nutrition values by sampling from the given player’s distribution of nutrition responses.

In other words, each time we came across a missing value for nutrition, we would randomly choose from the distribution of that players nutrition responses. If that player tended to always have excellent nutrition, then we are likely to fill this missing value with excellent. Thus, we actually leverage the fact that we have many repeated measures in the data as for any instance of missing nutrition, we have many other nutrition responses for the given player to infer the value of the missing response. Using this method of data imputation, we were able to retain nutrition and those observations that were missing it in our dataset. The details on this imputation are in the “Nutrition” subheading under the “Cleaning and Visualizing Data” heading notebook #1: Wellness Exploratory and Factor Analysis.

To ensure that this method of imputation did not fundamentally alter the data, we performed factor analysis on the data with and without imputing the nutrition column. Wellness observations missing a value for nutrition would be excluded from the later factor analysis. The results of the factor analysis were the same before and after imputing the nutrition columns which tells us that we did not alter the essence wellness data by imputing. This process is outlined in the “Factor Analysis” of the Wellness Exploratory and Factor Analysis notebook.

PCA

As mentioned in the factor analysis, even though in PCA the components are not interpretable, it can be useful to apply PCA to see what clusters different data points with different features create, and if there is any visual relationship between the clusters they form. We know that we can visualize high dimensional data using a scatter plot of the first two principal components. If we do this with the relevant slices of the wellness, performance, and load datasets, we can see if there are any patterns in the relationships between players. For example, if we saw a cluster of players close to each other in the 3 PCA plots, this would suggest that the measures are related. This type of analysis is performed in notebook #6: PCA for GPS, RPE and Wellness.

We also applied PCA on wellness data separately to make sure that the summary measures for wellness make sense and players with similar summary measures end up in the same cluster. That is to say, if PCA provides a ground truth visualization of the wellness data, we can compare patterns in this visualization to patterns in our summary wellness measures. This analysis is performed in the “Visualize Wellness Using PCA” subheading in the Wellness Exploratory and Factor Analysis notebook.

We also visualized the training load data in general in the “PCA on RPE data” subheading in the RPE Exploratory and Factor Analysis notebook. The fact that the cluster in the wellness data did not carry over to the clusters in the training load data was the first hint that there is no strong relationship between wellness and training load.

Quantification of Performance

Noting that heart rate was not included in the dataset, to quantify the performance of each player in a game, two options were considered. Maximum speed and maximum acceleration. To calculate these first, the average speed and acceleration over each second (10 frames) was calculated and then for each player the maximum value was considered as her best

performance in the game. To make sure which of the two is a better measurement, we realized that the average maximum speed of players in games that they won was significantly higher than the average maximum speed in games that they lost. However, this was not the case with the acceleration.
We can see this difference in the boxplots below (there is no such difference in the accelerations though):

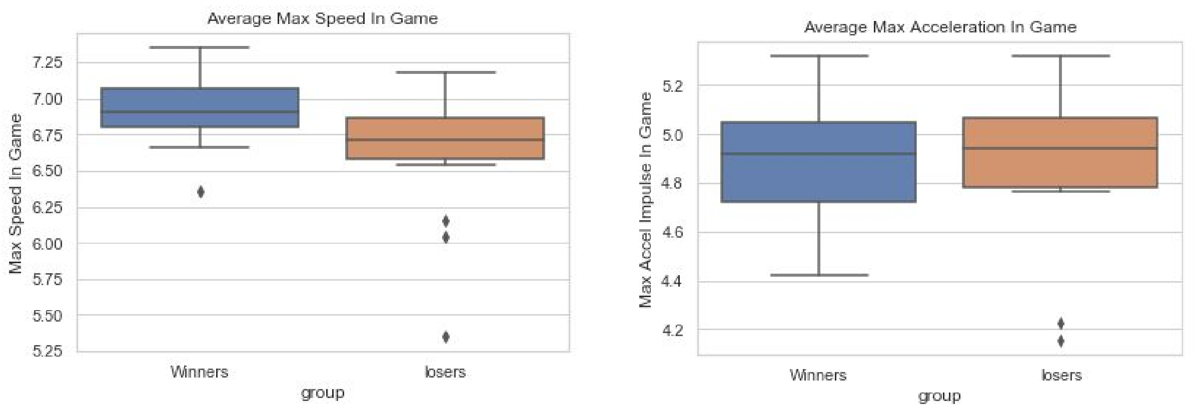


Fig. 1. Max Speed vs MAx acceleration comparison between winners and losers

Also, compared to maximum speed high acceleration was not necessarily an indicator of high performance (for example if a player changed her speed from 0 in a frame to 0.4 m/s in another frame her acceleration was suddenly 4 m/s^2 while barely even walked).
For more details refer to notebook #3: GPS Exploratory analysis.

Repeated Measure Correlation

We often want to get a basic idea of the correlation between measurements throughout our analysis. For example, in the Wellness Exploratory and Factor Analysis notebook, we wanted to see how all the measures for wellness were correlated. However, it would be incorrect to use regular correlation on this data as regular correlation assumes all observations are independent. Furthermore, we read that averaging the data over all the players would result in biased correlation estimations. We found a python package called “pingouin” that computes repeated measures correlations and also gives confidence intervals for the estimates of the correlation measure. This package is used throughout our analysis when the data contains repeated measures. For example, it is used in the “Repeated Measures Correlations” heading of the Wellness Exploratory and Factor Analysis notebook.

Visualization

Before we trained any regression models, we plotted some simple scatter plots for our single measure quantifications of performance versus wellness and performance versus training load. We did this to see if there was a clear relationship among the measure before moving to more complicated modeling. These scatter plots can be seen in the “Correlations Between Predictors and Response” heading of notebook #4: Performance vs Training Load and Wellness.

We also visualized performance, load, and wellness data as time series to see how the measures vary with time and to try and see any trends in the data that the regression models marginalized out. Here, we were looking for instances where wellness or training rises or fall in sync with performance, suggesting a relationship between them over the span of time or games.

As a specific example, for each player, we plotted performance versus acute chronic ratio over each game under the “Performance vs Training load graphs per player” of the Performance versus Training Load and Wellness notebook.

We also plotted performance vs the previous day’s load over time for each player in the “Performance vs yesterday's daily load” heading of notebook #5: Yesterday’s Load vs Performance and Wellness. This plot let us see the impact of spikes in dips in load on performance over time and is a good complement to regression analysis as it does not marginalize out the experience of individual players.

Regression

We performed many regressions during our analysis. We see regression as an important tool because it can be used to see which features act as predictors for a given target through having non-zero coefficients. Furthermore, we can see the independent effect of a given feature holding all other predictors equal. For example, this can give us insights into the specifics of the relationship between the elements of wellness and performance, where as a scatter plot of performance vs our wellness summary measure tells us about the relationship between wellness and performance in general. Given that we were working with data containing repeated measurements, we could not use standard linear regression as this has the assumption that each observation is independent. We could average the data over each player to get independent observations, but this would leave us with only a maximum of 17 data points: the number of distinct players in the data sets. This is too few data points to train a reliable model so we opted for regression tools that accept repeated measures.

In the Performance vs Training Load and Wellness notebook, we used regression to model the relationship between performance, wellness and training load in the “Regressions Of Wellness and Load on Performance” heading. Here, we used performance as the response and the relevant features of wellness along with acute chronic ratio as the predictors. Thus we can explore the relationship between performance and each element and wellness in addition to load. Our features for wellness comprised the elements of the summary wellness score: monitoring score, pain, illness, and menstruation in addition to sleep hours which is known to have an effect on performance as mentioned by Ming. In this same notebook, under the “Effect of Wellness On RPE” heading, we explore the relationship between perceived training load and wellness using RPE as the response and wellness followed by its elements mentioned above as the predictors. This regression lets us analyze the association between wellness and how players perceive their training. This is an interesting aspect of the general relationship between wellness and training load independent of performance.

In the Yesterday's Load vs Performance and Wellness notebook, we look at the effect of the previous day's daily load on the current day's performance and wellness. As such, we fit a regression with our summary measure for wellness as the response and the previous day's daily load as the predictor under the "Training load on the day before and today's wellness" heading. We also train a model with our measure for performance, max speed in game as the response and the previous day's daily load as the predictor under the "Performance vs yesterday's daily load" heading.

Model Selection

When performing regression, we needed to decide which predictors we should include in our final model that we use to infer the relationship between the predictors and the response. We knew we wanted to remove uninformative predictors from our model until we got a model with only informative predictors. However, given that p-values are not reliable in mixed models, we need some other selection criteria. We decided to use the Akaike information criterion (AIC). AIC is a measure of information about the processes that generated the data lost by a model. Thus, a lower AIC value represents a better model. We decided to follow a procedure where we remove the most uninformative coefficient given its standard error that also reduces the AIC. A coefficient would be considered uninformative if its standard error is large relative to its point estimate because we cannot be sure the coefficient is not actually 0. We continued this process until we were left with only informative coefficients.

Mixed Linear Models

Mixed linear models accept repeated measurements and model a regression using fixed effects and random effects. Random effects vary between different groups while fixed effects remain the same across all groups. In our case, all the observations of a given player for the different groups, as they are all correlated to each other. We chose to let the intercepts of the model be random effects, which means each player can have their own base level of performance for example. However, we modeled the predictor slopes as fixed effects because we wanted to get coefficient estimates that are generalizable across the entire population of players. We use mixed linear models throughout the latter halves of the Performance vs Training Load and Wellness and Yesterday's Load vs Performance and Wellness notebooks. The context of each instance is explained inline in these notebooks.

Generalized Estimating Equations

Another type of model that accepts repeated measures is called generalized estimating equation (GEE) model. GEE models require that we provide the structure of the covariance between groups. We used the Exchangeable covariance structure which assumes the measurements in a given group are correlated in the same way. In our case, that is to say the measurements for each player are correlated in the same way which makes sense. In fact, GEE with exchangeable correlation structure is analogous to a mixed model with random intercepts, however, GEE uses a different algorithm to estimate the value of the parameters. For example, in this paper[1], Hanley et. al. describe how GEE uses a quasi-liklihood approach and uses weighted combinations of observations to extract the appropriate amount of information from correlated data. If we get similar conclusions from this alternative measure of analysing correlated data, it will add more support to our conclusions. Some research indicates that GEE is actually better than mixed models, for example, in this paper[2], Hubbard et. al. argue that: "in general mixed models involve unverifiable assumptions on the data-generating distribution, which lead to potentially misleading estimates and biased inference. We conclude that the estimation-equation approach of population average models provides a more useful approximation of the truth."

We use GEE models throughout the latter halves of the Performance vs Training Load and Wellness and Yesterday's Load vs Performance and Wellness notebooks.

Assessing Quality of Fit

Once we selected a regression model we wanted to access its fit to the data. While it is true that we are not trying to use these models for predictions, they must be a reasonable fit to the data to allow us to use them to infer the relationships between the predictors and the response.

QQ Plots

One tool to access the quality of fit we used was QQ Plots. A model is said to fit the data well if the residuals are normally distributed. If this is not the case, then there must be some transformations of the predictors that can be done to make the model fit better. For example, the relationship between one of the predictors and the response might not be linear and we should add some non-linear terms as predictors.

A QQ plot transforms the standardized residuals to 2 dimensions that when plotted in a scatter plot for form a straight line with positive slope. If we see something close to this in our QQ plot, it means the residuals are normally distributed and our model should be a good fit to the data. We perform a QQ plot for each of our regressions in the Performance vs Training Load and Wellness and Yesterday's Load vs Performance and Wellness notebooks.

Standardized Residual Plots

Another measure of fit is to plot the standardized residuals versus a single predictor in a scatter plot. While QQ plots signal an issue with out model in general, these residual plots can tell us

---

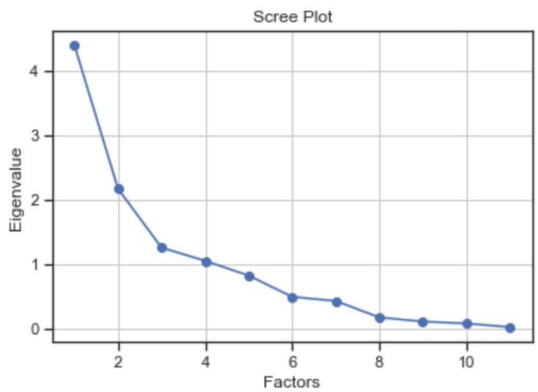
[1] Hanley, J. A. (2003). Statistical Analysis of Correlated Data Using Generalized Estimating Equations: An Orientation. *American Journal of Epidemiology*, *157*(4), 364–375. doi: 10.1093/aje/kwf215
[2] Hubbard, A. E., Ahern, J., Fleischer, N. L., Laan, M. V. D., Lippman, S. A., Jewell, N., … Satariano, W. A. (2010). To GEE or Not to GEE. *Epidemiology*, *21*(4), 467–474. doi: 10.1097/ede.0b013e3181caeb90

the predictors the issue is related to. If a model is a good fit to the data, then there should be no pattern in the standardized residuals with respect to each predictor. If this is not the case, then we can perform some kind of transformation to that predictor to make the model fit better. As with QQ plots, we perform a residual plot for each of the predictors for each of our regressions in the Performance vs Training Load and Wellness and Yesterday's Load vs Performance and Wellness notebooks.

Results

Wellness Exploratory Analysis

The scree plot for the wellness data was:



Given that the elbow occurs at 3 factors, we know we are looking for 3 latent factors in the wellness data.

Performing factor analysis using 3 factors left us with the following loadings matrix (small correlations less than 0.4 marked as N/A):

Column	**Factor 1**	**Factor 2**	**Factor 3**
Fatigue	0.99335799	N/A	N/A
Soreness	0.86689953	N/A	N/A
Desire	0.76959397	N/A	N/A
Irritability	0.88167378	N/A	N/A
SleepHours	N/A	N/A	N/A
SleepQuality	0.7844147	N/A	N/A
Pain	N/A	0.76986303	N/A

Illness	N/A	0.97599987	N/A
Menstruation	N/A	N/A	0.76518849
Nutrition	N/A	N/A	0.494086
TrainingReadiness	N/A	N/A	N/A

From this we see that Fatigue, Soreness, Desire, Irritability, and SleepQuality all relate to the first factor. This is interesting because these are the observations that made up the Monitor Score. We can interpret this as subjective wellness as they are all self reported values that, taken together, related to general wellness.

Pain and Illness are related to the second factor which we can interpret as objective wellness.

Mensuration is related to the third factor such that we can interpret it as mensuration itself. However, Nutrition is also related to this factor so we included this in the summary of wellness as we did not want to prematurely exclude features of wellness at this stage.

Given that the features mentioned above are all on different scales, we standardized them and added them up to get a single summary measure of wellness. This means that a feature is not overweighted in the summary score simply due to its scale.

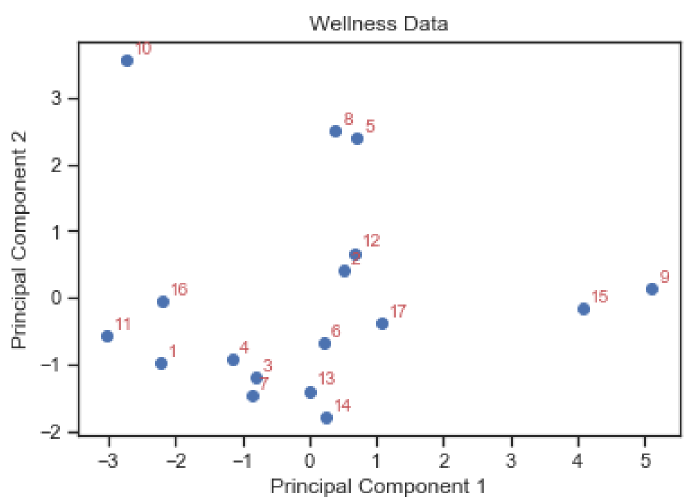
This process lead to the following summary scores for wellness for each player:

Player ID	**Wellness Score**
1	-0.319645
2	0.114238
3	0.546423
4	0.363299
5	-0.393185
6	0.731991
7	0.097324
8	-0.974241

9	2.216492
10	-3.505250
11	-0.840302
12	0.673359
13	0.481330
14	1.021188
15	1.802513
16	-0.664738
17	0.726474

Given that we always coded our wellness data using a scale where lower values represented worse wellness and vice versa, we can interpret these wellness score in the same way. Thus, player 10 has, by far, the worst average wellness on the team while player 9 has the best wellness followed closely behind in second by player 15.

We performed PCA on the entire averaged wellness dataset for each player:

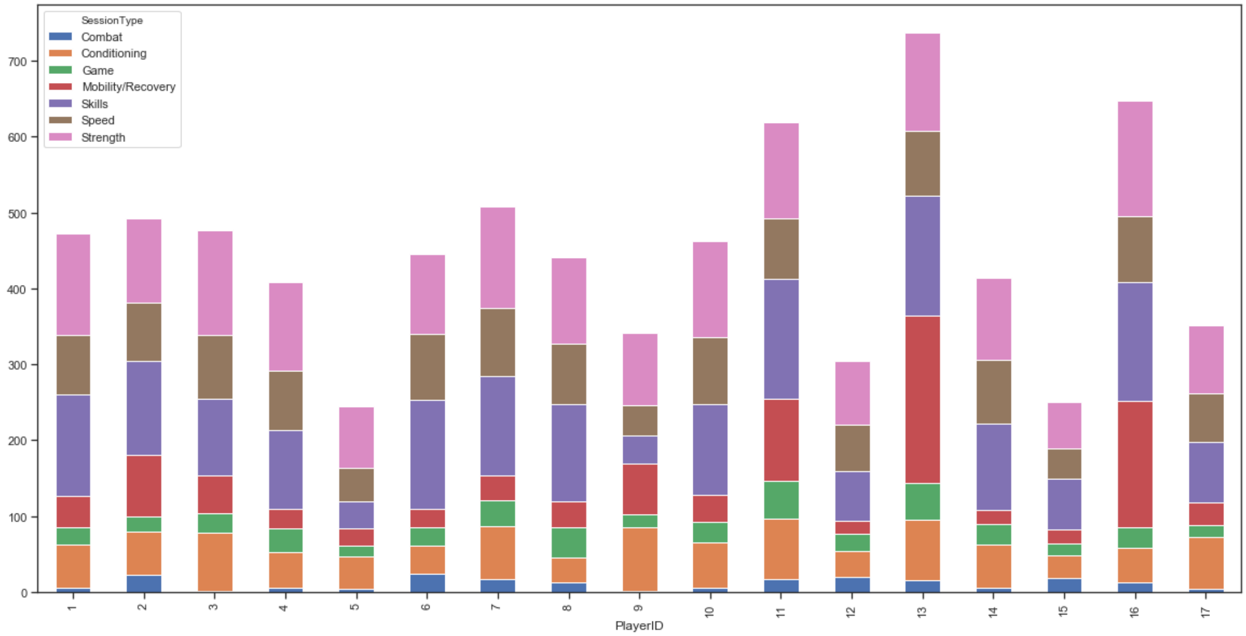


If we consider this PCA visualization as the ground truth of the wellness data, then patterns in this plot should appear in our wellness summary score if they are in fact good summaries of the wellness data.

We see that many patterns in the PCA visualization do in fact carry over. For example, player 10 is an outlier, players 2 and 12 have similar yet middling scores, players 8 and 5 have similar scores, players 15 and 9 are distinct from other players yet also similar to each other, and players 4, 3, and 7 are all similar.

RPE Exploratory Analysis

One of the interesting results that was observed after cleaning the RPE data, was the distribution of training sessions over all players which is shown below:



At a first glance, it might look like player 13 had the highest number of training sessions, however after paying attention to the session types we can see that the only session type that varies a lot among players is the Mobility/Recovery which is significantly larger for player 13 compared to other players.

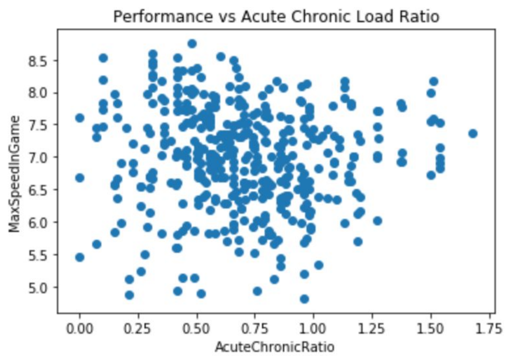
Max Speed vs Max Acceleration

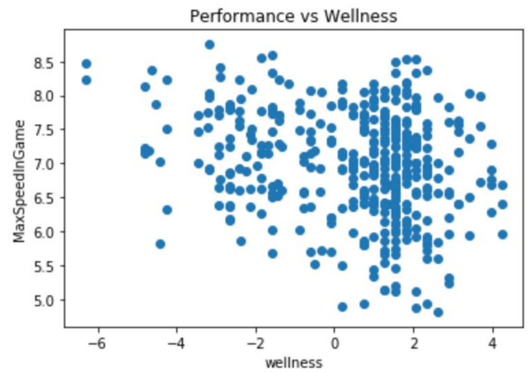
One of the interesting things that we noticed in the middle of our analysis was that whenever the average max speed of the players in the games they won was significantly higher than the average max speed of the players in the games they lost. This was shown in Fig. 1. As it was also mentioned such difference was not statistically significant for accelerations. Given that high max speed is associated with winning games, we choose to use max speed in a game as our measure for performance.

Performance as a Function of Wellness and Training Load

With out summary measure of wellness in addition to the summary measure for performance in max speed per game as well as our summary of training load in acute chronic ratio, we could explore the relationship between these 3 elements.

We started by plotting scatter plots of performance vs wellness and performance vs training load to get a high level idea of the relationships.





We see no particular pattern in either of the scatter plots above. This is a sign that there is no general relationship between performance and wellness nor performance and training load.

To see if this were true, we fitted mixed linear and GEE models to our data. In this case the response was performance in the form of max speed in game and the predictors were the wellness training load. Specifically, the predictors for wellness were: monitoring score, pain, illness, menstruation and sleep hours. This ensured we could find any effects of wellness elements obscured by the wellness summary measure. We removed uninformative predictors as per the AIC-based model selection procedure outlined in the methods section. This left a model containing only pain as a predictor. This was the model output:

```
          Mixed Linear Model Regression Results
============================================================
Model:              MixedLM Dependent Variable: MaxSpeedInGame
No. Observations:   420     Method:             REML
No. Groups:         17      Scale:              0.4103
Min. group size:    4       Likelihood:         -431.3861
Max. group size:    38      Converged:          Yes
Mean group size:    24.7
------------------------------------------------------------
                    Coef.  Std.Err.   z    P>|z| [0.025 0.975]
------------------------------------------------------------
Intercept            7.094    0.153 46.299 0.000  6.794  7.394
C(Pain)[T.1.0]      -0.202    0.111 -1.817 0.069 -0.420  0.016
Group Var            0.205    0.129
============================================================
```

The result for the GEE model for the same regression equation had similar parameter estimates.

Looking at the coefficients of our model above, we see moderately strong evidence that Pain has a negative relationship to performance. The standard error of this estimate is about half of the point estimate itself. Recalling how we coded pain, this is interesting because pain equal to 1 refers to not being in pain. Thus, better wellness with respect to pain is leading to worse performance.

We explored which players tended to play while reporting pain:

	PlayerID	AverageMaxSpeedInGame	ProbabilityOfPlayingNotInPain	GamesPlayedInPain
0	1	7.683961	0.964286	1
1	2	6.902797	0.685714	11
2	3	7.230131	1.000000	0
3	4	6.799333	1.000000	0
4	5	7.721914	0.714286	8
5	6	6.511335	1.000000	0
6	7	7.309137	0.818182	6
7	8	7.159093	0.818182	6
8	9	6.931128	1.000000	0
9	10	7.614805	0.384615	16
10	11	6.920157	1.000000	0
11	12	6.215714	1.000000	0
12	13	6.737886	1.000000	0
13	14	6.304165	1.000000	0
14	15	6.434150	1.000000	0
15	16	6.466877	0.875000	3
16	17	6.250989	1.000000	0

We see that the games played in pain come from a subset of players (2, 5, 7, 8, and 10). These players tended to play a substantial amount of their games in pain and they also tend to be higher performing players on average. In the extreme case, player 10 played most of their games in pain, yet they are one of the fastest players on the team.

There are many possible reasons for this imbalance. Some players might have different definitions of pain. For example, player 10 might have a more liberal view of pain and marks themselves as in pain often while a player like player 12, who was never in pain while playing, might have a very strict view of pain and must be very hurt to mark themselves as in pain. Alternatively, some players might not be as affected by being in pain and encouraged to play while in pain, while other players who are affected by being in pain are not played while they are hurt.

This example highlights the selection bias in our dataset. There were some choices that went into which players played a given game so we cannot see the true effect of load and wellness on performance. In other words, we might not see performance issues due to poor wellness because we do not have performance data for instances where wellness was low, since a player with low wellness might not be played. Had we done a controlled experiment where we very different factors of load and wellness and measure subsequent performance, we might be able to find stronger relationships between load, wellness, and performance. Given that many factors of wellness cannot be controlled for an experiment, we could also use data from more controlled performance testing that is completed regardless of the wellness of players so we could see the effect of all different values of wellness on performance.

Thus, our coefficient for pain is misleading as the pain data in imbalanced because the inevitable instances of poor performance while in pain are missing from the dataset. The player's max speed in a game has a lot of variance in it and it is hard to model the relationship between wellness, load, and performance using the data we have. Furthermore, the selection bias present in the data we have on performance makes it hard to draw useful conclusions from our analysis.

Relationship Between Wellness and Perceived Training Load

Now that we have explored the relationships between load and wellness on performance, we want to explore the relationship between wellness and load independent of performance. We will continue to use the same wellness predictors as before, however, for load we will use RPE as the target. This is because we want to see how wellness affected the perceived intensity of workouts.

After following our AIC-based model selection procedure we were left with the following model:

```
           Mixed Linear Model Regression Results
===========================================================
Model:              MixedLM Dependent Variable: RPE
No. Observations:   4822    Method:             REML
No. Groups:         17      Scale:              4.6860
Min. group size:    151     Likelihood:         -10595.2298
Max. group size:    452     Converged:          Yes
Mean group size:    283.6
-----------------------------------------------------------
                  Coef. Std.Err.   z    P>|z| [0.025 0.975]
-----------------------------------------------------------
Intercept         3.102    0.277 11.187 0.000  2.558  3.645
TrainingReadiness 0.004    0.002  2.718 0.007  0.001  0.007
Pain              0.247    0.120  2.062 0.039  0.012  0.482
Nutrition         0.244    0.079  3.087 0.002  0.089  0.399
Group Var         0.200    0.036
===========================================================
```

Model Interpretation

Training Readiness

There appears to be a very small positive relationship between RPE and Training Readiness but the effect is quite small. Moving training readiness from 0 to 100, which is the entire range of this predictor, only moves RPE up by 0.5, which is on the smaller size for a measure on a 0-10 scale. In terms of direction, it seems logical that higher training readiness is associated with loads that are perceived to be more intense.

Pain

Recall that pain is encoded so 1 relates to not being in pain and 0 relates to being in pain. Thus our model tells us that not being in pain is associated with an increase of 0.25 in RPE. Thus, not being in pain is associated with a small increase in perceived load.

Nutrition

Nutrition is coded as an ordinal variable with 3 levels. Moving up a single level is associated with an increase in RPE of about 0.25.

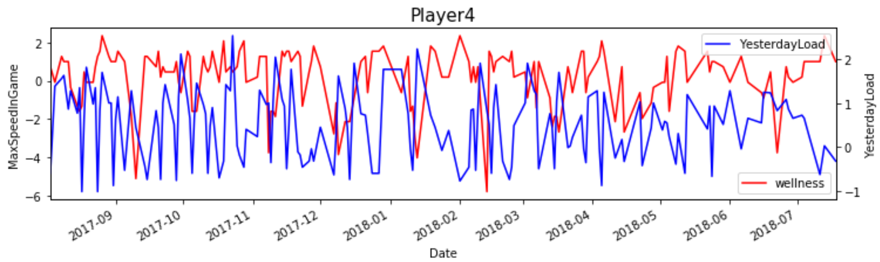
General Conclusions

In general, increased wellness in the form of training readiness, pain, and nutrition is associated with increased perceived loads. It is possible that players are only taking on harder workouts when their wellness in terms of these factors is increased, whereas they cannot push themselves as hard when their wellness is low. We originally thought decreased wellness would increase perceived load and workouts would seem hard when wellness is low. However, we see that hard workouts to not occur as often if wellness is not high. Thus, if the coaches want harder workouts to occur, they need to make sure these are occurring during periods of higher wellness for the given player.

Yesterday’s training load effect on wellness and performance

One of the interesting things that we explored was seeing if there is any specific relationship between yesterday's training load on the performance and wellness.

In general, we could see that the measurements for wellness and yesterday’s training load varied a lot with time and it was hard to see an obvious pattern at the first glance:

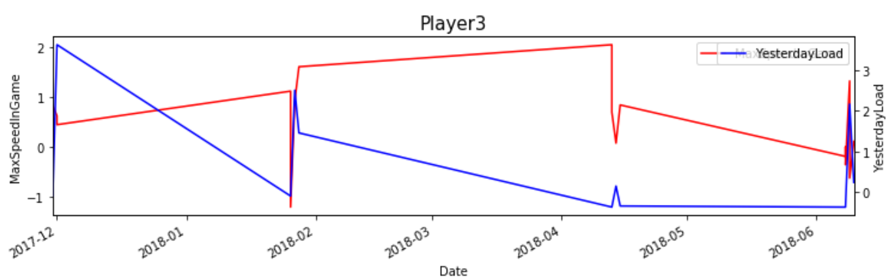


After applying mixed models on wellness and yesterday's training load we noticed that there was a statistically significant negative relationship (with a very small standard error) between yesterday's training load and wellness, which means that as the previous day's training load increases, the wellness on the next day decreases.

```
          Mixed Linear Model Regression Results
========================================================
Model:             MixedLM Dependent Variable: wellness
No. Observations:  3919    Method:             REML
No. Groups:        17      Scale:              2.5730
Min. group size:   120     Likelihood:         -7452.9572
Max. group size:   343     Converged:          Yes
Mean group size:   230.5
--------------------------------------------------------
              Coef.  Std.Err.   z    P>|z| [0.025 0.975]
--------------------------------------------------------
Intercept      0.244    0.248  0.984 0.325 -0.242  0.729
YesterdayLoad -0.134    0.026 -5.124 0.000 -0.185 -0.083
Group Var      1.031    0.231
========================================================
```

On the other hand for performance vs yesterday's load, even though some of the visualizations suggest that there should be a negative relationship between performance and yesterday's load,

after applying both GEE and mixed models our results showed that there was no statistically significant relationship between them. The following shows the visualization for one of the players as well as the table of the results for the GEE and random mixed models:



```
                               GEE Regression Results
===================================================================================
Dep. Variable:              MaxSpeedInGame   No. Observations:                  314
Model:                                 GEE   No. clusters:                       17
Method:                        Generalized   Min. cluster size:                   4
                      Estimating Equations   Max. cluster size:                  38
Family:                           Gaussian   Mean cluster size:                18.5
Dependence structure:         Exchangeable   Num. iterations:                     5
Date:                     Tue, 10 Dec 2019   Scale:                           1.008
Covariance type:                    robust   Time:                         22:02:44
=================================================================================
                    coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------
Intercept        -0.0389      0.160     -0.243      0.808      -0.353       0.275
YesterdayLoad    -0.0012      0.057     -0.021      0.983      -0.112       0.110
==============================================================================
Skew:                        -0.2973   Kurtosis:                        -0.2851
Centered skew:               -0.3635   Centered kurtosis:                0.0404
==============================================================================
```

```
           Mixed Linear Model Regression Results
============================================================
Model:             MixedLM Dependent Variable: MaxSpeedInGame
No. Observations:  314     Method:             REML
No. Groups:        17      Scale:              0.6946
Min. group size:   4       Likelihood:         -409.9228
Max. group size:   38      Converged:          Yes
Mean group size:   18.5
------------------------------------------------------------
                  Coef.  Std.Err.   z    P>|z| [0.025 0.975]
------------------------------------------------------------
Intercept         -0.050    0.169 -0.298 0.765 -0.382  0.281
YesterdayLoad     -0.002    0.053 -0.039 0.969 -0.106  0.102
Group Var          0.426    0.216
============================================================
```

Conclusions / Discussion

Summary of conclusions

Despite the limitations, our analysis has shown us that we cannot make solid conclusions about performance in terms of wellness or training load, however we can say that perceived training load increases slightly with better wellness in terms of training readiness, lack of pain, and better nutrition. If coaches want high intensity workouts to occur, they should make sure their players are doing well in these 3 areas.

Furthermore, we saw that in general, there is a lot of variance in the performance of players that cannot be attributed to the elements in this dataset.

Also, the analysis on the wellness and the day before’s training load showed that there is a slightly negative relationship between the two, meaning that as the training load increases, the next day’s wellness decreases. Such relationship however was not observed between performance and yesterday’s training load.

Discussion and Limitations

One of the main limitations was the existence of selection bias in our dataset. There were some choices that went into which players played a given game so we cannot see the true effect of load and wellness on performance. In other words, we might not have seen performance issues due to poor wellness because we did not have performance data for instances where wellness was low, since a player with low wellness might not have been played in a game. Had we done a controlled experiment, where we vary different factors of load and wellness and measure subsequent performance, we might have been able to find stronger relationships between load, wellness, and performance. Given that many factors of wellness cannot be controlled for an experiment, we could also use data from more controlled performance testing that is completed regardless of the wellness of players so we could see the effect of all different values of wellness on performance. Using such a performance measure may be less volatile and would serve as an improvement over our current performance measure.

Another limitation in this project was not having access to the heart rate data of the players. Most of the available data were subjective which made it hard to make conclusions in some cases. Having access to heart data could have given us the option to come up with better performance metric or even to apply the Banister model for the relationship between training load and performance.