JOURNAL OF
EMERGING INVESTIGATORS

Article

An analysis of junior rower performance and how it is affected by rower's features

Audrey Biller[1], Nagiza Samatova[2]
[1] East Chapel Hill High School, Chapel Hill, North Carolina
[2] North Carolina State University, Raleigh, North Carolina

SUMMARY

High-school student participation in rowing has increased over the last two decades. Due to the physical intensity of rowing, this sport may be expected to be male dominant while the female lightweight category may be expected to be non-competitive. Over the years, junior rowers may have become faster, and they may be expected to demonstrate their largest improvement by age 16. The purpose of this study was to test these expectations with hypothesis-driven experiments while presenting a glimpse into the world of high-school rowing by analyzing World Indoor Rowing Championship data. This dataset was obtained from ergometer machines used for rowing on land. The analysis showed that the participation of junior women has been exceeding the participation of junior men in the heavyweight class. We found that the key feature determining the performance of a rower is first sex and then weight class, but it is possible for a female rower to be faster than a male rower. Rowing has become more competitive in all categories except male heavyweight. We tested whether a junior rower achieves the largest ergometer time improvement by age 16 and accepted this hypothesis except for the female lightweight rowers. Additionally, for each category, we built the 2000-meter ergometer time distribution, which junior rowers can use to assess current performance and understand what they should achieve for top placement. Finally, we developed models predicting future ergometer times based on current performance, sex, weight class, and target percentile rank to guide junior rowers in their journey.

INTRODUCTION

Rowing on land is a sport where an athlete tries to take fast and efficient strokes on a rowing ergometer. A stroke is a series of movements that would propel a boat with oars while an ergometer is a machine that mimics taking strokes on land. Junior rowing includes athletes between the ages of 14 and 18 and is classified into two weight classes: heavyweight (HW) and lightweight (LW). A male junior athlete weighing more than 150 pounds is considered a HW male rower, while a female junior athlete weighing more than 130 pounds is considered a HW female rower.

Over the last two decades, there has been an increase in interest in the rowing sport at the junior level (1, 2). We wondered how performance is affected by age, sex, and weight class. We were also interested in understanding how individual performance compares to the performance of other junior rowers of the same age and in the same category (sex and weight class). Motivated to answer these questions, we used the time to complete a 2000-meter distance on a rowing ergometer as a measure for the performance of an individual rower. We performed statistical analysis on the ergometer times recorded over a 25-year period at the **C**harles **R**iver **A**ll **S**tart **H**as-**B**eens (C.R.A.S.H.-B) Sprints for junior male and female rowers in HW and LW classes. We downloaded the records of these ergometer times from the C.R.A.S.H.-B. Sprints website (3).

There is an increasing number of organizations using statistical analysis to help athletes and sports teams become successful (4). Because this study considered junior rowing, it was closely related to work from Keenan et al. (5). The authors studied sex differences in rowing performance and participation using junior rowing data collected between 1997 and 2016. However, they considered only the rowers ranked in the top 16. Therefore, the authors developed their insights from a dataset of 1,280 rowers and columns of sex, weight class, and finish time. We, on the other hand, analyzed a richer dataset with 12,596 rowers and 10+ characteristics about each rower, presenting new information hidden in the C.R.A.S.H.-B. Sprints data to tell the story of junior rowing. The dataset we used led to the identification of a 2000-meter ergometer time distribution that junior rowers could use to assess current performance and understand what they should achieve for top placement to get ahead in the recruitment process. As a result of analyzing the ergometer times of all C.R.A.S.H.-B. Sprints participants, we also identified the individual rowers that returned to the championship for at least a second year and developed models that predicted the future ergometer times of junior rowers. Furthermore, we discussed the role of age on the junior rower's performance based on sex and weight class.

Today, it is widely recognized by the rowing community that participation of high-school students in rowing has increased over the years (6). Rowing is well known to be intense both mentally and physically (7). Therefore, this sport may be expected to be male dominant (8). Furthermore, junior rowing has become more competitive, driving down the ergometer times as rowers get faster (5).

In this study, we tested whether these opinions on participation and performance are supported by World Indoor Championship data. Specifically, we posed five different testable questions about junior rowing and answered them by performing hypothesis-driven experiments. The first experiment investigated the hypothesis that the annual male rower participation at the C.R.A.S.H.-B Sprints exceeds the female participation. We also hypothesized that sex is the only feature that determines the ergometer time of a rower. The third experiment searched for numerical evidence on the competitiveness of the female LW category while the fourth experiment explored the hypothesis that rowers of each category have exceeded performance of the past. Motivated by prior work on performance development in adolescent track and field athletes (9), the final experiment investigated the hypothesis that a junior rower of each category attains the largest improvement by age 16. Data from these experiments showed that the participation of female HW rowers in the C.R.A.S.H.-B Sprints exceeds the participation of male HW rowers. The data also showed that the key features determining the ergometer time of a rower are primarily sex and then weight class. However, we found numerical evidence that a female rower in the HW class could be faster than a male rower in the LW class. In addition, junior women rowing has been significantly more competitive, including the LW class. Mean ergometer time decreases with rower's age, independent of sex and weight class. However, HW rowers experience their largest ergometer time improvement by age 16 while female LW rowers experience the largest jump in performance from age 17 to age 18. These findings present a new glimpse into the world of junior rowing. Also, for the first time, junior rowers have access to the 2000-meter ergometer-time distribution to help them assess current performance and the models to predict future ergometer times based on current performance, sex, weight class, and target percentile rank.

RESULTS

Descriptive Analysis of Rowing Data

The C.R.A.S.H.-B Sprints data provide junior men and women ergometer times for both HW and LW classes (3). We investigated the demographics of the junior rowers, who raced between 1996 and 2017, by sex and weight class and found that 45% of these rowers competed in the female category **(Table 1)**. We also studied the demographics of the junior rowers who competed between 1997 and 2003 and obtained a decomposition of participation by age, sex, and weight class **(Table 2)**.

Female (n=5,325)		**Male** (n=6,468)	
Heavyweight	**Lightweight**	**Heavyweight**	**Lightweight**
31%	14%	30%	25%

Table 1: Demographics of rowing data by sex and weight class (1996–2017). The dataset includes the rowers who participated in the C.R.A.S.H.-B Sprints between 1996 and 2017.

Age	**Female Heavyweight**	**Female Lightweight**	**Male Heavyweight**	**Male Lightweight**
Age 14	1%	1%	1%	1%
Age 15	5%	2%	4%	3%
Age 16	9%	4%	8%	6%
Age 17	12%	4%	13%	9%
Age 18	4%	2%	8%	5%

Table 2: Demographics of rowing data by age, sex, and weight class (1997–2003). The dataset includes the rowers who participated in the C.R.A.S.H.-B Sprints between 1997 and 2003.

We identified the full distribution of the ergometer times for each category of junior rowing **(Figure 1)**. We also obtained a summary of descriptive statistics, each of which is calculated in the traditional minute:second (mm:ss) format **(Table 3)**. We found that male HW is the fastest category of junior rowing, followed by LW junior men and HW and LW junior women categories.

Female HW Rower Participation Exceeds Male HW Rower Participation

We found a statistically significant difference between the annual participation of female and male rowers who competed at the C.R.A.S.H.-B Sprints between 1995 and 2017 (p-value = 1E-07). We calculated the mean difference between female and male participation to be -49 rowers and the 95% confidence interval (CI) for this mean difference to be (-62 rowers, -35 rowers); thus, between 1995 and 2017, the total male participation exceeded the total female participation. Next, we restricted analysis to the HW class starting from 1998 (i.e., the year in which rowing is recognized as a National Collegiate Athletic Association sport for junior women). We found that there was also a statistically significant difference between the annual participation of female and male rowers in the HW class (p-value = 0.026). However, we found that the mean difference between female HW and male HW participation between 1998 and 2017 was 11 rowers and the 95% CI for this mean difference was (2

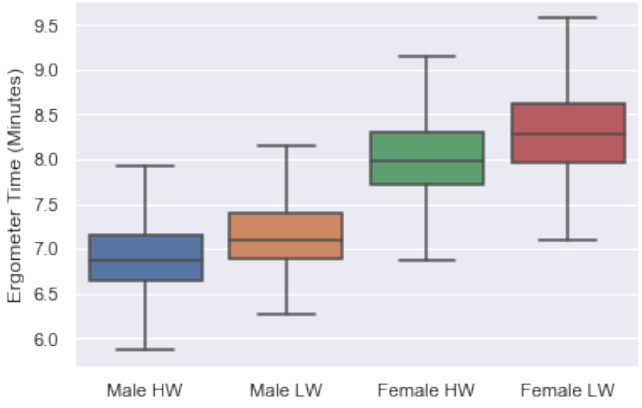

Figure 1: Displaying the distributions of ergometer times (in minutes) using box plots. The middle bar is the median, and the bottom and top of the box are the 25th and 75th percentiles in each of the box plots illustrating that the key feature determining the ergometer time of a rower was first sex and then weight class. Thus, the fastest category was male heavyweight and the slowest category was female lightweight.

Descriptive Statistic	Male Heavyweight	Male Lightweight	Female Heavyweight	Female Lightweight
Mean	06:56	07:11	08:02	08:22
Standard Deviation	00:27	00:27	00:30	00:35
Minimum	05:52	06:16	06:30	07:06
1st Percentile	06:07	06:27	06:59	07:23
2nd Percentile	06:13	06:30	07:08	07:27
5th Percentile	06:21	06:36	07:20	07:35
10th Percentile	06:27	06:42	07:29	07:44
20th Percentile	06:35	06:50	07:40	07:56
25th Percentile	06:38	06:53	07:43	07:58
50th Percentile	06:52	07:07	07:59	08:17
75th Percentile	07:09	07:23	08:18	08:37
90th Percentile	07:27	07:42	08:40	09:04
95th Percentile	07:42	07:58	08:53	09:25
99th Percentile	08:16	08:47	09:29	10:18

Table 3: Descriptive ergometer time statistics presented in the traditional minute:second format. The C.R.A.S.H.-B Sprints data are used for calculating mean, standard deviation, minimum value, and various percentiles of ergometer times for each category of sex, and weight class.

rowers, 21 rowers). Therefore, between 1998 and 2017, the annual female HW rower participation at the C.R.A.S.H.-B Sprints exceeded the annual male HW rower participation. We made the same observation when we illustrated how junior participation has changed over time as a function of sex and weight class **(Figure 2)**. We found an upward trend in the total participation. We also observed that the HW participation is higher than the LW participation and the participation of junior women has been exceeding the participation of junior men in the HW class.

The Key Feature Determining Ergometer Time is Sex, Followed by Weight Class

First, we identified a statistically significant difference between the ergometer times of male and female junior rowers between 1996 and 2017 (p-value < 1E-14). We found that the mean difference was -1.099 minutes and lower and upper bounds of the 95% CI for this mean difference were -1.118 minutes and -1.081 minutes. Thus, male rowers are faster than female rowers. We also found a statistically significant difference between the ergometer times of HW and LW classes (p-value = 4E-14). We calculated the mean difference and the 95% CI to be -0.107 minutes and (-0.134 minutes, -0.079 minutes), respectively: the HW rowers are faster than LW rowers. Therefore, sex is not the only feature that determines the ergometer time of a rower. Instead, the key feature that determines the ergometer time of a rower is first sex and then weight class. We found support for this result with box plots of the ergometer times **(Figure 1)** whose descriptive statistics are in **Table 3**. These results show that the mean ergometer time for a male LW rower is 4% slower than for a male HW rower. However, male HW is a very competitive category as 76.5% of all the participants in this category are faster than an average male LW rower.

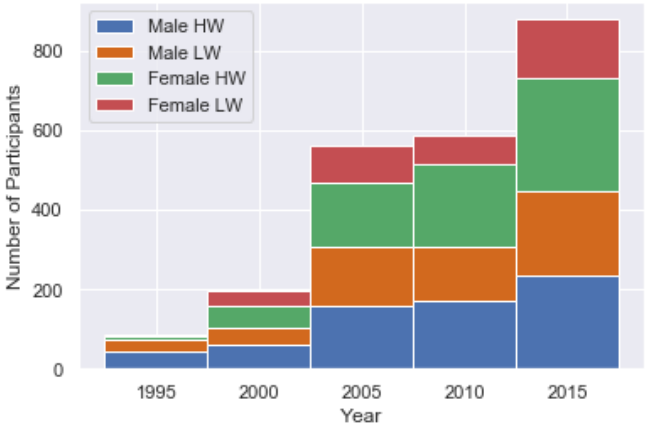

Figure 2: Participation by category. The total number of junior rowers participating in the C.R.A.S.H.-B Sprints World Indoor Rowing Championship increased from 1995 to 2017.

We also observed that the fastest female HW rowers can be as competitive as a male LW rower. Although an average female HW rower is 12% slower than an average male LW rower, 2.5% of all the female HW rowers are faster than an average male LW rower. Thus, it is possible for a female rower to be faster than a male rower, but this observation holds only for the fastest of the female rowers in the HW class. We tested this observation by comparing the male LW ergometer times to the ergometer times of the top five female HW rowers. We found that the difference between the ergometer times of male LW rowers and the top five female HW rowers was statistically significant (p-value = 0.014). Furthermore, we calculated the mean ergometer time difference between these two groups to be 4.902 seconds and the 95% CI for the mean ergometer time difference between these two groups to be (1.099 seconds, 8.704 seconds). Thus, HW junior women with top placements row faster than some LW junior men.

Finally, we observed that the mean ergometer time of a female LW rower was 4% slower than the mean ergometer time of a female HW rower. Only 30% of the female LW rowers were faster than an average female HW rower. However, being a competitive female LW rower requires an ergometer time that falls below 8 minutes. Each row of Table 3 associated with less than or equal to the 20th percentile presents an ergometer time that is faster than 08:00 mm:ss.

Junior Women Rowing Has Become Significantly More Competitive

First, we obtained the annual ergometer times of the top five rowers for each category. Then, we fit a linear regression line to the ergometer times of each category by choosing year as the independent variable and ergometer time as the dependent variable. For each category, we tested the null hypothesis that there is no statistically significant relationship between year and ergometer time and calculated the p-value, the mean (in minutes), and the 95% CI (in minutes) **(Table 4)**. Except for the male HW class, we determined that the relationship between time and ergometer time was statistically significant. We also found the values in the 95% CIs to be negative. This means that the ergometer times of the competitive rowers have decreased over the years, except for the male HW class. We illustrated this finding by plotting the ergometer times of junior rowers that placed in the top five in each year by category **(Figure 3)**. The linear

Parameter	Male HW	Male LW	Female HW	Female LW
p-value	0.644	3.78E-11	1.06E-04	2.36E-08
mean (minutes)	-0.001	-0.010	-0.012	-0.014
95% CI (minutes)	-0.004 – 0.002	-0.013 – -0.007	-0.017 – -0.006	-0.019 – -0.009

Table 4: Testing significance of relationship between time and performance. The *p*-value is calculated as less than 0.05 except for the male heavyweight category; therefore, the relationship between time and ergometer time is found to be statistically significant for each of the male lightweight, female heavyweight, and female lightweight categories. The means and the 95% CIs are given in minutes.

regression lines of **Figure 3** support that junior rowing has become more competitive over the years, except for the male HW category. As also reflected in the 95% CIs in **Table 4**, the largest increase in competitiveness has been in the female LW category.

Heavyweight Rowers Experience Largest Ergometer Time Improvement by Age 16

We used the ergometer times recorded between 1997 and 2004 and tested the null hypothesis that there is no statistically significant difference in the ergometer times recorded at two consecutive ages, starting at age 14 and ending at age 18, for each category. We calculated the *p*-values, means, and 95% CIs **(Table 5)**. We highlighted the entries of **Table 5** that indicate the age at which the performance improvement was statistically significant. We also investigated the impact of age using box plots with age on the x-axis and performance on the y-axis **(Figure 4)**. The arrows in **Figure 4** indicate the age at which the performance improvement was identified as statistically significant for the first time in **Table 5**. We observed that the largest improvement in mean ergometer time occurred from age 14 to age 15 in the male HW class. The largest improvement in the mean ergometer time happened with a year delay both in the male LW category and in the female HW category. For female LW rowers, the largest improvement took place by age 18.

Figure 3: Ergometer times of the top five rowers (in minutes). There was a downward trend in the linear regression fits obtained for ergometer times of the fastest rowers in each category except for the male heavyweight category.

Age	Parameter	Male HW	Male LW	Female HW	Female LW
14 – 15	*p*-value	**0.006**	0.569	0.074	0.420
	mean (minutes)	0.493	0.073	0.289	0.152
	95% CI (minutes)	0.215–0.770	-0.151–0.297	0.040–0.538	-0.277–0.581
15 – 16	*p*-value	**0.024**	**0.001**	**0.005**	0.087
	mean (minutes)	0.138	0.189	0.178	0.261
	95% CI (minutes)	0.020–0.257	0.075–0.302	0.056–0.301	-0.016–0.537
16 – 17	*p*-value	**1.07E-06**	**0.008**	**0.001**	0.529
	mean (minutes)	0.206	0.114	0.157	0.086
	95% CI (minutes)	0.127–0.284	0.028–0.200	0.126–0.188	-0.187–0.359
17 – 18	*p*-value	0.431	0.073	0.145	**0.009**
	mean (minutes)	0.028	0.084	0.083	0.368
	95% CI (minutes)	-0.044–0.100	-0.012–0.180	-0.034–0.201	-0.013–0.749

Table 5: Testing significance of performance improvement with age using *p*-values, means, and 95% CIs. The performance improvement is statistically significant when the *p*-value is less than 0.05. Therefore, the performance improvement is statistically significant till age 17 for the male heavyweight category; the performance improvement is statistically significant from age 15 to age 17 both for male lightweight and female heavyweight categories; and performance improvement is statistically significant only from age 17 to age 18 for the female lightweight category. Any age at which the performance improvement is statistically significant is in bold. Both the mean and the 95% CIs are presented in minutes for each age group and each category.

Performance Prediction

The World Indoor Championship data can be further used to develop a model that predicts the ergometer time of a rower one year later given current ergometer time, target percentile rank, sex, and weight class. We show how this can be done for a female LW rower with an ergometer time of 08:15 mm:ss − seven seconds faster than an average female LW rower – and a target percentile rank of 75%. We used the ergometer times of the female LW C.R.A.S.H.-B Sprints participants that had returned to race for at least one more time between 1996 and 2019. We identified a high positive correlation − a Pearson's correlation coefficient of 0.91 − between the ergometer times of these female LW rowers that had participated at the C.R.A.S.H.-B Sprints for at least two consecutive years.

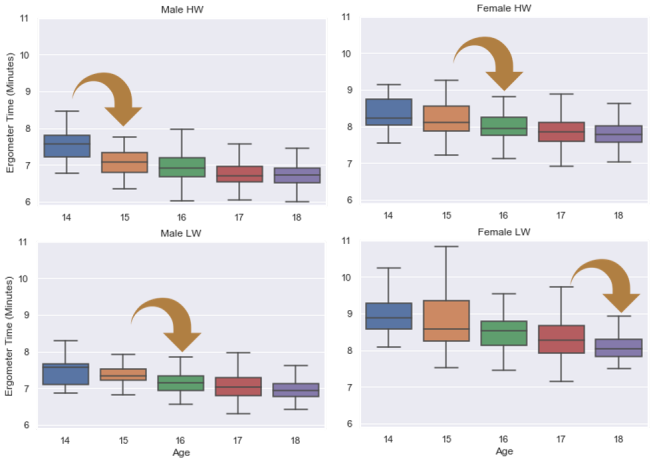

Figure 4: Ergometer times (in minutes) by category and age. The largest improvement in ergometer time was observed from age 14 to age 15 in the male heavyweight category, from age 15 to age 16 in the male lightweight and female heavyweight categories, and from age 17 to age 18 in the female lightweight category.

We treated the next year's ergometer time as a continuous target variable and identified linear regression as the best approach to model the relationship between the ergometer times of consecutive years and the target percentile rank for performance prediction. Although a caveat of linear regression is that it is limited to represent linear relationships, it is a suitable choice for modeling the ergometer times of the female LW rowers because this data follows a linear trend. By comparing the predictions from the linear regression fit to the observations in the validation and test datasets, we calculated the mean-squared errors to be 0.014 and 0.012 and the accuracies – the percentages of correct predictions – to be 92% and 94%. Despite the successful performance of the linear regression fit for the female LW rowers' ergometer times, a linear model is known to be applicable for a given range of data and to be sensitive to outliers. This issue did not appear to be a concern here because the range of the female LW ergometer times corresponds to an approximately three-minute duration **(Table 3)**. Instead, a linear model readily quantifies the relation between the prediction and an explanatory variable. The resulting model predicts the next year's ergometer time to be 07:57 mm:ss for the 08:15 mm:ss current ergometer time and the 75th target percentile rank. Thus, reducing the ergometer time to below 08:00 mm:ss within the next year is an appropriate measure to target for this female LW rower to remain competitive (in the top 25% of the female LW category). If this rower finishes the 2,000 meters in 07:57 mm:ss at the end of the next year and aims for a 95% target percentile rank for the following year, the model predicts that year's ergometer time to be 07:37 mm:ss. Thus, the prediction 07:57 mm:ss is for Year 1 and the prediction 07:37 mm:ss is for Year 2 **(Table 6)**. We also investigated how

Year	Target Percentile Rank	Male HW (Current Time) Prediction	Male LW (Current Time) Prediction	Female HW (Current Time) Prediction	Female LW (Current Time) Prediction
1	50%	(06:52) 06:53	(07:07) 07:08	(07:59) 08:03	(08:15) 08:13
1	55%	(06:52) 06:52	(07:07) 07:04	(07:59) 07:57	(08:15) 08:10
1	60%	(06:52) 06:50	(07:07) 07:07	(07:59) 07:53	(08:15) 08:07
1	65%	(06:52) 06:43	(07:07) 07:00	(07:59) 07:51	(08:15) 08:03
1	70%	(06:52) 06:43	(07:07) 07:00	(07:59) 07:47	(08:15) 08:00
1	75%	(06:52) 06:38	(07:07) 06:55	(07:59) 07:46	(08:15) **07:57**
2	80%	(06:38) 06:35	(06:55) 06:49	(07:46) 07:39	(07:57) 07:47
2	85%	(06:38) 06:33	(06:55) 06:45	(07:46) 07:34	(07:57) 07:44
2	90%	(06:38) 06:25	(06:55) 06:41	(07:46) 07:28	(07:57) 07:40
2	95%	(06:38) 06:23	(06:55) 06:38	(07:46) 07:21	(07:57) **07:37**

Table 6: Two-year ergometer time predictions for given combinations of current ergometer time, target percentile rank, and junior rowing category. The second year starts with the rower achieving the first-year prediction obtained for the 75% target percentile rank. The model predicts the year-1 female lightweight ergometer time as 07:57 mm:ss for the 08:15 mm:ss current ergometer time and the 75% target percentile rank. If this rower finishes the 2,000 meters in 07:57 mm:ss at the end of the first year and aims for the 95% target percentile rank for the second year, the model predicts year-2 ergometer time as 07:37.

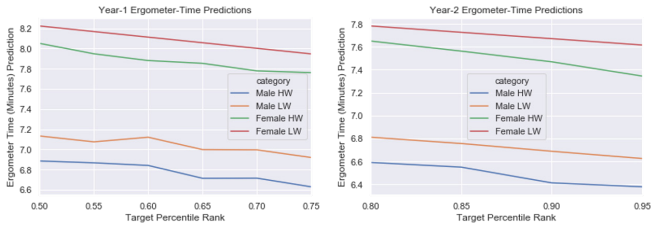

Figure 5: Ergometer time predictions (in minutes) for given combinations of current ergometer time, target percentile rank, and junior rowing category. The female LW category predictions were obtained from a linear regression fit while the remaining predictions were obtained from random forest fits. There was a downward trend in the ergometer time predictions with the increasing values of the target percentile rank for each rowing category.

the predictions would change with different values of target percentile rank **(Table 6)**. Furthermore, we obtained these findings not only for the female LW category, but also for the other three categories of junior rowing: female HW, male HW, and male LW.

Although the female LW category predictions were obtained by using a linear regression model, the predictions for the female HW, male HW, and male LW categories were obtained by using random forest models. A random forest is an ensemble of decision trees, which are tree-shaped representations of all possible outcomes for given decisions. The strength of a random forest model comes from a combination of different decision-tree models performing better than the individual decision-tree models. For example, it is possible for an individual model to overfit to a specific portion of the data. However, combining different models into an ensemble is expected to average out the poor performances of the individual models, reduce overfitting, and improve the overall model performance. It is also this ensemble feature of the random forest model that helps it outperform linear regression models in capturing nonlinear relationships. Nonlinearity appears in the ergometer times of female HW, male HW, and male LW categories because the ergometer times recorded for faster junior rowers are more strongly dependent on each other than those recorded for the slower junior rowers. Although a caveat of a random forest model is the need for many decision trees, for this study we built random forest models as ensembles of ten different decision trees. By comparing the predictions from the random forest fits to the observations in the validation and test datasets, we calculated the accuracies to be 97% and 96% for the female HW category, 94% and 93% for the male HW category, and 94% and 95% for the male LW category. In **Table 6**, the median value in **Table 3** was chosen to be the first-year ergometer time (current time) for each of these three categories. Finally, we plotted both the first-year ergometer time predictions and the second-year ergometer time predictions **(Figure 5)**.

DISCUSSION

Performance

We showed that the female LW category has been increasingly more competitive. We also showed that the rowers in this category were expected to achieve their largest ergometer time improvement by age 18, while rowers in the other categories accomplished the same achievement by age 16. Thus, the female LW rowers should not be discouraged by the smaller early improvement in their ergometer times. If a rower is in a category that may not be as strong as another one, then it may take longer for that rower to achieve their greatest improvement. The most important thing for rowers in such a situation is not to give up; instead, they should persevere and believe that hard work will eventually pay off.

Performance Prediction

In this study, we built prediction models by only using the data from the C.R.A.S.H.-B Sprints. However, a significant number of rowing competitions take place, especially in the winter months. Combining the World Indoor Championship data with the records of other indoor competitions could be a goal of future work to improve prediction accuracy with the collection of additional data.

MATERIALS AND METHODS

Rowing Data Description

The rowing dataset that we downloaded from the C.R.A.S.H.-B. Sprints website (3) contains 12,596 rows and 12 columns. Each row represents a specific contest of a rower. Each column represents a specific feature of the rower and the race including year, sex, weight class, ranking (representing the placement with respect to the ergometer time in a specific race), rower (name), organization, college, age, state, country, ergometer time, and category. The dataset includes data from 1995 to 2019 while the ages of the junior rowers are provided between 1997 and 2003. The C.R.A.S.H.-B Sprints held the title of "World Indoor Rowing Championship" until 2018 (9). The lack of the title in 2018 and 2019 led to a significant reduction in the C.R.A.S.H.-B. Sprints participation. It is for this reason that we studied participation and performance by using the data recorded until 2017. However, we included the data from 2018 and 2019 for predictive modeling because predictive model development requires the tracking of the ergometer times for individual rowers returning to the C.R.A.S.H.-B Sprints.

Web Scraping and Data Cleaning

We extracted the results of 100 different races (4 races per year over the course of 25 years) from the C.R.A.S.H.-B Sprints website (3). The format of the data was not the same for each year. Therefore, we extracted the data by using a combination of urllib3 (version 1.24.2), BeautifulSoup (version 4), requests (version 2.22.0), and lxml.html (version 3.7.2) Python packages. We also used Python 3.7.4 to perform descriptive and predictive rowing analytics for this study. We cleaned each of the extracted data sets by filling in the missing information and correcting the category misclassifications and the misspellings, which were mostly associated with the rower, organization, state, and country columns, by relying on the information at the RegattaCentral (10).

Selecting Participation and Performance as the Output Variables

We analyzed the cleaned C.R.A.S.H.-B. Sprints data for the following two variables: participation and performance. We defined participation as the number of rowers enrolled annually in each of the male HW, male LW, female HW, and female LW categories. We measured performance with the ergometer time. To study participation, we restricted our focus to the data available from 1995 to 2017, when C.R.A.S.H.-B Sprints had the World Indoor Rowing Championship title. To study performance, we switched focus to the data starting in 1996 because it was the first year when the ergometer times were collected for the 2,000-meter distance.

Hypothesis Testing

We tested each hypothesis of this study at a significance level of 0.05. We used the method of paired two-sample t-test for the means to test the hypothesis that the annual male rower participation at the C.R.A.S.H.-B Sprints exceeds the female participation. We used the method of two-sample t-test assuming unequal variances to test the hypothesis that sex is the only feature that determines the ergometer time of a rower and to test the hypothesis that a junior rower of each category attains the largest improvement by age 16. We tested the remaining hypotheses by using the method of testing the slope of a linear regression line. For each test, we further provided the accompanying 95% CIs. We obtained all the results for hypothesis testing by using Data Analysis Tools of Microsoft Excel Version 2008.

Performance Prediction

Using Python's machine learning library (scikit-learn version 0.21.3), we first split the data into training, validation, and testing datasets. We used a 60% split for training and a 20% split each for validation and testing. Then, we identified the best model as the linear regression model for the female LW category and the random forest model for each of the female HW, male HW, and male LW categories. Finally, we used the LinearRegression and RandomForestRegressor implementations available in Python's machine learning library for fitting linear regression and random forest models to the data. We computed training, validation, and test accuracies to be 0.92, 0.92 and 0.94 for the female LW category, 0.98, 0.97 and 0.96 for the female HW category, 0.99, 0.94 and 0.95 for the male LW category, and 0.99, 0.94 and 0.93 for the male HW category.

Received: September 20, 2020
Accepted: October 18, 2021
Published: January 7, 2022

REFERENCES

1. SFA-SFM (2016). The Resurgence of Rowing. Sports Facilities Advisory and Sports Facilities Management (SFA-SFM). July 6th, 2016.
2. NCAA (2020). NCAA Sports Sponsorship and Participation Rates Report (1956-57 through 2019-20). Posted: September 1, 2020. Indianapolis, IN.
3. C.R.A.S.H.-B World Indoor Rowing Contest Website. http://www.crash-b.org. Accessed 1 Sep. 2019.
4. SAS® Analytics. "Ahead of the Game: Learn how Professional Sports Teams Win Fans and Games with Analytics." 2017, https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/ahead-of-the-game-108813.pdf. Accessed December 29th, 2019.
5. Keenan, Kevin G., et al. "Girls in the Boat: Sex Differences in Rowing Performance and Participation." *PLoS One*, vol. 13, 19 January 2018, pp. 1-14.
6. Canaley, Jess. "As Popularity of Rowing Increases, Athletes Look to Continue in College Despite Limited Scholarships." 2018, https://hilite.org/60327/sports/as-popularity-of-rowing-increases-athletes-look-to-continue-in-college-despite-limited-scholarships/. Accessed January 5th, 2021.
7. Brown, Daniel James. *The Boys in the Boat: Nine Americans and Their Epic Quest for Gold at the 1936 Berlin Olympics*. The Penguin Books, 2013.
8. Deaner, Robert O., et al. "A Sex Difference in the Predisposition for Physical Competition: Males Play Sports Much More Than Females Even in the Contemporary U.S." *PLoS One*, vol. 7, 14 November 2012, e49168.
9. Tonnessen, Espen, et al. "Performance Development in Adolescent Track and Field Athletes According to Age, Sex and Sport Discipline." *PLoS One*, vol. 10, 4 June 2015, e0129014.
10. RegattaCentral. https://www.regattacentral.com/clubs/. Accessed 1 Sep. 2019.