ABSTRACT

Introduction: The 100-meter dash (100 m) event holds particular appeal. Coaches and researchers seek to understand the determinants of performance in this task. Although information has been produced over the years, it is not fully applied by coaches who generally assess the success of employed training methods through objective field tests, such as 60 m dash test performance. Objective: Investigate 100 m performance based on 60 m performance. Methods: Two hundred and forty six men and 153 women divided into two subgroups were evaluated for estimation (F_{validation}; n=123 and M_{validation}; n=204) and validation of predictive models (F_{cross-validation}; n=30 and M_{cross-validation}; n=42) for 100 m dash performance (time take to cover 100 m). Partial time was measured based on the 100 m distance marked previously every 10 meters from the starting line on both sides of the track. The predictive models were based on the interval in the 60 meters with a time interval of 10-10 m. Results: Magnitude of correlation was very high. High coefficients of determination and differences of no statistical significance (p <.001) were found between the criteria and predicted values. The predictive equations presented constant error values below 0.001 s; total absolute error of 0.12 s; 0.10 s for M_{validation} and F_{validation}, respectively, and 1.13% and 0.85% of total relative error for M_{validation} and F_{validation}, respectively. Bland-Altman analysis showed an increase in the level of concordance between the criteria and predicted values of F_{validation} and M_{validation}. Similar responses were found when the proposed models were applied to F_{cross-validation} and M_{cross-validation}. Conclusion: The estimation models were able to accurately predict 100 m performance based on 60 m performance. Level of evidence: II; Diagnostic studies - Investigating a diagnostic test.

Keywords: Running; Athletic performance; Biomechanical phenomena.
previously signaled to every 10 meters, to start the line of the track in both sides of the track. The models were estimated in the interval in the 60 meters to the start of the interval of time of 10-10m. Results: The magnitude of the correlation estimated was very high. The models used to predict the values were significant at p<0.001. The equations that predicted the values were shown to be significant at 0.01; error total absolute of 0.12s; 0.10s for M_validación and F_validación respectively, and 1.13% and 0.85% of error relative total for M_validación and F_validación respectively. The analysis of Bland-Altman showed an increase in the level of concordance between the criteria and the values predicted for F_validación and M_validación. The same criterion at the round and the criterion to predict the values of errors in studies with similar criteria. The models were able to predict performance with high precision in 100m. The objective of the present study and its applicability thereof, and when the criterion is to be predicted to some subsequent behavior analysis, the predictive validity becomes the point of greatest concern.12 With this in mind, each group was randomly divide into mutually exclusive subsets, used to estimate the parameters for M_validación [n=123 and M_validación with n=153)] and to cross validate the models for F_validación [F_cross-validation = 30 and M_cross-validation = 42]). The cross-validation subsets exhibited a representative sample size of 20% of the sample size for sex. Kinematic analysis was performed by the technique frame by frame. The high acquisition rate cameras (EXILIM® - Câsio) were used at a sampling rate of 240 frames per second (fps). Although 10014 and 50 fps15 are reported in literature for similar analysis with acceptable error magnitude,24 240 fps was used in order to obtain greater discretization and minimizing errors in the identification of desired events. Partial distances of 10 meters were previously flagged from the starting line on both sides of the running track, allowing the quantification of the time interval in which each athlete concluded the distance. Cameras were fixed in tripods and positioned 12 meters in height relative to the ground, with a distance of 35 meters starting from the outer edge of the lap number eight with 20 meters between each other. The calibration of the collection of each partial volume was performed with 10-cm diameter markers set at known distances. Despite being compulsory in official sprints to use blocks starting strategy in order to assign ecological validity and therefore, closer proximity to the usually adopted in monitoring and control procedures of the training loads, the moment when the foot touched the ground after the beep-on test was considered to start the timing. The time interval at which each participant completed each partial distance of 10 meters to 60 meters long was used for generation of predictive models. To validate the predictive models the sum of all the time slots of each partial measurement was used as the criterion. The temporal parameters were quantified with Kinovea Video Editor 0.8.15 software. After confirmation of adherence to normal distribution of data, descriptive statistics used were the average and standard deviation. Considering the major changes of the magnitude of time of each partial 10m for the development of a predictive model, multivariate...
regression analysis was used. The adequacy of the data to the assumptions was previously checked.\textsuperscript{16,17}

The normal distribution of the residuals was performed by Kolmogorov-Smirnov test, considering the standardized residuals; homoscedasticity was checked by the square of the standard waste of Pesarán-Pesará and Bartlett sphericity.

The absence of autocorrelation was analyzed by Durbin-Watson test.\textsuperscript{17,18} The linearity of the residues was assessed by visual inspection of the scatter plot constructed with standardized predicted values and standardized residuals.\textsuperscript{17} For diagnosis of multicollinearity the quantification criteria of the parameters Variance Inflation Factor (VIF), eigenvalue, incremental percent, condition index was taken. For VIF a value of up to 1 was assumed as absence of collinearity between 1 and 10 with acceptable collinearity > 10 with collinearity problem.\textsuperscript{19} For eigenvalue and incremental percent, values close to zero, and for condition index, values close to 10 are all indicative of collinearity.\textsuperscript{16}

There was still necessary to use the multivariate technique of main components analysis and Ridge Regression Estimator regression analysis.\textsuperscript{20}

The Pearson (r) and determination (R\textsuperscript{2})\textsuperscript{12,17} correlation coefficient, constant error (CE), standard error of measurement (SEM), absolute total error (TE\textsubscript{abs}), and overall relative error (TE\textsubscript{rel})\textsuperscript{24} were calculated.

The correlation coefficients were ranked as trivial (r<0.1), low (0.1<r<0.29), moderate (0.31<r<.50), large (0.51<r<.70), very large (0.71<r<.90), near perfect (0.91<r<.99) and perfect (r=1).\textsuperscript{17}

Student’s t-test for paired samples (t) was used to check for differences between the predicted and criteria values, and Bland-Altman\textsuperscript{23} was conducted to verify the agreement between the predicted and criteria values, including the limits of agreement at 0.95%.

Whereas the estimated regression coefficients can vary from one sample to another when the independent variables are correlated (multicollinearity) with inaccuracy of the predicted value,\textsuperscript{24} we investigated the possibility of extrapolating models proposed to another context (cross-validation).\textsuperscript{12,19}

The values of r, R\textsuperscript{2}, t, CE, SEM, TE\textsubscript{abs}, TE\textsubscript{rel} and Bland-Altman analysis\textsuperscript{23} were also carried out for cross-validation analysis.

The NCSS\textsuperscript{®} (NCSS LLC, USA) and SPSS\textsuperscript{®} (SPSS Inc, USA) softwares were used. The significance level equivalent to a Type I error was 0.5%.

RESULTS

Analysis of the standardized residuals showed values of p=0.94 and p=0.58 for F and M, respectively, given the normal distribution assumption. This response is similar to the square analysis of standardized residuals (p=0.31 and p=0.97 for F and M, respectively) and sphericity test (p<0.001).

Durbin-Watson test\textsuperscript{18} indicated the absence of autocorrelation from the chosen confidence level (p<0.05), sample size (by sex) and number of independent variables. Visual inspection of the dispersion graph between standardized predicted value and standardized residual value showed randomly dispersed distribution around zero, then linearity and homoscedasticity.\textsuperscript{17}

Given the eigenvalue, incremental percent, condition index parameters, different results were indicative of multicollinearity. (Table 1)

With this, the multivariate technique of main components analysis was used to observe the formation of components and their contributions of parameters for each component, with the intention of listing variable passive exclusion; however, the analysis showed the formation of just one factor, not allowing the exclusion of parameters.

With the identification of multicollinearity and impossibility of reducing independent variables, we adopted the Ridge Regression Estimator analysis given the ability to produce more accurate observations in this condition.\textsuperscript{24} Thus, multicollinearity was controlled (VIF near zero (p<0.001) with estimator 0.92 M to 0.36 F). With all assumptions met, prediction models were generated.

Female model = 1.383695\textsuperscript{+}1.047914\textsuperscript{*}1.531326\textsuperscript{*}X1+X2+X3\textsuperscript{+}1.582101\textsuperscript{*}1.394641\textsuperscript{*}1.458194\textsuperscript{*}X4+X5+X6\textsuperscript{+}1.564195

Male model = 2.585927\textsuperscript{+}0.538856\textsuperscript{+}9.107026\textsuperscript{*}X1\textsuperscript{+}X2\textsuperscript{+}X3\textsuperscript{+}0.9878149\textsuperscript{*}1.690163\textsuperscript{*}1.69796\textsuperscript{*}X4+X5+X6\textsuperscript{+}1.605068

Where: X1 = time for 0-10 meters, X2 = time for 10-20 meters, X3 = time for 20-30 meters, X4 = time for 30-40 meters, X5 = time for 40-50 meters, X6 = time for 50-60 meters.

Whereas M\textsubscript{validation}, and F\textsubscript{validation}, the predictive models presented EC below 0.001s (both sexes), TE\textsubscript{abs} of 0.12s and 0.10s respectively, and TE\textsubscript{rel} of 1.13% and 0.85% respectively. (Table 2)

In these subgroups were also found near perfect correlation coefficients and determination coefficient high values, besides not having statistically significant differences (p<0.001) between predicted and criteria values. (Table 2)

Bland-Altman analysis showed high level of agreement between the predicted and criteria values in the subgroups F\textsubscript{validation} and M\textsubscript{validation} (Figure 1A and 1C) and F\textsubscript{cross-validation} and M\textsubscript{cross-validation} (Figure 1B and 1D).

Table 1. Ridge regression analysis results.

<table>
<thead>
<tr>
<th>Partial (m)</th>
<th>VIF</th>
<th>Eigenvalues of correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Eigenvalue</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>F</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 10</td>
<td>1.4</td>
<td>2.6</td>
</tr>
<tr>
<td>10 to 20</td>
<td>2.2</td>
<td>3.5</td>
</tr>
<tr>
<td>20 to 30</td>
<td>2.5</td>
<td>4.5</td>
</tr>
<tr>
<td>30 to 40</td>
<td>4.4</td>
<td>6.2</td>
</tr>
<tr>
<td>40 to 50</td>
<td>5.7</td>
<td>9.9</td>
</tr>
<tr>
<td>50 to 60</td>
<td>5.0</td>
<td>9.1</td>
</tr>
</tbody>
</table>

Male (M), Female (F), variance inflation factor (VIF).

Table 2. Performance characteristics (mean ± SD) of the validation (M\textsubscript{validation} and F\textsubscript{validation}) and cross-validation groups (M\textsubscript{cross-validation} and F\textsubscript{cross-validation}).

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>CE</th>
<th>TE\textsubscript{abs}</th>
<th>TE\textsubscript{rel}</th>
<th>SEM</th>
<th>r</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>M\textsubscript{validation}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.95 (±4.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F\textsubscript{validation}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.95 (±5.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M\textsubscript{cross-validation}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10.87 (±3.8)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>F\textsubscript{cross-validation}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.87 (±5.9)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*p<0.05 Pearson linear correlation coefficient (r) between the measured criteria values and predicted ones, determination coefficient (T), constant error (CE), standard error of measurement (SEM), absolute total error (TE\textsubscript{abs}), total relative error (TE\textsubscript{rel}).

DISCUSSION

The 60m test has been commonly used by coaches and trainers for performance prediction of 100m activities, however empirically explaining the embodiment of the present study.

Performance prediction models are promising given that predicted values (10.95±0.38s and 11.84±0.55s) are close to the criteria values (10.95±0.45s and 11.84±0.59s) for M\textsubscript{validation} and F\textsubscript{validation}, respectively, and for presenting extremely low SEM values (0.02s for both sexes), CE (<0.001s for both sexes), TE\textsubscript{abs} of 0.12s and 0.10s (for M\textsubscript{validation} and F\textsubscript{validation}, respectively) and TE\textsubscript{rel} 1.13% and 0.85% (for M\textsubscript{validation} and F\textsubscript{validation}, respectively). (Table 2) These results indicate the accuracy of the model generated in the prediction.

The SEM value of 0.02s suggests a range within which the “true value” can be expected considering the error associated.\textsuperscript{26} Considering
a hypothetical condition of an athlete completing the distance in 11.20s, one can say with 0.95% confidence (2±SEM) the "true value" for this individual is between 11.24s and 11.16s. Thus, major changes to 0.02 seconds represent real change in performance rather than measurement error.

This approach contributes greatly to coaches who use the tests as performance indicators, since it can be considered a "real" condition of the athlete.

CE lower than 0.001s, TE_rel of 1.13% and 0.85% (for M_validation and F_validation, respectively), correlation coefficients near perfect (R=.992 and 0.98% determination, besides absence of differences (p<.001) between the predicted and criteria values indicate that the models present accuracy and validity.

Bland-Altman analysis (Figure 1) confirms this condition given to substantially close agreement observed between criteria and predicted time.

With this, it can be considered that the proposed models are valid for predicting performance in the 100 meters, for both men and women.

Considering the possibility of extrapolating the use of the generated models, knowledge of the responses found in other population was necessary. In this regard, for M_validation and F_validation subgroups, the models application generated predicted values of 10.89±0.31s and 11.85±0.31s, respectively, consistent with criteria values (10.87±0.38s and 11.84±0.33s, respectively).

Following the results found for M_validation and F_validation groups, SEM values were extremely low (0.02 and 0.03s seconds for M_validation and F_validation, respectively), as well as CE (-0.02s and -0.03s for M_validation and F_validation, respectively), TE_abs of 0.10s to 0.80s (for M_validation and F_validation, respectively) and TE_rel of 0.88% and 0.64% (for M_validation and F_validation, respectively). (Table 2)

The correlation and determination coefficients also showed values close to perfect (0.98<r<.99, 96% <R<98%) in both subgroups. (Table 2)

These results demonstrate that when generated models were applied in a separate sample, results with low error magnitude were found with high levels of agreement between the predicted and actual values. (Figure 1B and 1D)

Bland-Altman analysis and M_validation and F_validation subgroups (Figure 1A and 1C) also revealed a regression models tendency for producing larger SEM values in lower performance athletes. These results are in agreement with the literature, since when distinct 100m paths of national level athletes and runners of world championships were compared, a significant reduction of speed (p<.05) after 60 meters was found in the lower performance athletes. Thus, higher performance athletes tend to produce a lower change rate in the final stretches of proof than lower performance level ones.

In this scenario, the results suggest models with accepted external validity, given the possibility of applying these models in other samples from runners with accurate response.

Nevertheless, despite having found relevant answer of practical nature, possible limitations on the construction of the proposed models can be found on a proposed relationship basis between the number of predictor variables and the number of observed cases, suggested from 10 cases of data for each predictor model.

Although a very close recommended relationship has been used in this study (12 cases for each model predictor in F_validation and 20 cases for each model predictor in M_validation), this does not compromise the generated model, since it is also recommended that the sample size depends on the effect size that you want to detect. Whereas it has been proposed to calculate the effect of sample size on research by the equation k / (n-1), being K the number of predictor variables, n the number of observed cases, in this study, there was obtained an effect
size of 0.05 and 0.03 for $F_{\text{val}}$ and $M_{\text{val}}$, respectively. Therefore, it was classified as very small or insubstantial magnitude of effect, signaling that the sample size did not influence the modeling equations, and so, the prediction of the desired performance.

**CONCLUSION**

Results suggest models with external validity, given the possibility of application to other sprinter samples with accurate response when purpose is to predict tests of 100m from parameters obtained in the distance of 60m. This extrapolation possibility is justified by reduced errors in predicted values with no significant differences between the criteria and predicted value, high levels of agreement, Pearson product-moment correlation coefficient and high coefficients of determination values.

All authors declare no potential conflict of interest related to this article.

**AUTHORS’ CONTRIBUTIONS:** Each author made significant individual contributions to this manuscript. RMA (0000-0002-4816-5635)**: substantial contribution to the conception or design of the work, acquisition, analysis or interpretation of data for the work; drafting of the work; final approval of the version of the manuscript to be published. AJF Jr (0000-0002-6635-8019)**: substantial contribution to the conception or design of the work, critical revision of its intellectual content; final approval of the version of the manuscript to be published. JC (0000-0002-3646-3387)**: substantial contribution to the conception or design of the work; interpretation of data for the work; critical revision of its intellectual content; final approval of the version of the manuscript to be published. *ORCID (Open Researcher and Contributor ID).

**REFERENCES**

2. Mackala K. Optimisation of performance through kinematic analysis of the different phases of the 100m. NSA. 2007;22(2):7-16.