Construction of an anthropometric discriminant model for identification of elite swimmers: An adaptive lasso approach

Qile Pan [1], **Rongxin Zhu** [1], **Jun Qiu** [1], **Guang Cai** Corresp. [1]

[1] Shanghai Research Institute of sports science, Shanghai, China

Corresponding Author: Guang Cai
Email address: leng8yang@126.com

Background. Anthropometric characteristics are important factors that affect swimming performance. The aim of this study is to build a discriminant model using anthropometric factors to identify elite short-to-medium-distance freestyle swimmers through an adaptive Lasso approach.

Methods. The study recruited 254 swimmers (145 males and 109 females) who were divided them into elite (aged 17.9 ± 2.2 years, FINA points 793.8 ± 73.8) and non-elite (aged 17.1 ± 1.3 years, FINA points 560.6 ± 78.7) groups. Data for 73 variables were obtained, including basic information, anthropometric and derivative indicators. After filtering out highly correlated variables, 24 candidate variables were retained to be used in adaptive Lasso to select variables for prediction of elite swimmers. Deviance and area under the curve (AUC) were applied to assess the goodness of fit and prediction accuracy of the model, respectively.

Results. The adaptive Lasso selected 12 variables using the whole sample, with an AUC being 0.926 (95% CI: 0.895-0.956; P=2.42×10^{-29}). In stratified analysis by gender, nine variables were selected for male swimmers with an AUC of 0.921 (95% CI: 0.880-0.963; P=8.82×10^{-17}), and eight variables were for female swimmers with an AUC of 0.941 (95% CI: 0.898-0.984; P=7.67×10^{-15}).

Conclusion. The adaptive Lasso showed satisfactory performance in selecting anthropometric characteristics to identify elite swimmers. Additional studies with longitudinal data or data from other ethnicities are needed to validate our findings.

Construction of an anthropometric discriminant model for identification of elite swimmers: An adaptive Lasso approach

Qile Pan[1], Rongxin Zhu[1], Jun Qiu[1], Guang Cai[1]

[1] Shanghai Research Institute of Sports Science, Shanghai, China

Corresponding Author:
Guang Cai[1]
87 Wuxing Road, Shanghai, 200030, China
Email address: leng8yang@126.com

Abstract

Background. Anthropometric characteristics are important factors that affect swimming performance. The aim of this study is to build a discriminant model using anthropometric factors to identify elite short-to-medium-distance freestyle swimmers through an adaptive Lasso approach.

Methods. The study recruited 254 swimmers (145 males and 109 females) who were divided them into elite (aged 17.9 ± 2.2 years, FINA points 793.8 ± 73.8) and non-elite (aged 17.1 ± 1.3 years, FINA points 560.6 ± 78.7) groups. Data for 73 variables were obtained, including basic information, anthropometric and derivative indicators. After filtering out highly correlated variables, 24 candidate variables were retained to be used in adaptive Lasso to select variables for prediction of elite swimmers. Deviance and area under the curve (AUC) were applied to assess the goodness of fit and prediction accuracy of the model, respectively.

Results. The adaptive Lasso selected 12 variables using the whole sample, with an AUC being 0.926 (95% CI: 0.895-0.956; $P=2.42\times10^{-29}$). In stratified analysis by gender, nine variables were selected for male swimmers with an AUC of 0.921 (95% CI: 0.880-0.963; $P=8.82\times10^{-17}$), and eight variables were for female swimmers with an AUC of 0.941 (95% CI: 0.898-0.984; $P=7.67\times10^{-15}$).

Conclusion. The adaptive Lasso showed satisfactory performance in selecting anthropometric characteristics to identify elite swimmers. Additional studies with longitudinal data or data from other ethnicities are needed to validate our findings.

Introduction

37   Anthropometric characteristics are widely considered as important factors affecting sport
38   performance (Barbosa et al., 2010; Saavedra et al., 2010; Ugarkovic et al., 2002).
39   Anthropometrics that matches the characteristics of a sport helps to master the movement
40   techniques, shorten training time, and plan targeted exercise program to achieve peak
41   performance (Putra et al., 2020). In swimming, arm length and girth, or the ratios (e.g., hand-
42   length/forearm-length and forearm-girth/wrist-girth) were found to be significantly associated
43   with the performance of breaststroke (Sammoud et al., 2018a), front crawl (Jurimae et al., 2007;
44   Latt et al., 2010) or freestyle (Nevill et al., 2015), backstroke (Sammoud et al., 2019), and
45   butterfly (Sammoud et al., 2018b), highlighting the limb advantage of longer levers. In addition,
46   leg length, trunk breadth (Sammoud et al., 2018a), and sitting height (Sammoud et al., 2019)
47   were also found to affect swimming performance, but their effect could vary by stroke. From
48   these findings, it is reasonable to hypothesize that anthropometric factors could be useful in the
49   identification of elite swimmers.
50
51   Indeed, previous studies have examined the feasibility of identifying elite swimmers by utilizing
52   anthropometric factors. However, in these studies, ordinary regression and discriminant analysis
53   were commonly used to explore the relationship between anthropometric factors and swimming
54   performance (Latt et al., 2009a, 2009b; Saavedra et al., 2010). The sample sizes are often very
55   limited, with a large number of different anthropometric factors available, and traditional
56   methods often perform poorly. In this study, we tackled the relationship between multiple
57   anthropometric factors and swimming performance using an adaptive Lasso (Least Absolute
58   Shrinkage and Selection Operator) method (Bhattacharya & McNicholas, 2014; Schneider &
59   Wagner, 2012), which has advantages in consistent variable selection. Thus, the main aim of this
60   study is to build a discriminant model using anthropometric factors to identify elite swimmers
61   through an adaptive Lasso approach.
62

Material & Methods

64   *Participants*

65   From 2016 to 2018, a total of 254 short-to-medium-distance (50-m event, 100-m event, 200-m
66   event, and 400-m event) (Robertson et al., 2009) freestyle swimmers were recruited, all of whom
67   were from two provincial teams and three university teams. Swimmers who attained the highest
68   level with a stroke other than freestyle or a distance of more than 400-m were excluded. Skeletal
69   age and menstrual status indicate that these swimmers have largely matured, minimizing the
70   effect of physical maturity on athletic performance (Vaeyens et al., 2008).
71   They were grouped into two categories, i.e., elite (achieving national grade or higher, FINA
72   points: 793.8 ± 73.8 at long course 50~400-m freestyle) and non-elite (achieving 1st grade or
73   lower, FINA points: 560.6 ± 78.7 at long course 50~400-m freestyle), according to their highest
74   competitive level which was achieved between 2009 and 2018 and was retrieved from the
75   official website of the General Administration of Sport of China

76 (http://app.gjzwfw.gov.cn/jmopen/webapp/html5/tyzjydyjsdjcx/index.html). The rating standards
77 of the training levels were shown in Supplementary **Table 1**.
78 All the study participants provided written informed consent. Ethical approval of the study was
79 obtained from the Ethics Committee of the School of Life Sciences of Fudan University
80 (No.497).
81
82 *Anthropometry*

83 Somatic measurements were taken by qualified anthropometrists following standardized
84 procedure as specified in the "Shanghai Sport Talent Selection Test Guide" (manual for internal
85 training, Supplementary **File 4**). There were 36 assessments: 17 for length, seven for breadth,
86 and 12 for circumference. A second-generation digital display Martin ruler (Martin-CHN,
87 Shanghai Quyi Technology Co., Ltd, Shanghai, China) and a standard soft tape measure were
88 used. The intra-assessment agreement (technical error of measurement, TEM) of these
89 measurements was 0.26-cm (95% CI: 0.20-0.32) (Hardy et al., 2018). Body composition data
90 were obtained by a multi-frequency body composition analyser (Inbody 720, Biospace, Seoul,
91 Korea).
92 Considering that season had minimal effect on our measurements (Putra et al., 2020), and the
93 cooperativeness of the athletes and coaches, each swimmer was measured only once in
94 December (i.e., off-season) during the study. Measurements were taken in the morning, starting
95 with a body composition test in a fasting situation. The rest of the anthropometric measurements
96 were conducted half an hour after eating. Of the swimmers tested during a day, we randomly
97 selected 3% for re-testing to verify reliability. Other basic information of the swimmers was
98 collected from the test record form.
99 In addition, 28 derivative indexes were calculated, and basic information, such as age of
100 attending the highest level, gender, and competitive level, were also collected, leading to a total
101 of 73 characteristic variables (Supplementary **Table 2**).
102
103 *Statistical analysis*

104 Among the 73 variables in this study, competitive level (elite vs. non-elite) is a binary dependent
105 variable, gender is a binary predictor, and the remaining 71 are continuous variables. The
106 pairwise correlation between the variables were calculated to eliminate variables that were highly
107 correlated (i.e., the absolute value of the correlation coefficient $|r| > 0.7$ in the corresponding
108 column).
109 The adaptive Lasso-penalized logistic model was then adopted to identify optimal
110 anthropometric variables for the identification of elite swimmers. Although this method may
111 produce some degree of asymptotic bias for the non-zero regression parameters, it can provide
112 consistent variable selection and is efficient in coefficient estimation (Das et al., 2019; Zheng et
113 al., 2016). Moreover, it can reduce estimation bias and improve prediction accuracy (Alhamzawi
114 & Ali, 2018). The model takes the following form:

115 $logit(\pi_i) = ln\left(\frac{\pi_i}{1-\pi_i}\right) = X^T\beta,$

116 where $\pi_i = P(y_i = 1|\boldsymbol{x_i})$ represents the probability of being an elite swimmer, given the

117 anthropometric characteristics, X is a n×p design matrix, and β is the corresponding coefficient

118 vector of p×1. β can be estimated using adaptive penalty by:

119 $\beta = \arg_\beta \min\left\{-\sum_{i=1}^{n}\left[y_i\ln(\pi_i)+(1-y_i)\ln(1-\pi_i)\right]+\lambda\sum_{i=1}^{\rho}\omega_i|\beta_i|\right\},$

120 where the penalty parameter λ is commonly determined using the k-fold cross-validation or the

121 Bayesian information criterion (BIC) (Bhattacharya & McNicholas, 2014; Schwarz, 1978), and

122 the penalty weights ω_i is generally taken as the inverse of the absolute value of the estimated

123 coefficients from the ordinary least square (OLS) or ridge regression (Alhamzawi & Ali, 2018).

124 In this study, we used 10-fold cross-validation to determine λ and OLS to determine ω_i. The

125 goodness of fit and prediction accuracy of the constructed model was assessed by deviance and

126 area under the curve (AUC), respectively.

127

128 Results

129 **Table 1** illustrates the basic information of the swimmers. A total of 71 continuous variables

130 composed a 71×71 lower triangular matrix with the diagonal and the above elements being zero.

131 The pairwise correlation analysis eliminated 48 continuous variables, leading to a total of 24

132 candidate variables (23 continuous plus gender) for subsequent analyses (supplementary **Table**

133 **3**).

134 ****Table 1 about here****

135 In the whole-sample model, *gender* was firstly converted into a dummy variable, with 0 being

136 male and 1 being female. The obtained coefficient solution path and the corresponding BIC were

137 shown in **Figure 1** and **2**, respectively, and the minimum BIC was -465.387. The penalty

138 parameter was then chosen to shrink the coefficients of some of the candidate variables to zero

139 (i.e., excluded from the model), and finally 12 out of the 24 variables were selected (**Table 2**).

140 ****Figure 1-3 about here****

141 According to the chosen variables, a logistic model was built and statistical inference was

142 performed (**Figure 3**). The predicted values were between 0 and 1, representing the probability

143 of becoming an elite athlete. The estimated coefficients of the selected variables were shown in

144 **Table 2**. This model explained 48.23% of the deviance. The receiver operating characteristics

145 (ROC) curve was shown in **Figure 4**, with an AUC (obtained by the bootstrapping) of 0.926

146 (95% CI: 0.895-0.956; P=2.42×10^{-29}).

147 **** Table 2 and Figure 4 about here****

148 In stratified analysis by gender, nine out of 23 variables were selected for male swimmers (**Table**

149 **3**), while eight variables for female athletes (**Table 4**). The model for male explained 46.49% of

150 the deviance, with an AUC of 0.921 (95% CI: 0.880-0.963; P=8.82×10^{-17}; **Figure 5**). The model

151 for female explained 54.85% of the deviance, with an AUC of 0.941 (95% CI: 0.898-0.984;
152 P=7.67×10⁻¹⁵; **Figure 6**).

153 **** Table 3 and Figure 5-6 about here****
154

Discussion

156 In this study, we built a mathematical model using anthropometric factors to identify elite
157 freestyle swimmers (≤400 m) through an adaptive Lasso approach. The adaptive Lasso selected
158 12 variables using the whole sample, and six of them were consistently selected in sex-specific
159 analysis. The models showed satisfactory performance in the identification of elite short-to-
160 medium-distance freestyle swimmers. To the best of our knowledge, this is the first study that
161 adopted the adaptive Lasso approach to identify elite swimmers using anthropometric
162 characteristics.

163 Fat free mass, which primarily consists of skeletal muscle, was reported to be one of the best
164 predictors of swimming performance (Latt et al., 2009a; Nevill et al., 2015). For example,
165 freestyle swimming speed could increase by 33.1% with an increase of 1-kg lean body mass
166 (Nevill et al., 2015). More lean body mass often means more appendicular lean mass as well as
167 more skeletal mass in the limbs, leading to greater stroke rate and subsequent propulsive force,
168 and finally quicker swimming speed (Nevill et al., 2015; Sammoud et al., 2018b). Consistent
169 with these findings, in our study, skeletal muscle mass was selected as a significant predictor of
170 elite swimmers: with 1-kg increase in skeletal muscle mass, the odds of becoming an elite
171 swimmer increased by 33% (OR=1.33, 95% CI: 1.17-1.54; P<0.0001) and 67% (OR=1.67, 95%
172 CI: 1.24-2.43; P=0.002) in the analysis using the whole sample and female swimmers,
173 respectively (**Table 2 & 4**). Although skeletal muscle mass was not significantly associated with
174 being an male elite swimmers, it still contributed to the prediction accuracy (OR=1.19, 95% CI:
175 1.00-1.44; **Table 3**). These findings corroborated the importance of muscularity component in
176 performance prediction and talent identification.

177 The trunk is an essential segment for swimming performance. Long torso was found to positively
178 affect backstroke swimming speed (Sammoud et al., 2019). Expanding to the overall competitive
179 performance (i.e., not restricted to one stroke and/or distance), torso length (sitting height) was
180 the only anthropometric variable selected in the multiple lineal regression model in male
181 swimmers, and sitting height explained 34% of the variance in swimming performance
182 (Saavedra et al., 2010). In our study, we found that torso length contributed to the prediction of
183 being an elite swimmer: with 1-cm increase in torso length, the odds of becoming an elite
184 swimmer increased by 34% (OR=1.34, 95% CI: 1.15-1.59; P<0.0001). Stratification analysis by
185 gender yielded similar results (**Table 3 & 4**). These results were further supported by previous
186 findings regarding the advantage of having shorter legs in freestyle swimming, as wave drag
187 could be reduced due to longer torso, leading to increased hull speed and decreased energy

188   consumption (Nevill et al., 2015; Papic & Sanders, 2018; Sammoud et al., 2019). Long torso is
189   thus a promising indicator of elite swimmers.
190   In addition to length of the trunk, horizontal size of the trunk also plays an important role in
191   swimming performance. Biacromial breadth (r=-0.61, p<0.01) and biiliocristal breadth (r=-0.46,
192   p<0.01) were significantly correlated with swimming sprint performance in young male
193   swimmers (Geladas et al., 2005). Furthermore, each 1-cm increase in biacromial breadth and
194   biiliocristal breadth increased 100-m butterfly speed by 48.9% and 29.2%, and 100-m
195   breaststroke speed by 56.5% and 40.3%, respectively (Sammoud et al., 2018a, 2018b). In our
196   study, trunk breadth ratio (i.e., bitrochanteric breadth/biacromial breadth) and trunk
197   circumference ratios (i.e., waist circumference/chest circumference, and waist circumference/hip
198   circumference) were selected as significant predictors: in the whole sample, a 1-unit increase in
199   the first two ratios decreased the odds of being an elite swimmer by 18.2% (OR=0.818, 95% CI:
200   0.698-0.947; P=0.009) and 23.2% (OR=0.768, 95% CI: 0.647-0.897; P=0.001), respectively;
201   while a 1-unit increase in the last ratio increased the odds by 27% (OR=1.27, 95% CI: 1.107-
202   1.474; P=0.001) (**Table 2**). Similar results for trunk circumference ratios were observed in
203   stratified analysis by gender (**Table 3 & 4**). These observations suggested that greater biacromial
204   breadth and chest circumference were associated with higher likelihood of being an elite
205   swimmer, while bitrochanteric breadth and hip circumference showed a reverse association with
206   the likelihood of being an elite swimmer. Our results, combined with the negative association of
207   head circumference, supported that a 'streamlined' body shape was essential to an elite swimmer
208   as it could reduce resistance during the underwater phase (Sammoud et al., 2018b). Together, our
209   findings suggested that coaches should take trunk into account as a key anthropometric predictor
210   of being an elite swimmer.
211   We found different indicators for male and female elite swimmers, implying that there may be
212   differences in anthropometric characteristics that affect swimming performance in male and
213   female swimmers. Previous literature attributed gender differences in swimming performance to
214   differences in somatic features (Knechtle et al., 2020). For example, having a higher lean body
215   mass and lower body fat in male swimmers compared to females was found to be significantly
216   associated with shoulder strength and swimming performance (Avlonitou et al., 1997). In
217   addition, anthropometric characteristics also contributed to differences in swimming
218   biomechanics (e.g., upper and lower segment coordination, energy cost, and swimming
219   economy) between male and female swimmers. For instance, female swimmers have a smaller
220   body size, which allows them to have lower body torque due to reduced distance between the
221   center of mass and the center of volume, leading to a smaller body cross-sectional area in the
222   water and improved swimming economy (Seifert et al., 2011).
223   Early (de)selection and early specialization is a concern for long-term athlete development, while
224   early diversification is considered as a remedy measure for early specialization. Both early age of
225   peak performance of swimmers and the "10-year rule" suggested the need for early
226   specialization of swimmers (Born et al., 2022). However, the unique performance environment
227   of swimming limits effectiveness of early diversification practices in transferring performance

228   elements and cross-training from other diversified sports (Baker, 2003). Nonetheless, dry-land
229   strength training programs are still important for swimmers as upper-body strength was found to
230   contribute significantly to the full stroke cycle: with 1-kg increase in the upper-body dry-land
231   strength, the thrust is increased by 0.397 N (95% CI: 0.189-0.605; P=0.002) (Morais, Marques,
232   et al., 2020).
233   This study has some limitations. We only considered anthropometric factors in the model, and
234   the included somatic indicators were not comprehensive. Future studies are needed to explore
235   whether adding other factors (e.g., physiological and bioenergetical factors) and other
236   anthropometric factors (e.g., frontal surface area, trunk transverse surface area, maximal rate of
237   change in cross-sectional area, and posterior form gradient of the waist-hip) that indicate actual
238   athletic swimming performance can further improve the performance of the prediction model
239   (Morais et al., 2011; Morais, Sanders, et al., 2020; Papic et al., 2020). In addition, considering
240   test efficiency, we only used the traditional soft tape to collect circumference features, while
241   several studies reported that new procedures, such as 3D scanners, could provide reliable and
242   accurate data (Rumbo-Rodríguez et al., 2021). We only used cross-sectional data in model
243   building. It is possible that changes in anthropometric factors over time may be better indicators
244   for identification of elite swimmers (Abbott & Collins, 2002). Future studies utilizing
245   longitudinal data and advanced analytical methods, such as joint modeling (Wulfsohn & Tsiatis,
246   1997) and time-varying effects modeling (Tan et al., 2012), might provide better predictive
247   performance. Joint modeling takes both longitudinal follow-up data and time-to-event
248   information into account, leading to more precise estimation and more accurate prediction than
249   modeling the two types of data separately (Koslovsky et al., 2020; Li & Luo, 2019; Rizopoulos
250   et al., 2016), while time-varying effects modeling targets intensive longitudinal data, and can
251   provide dynamic predictions when more longitudinal data are available, making it ideal for
252   identification of elite swimmers. To represent the homogeneity of swimmers as much as
253   possible, event specificity was not strictly controlled for in this study. Future studies are needed
254   to explore anthropometric factors associated with each specific event by recruiting swimmers
255   who participate in the same events after sample size calculation. In addition, our study is of
256   retrospective nature. Prospective studies that align sampling time with the time at which an
257   athlete achieves his/her highest level could provide more robust and accurate prediction. This
258   study was conducted on Chinese swimmers, and it is not known whether our findings can be
259   generalized to swimmers of other ethnicities. More studies with data from other ethnicities are
260   needed to validate our findings.
261
262   **Conclusions**

263   The present study adopted the adaptive Lasso method to select anthropometric factors to
264   establish a model for the prediction of elite short-to-medium-distance freestyle swimmers. Our
265   analysis pinpointed several anthropometric factors that were associated with being an elite
266   swimmer, such as torso length, leg length A to H, head circumference, ratio of waist to hip, and
267   skeletal muscle mass. The models showed satisfactory predictive performance, implying that

268   they might be helpful to coaches in the identification of elite freestyle swimmers in the late
269   adolescence. They also indicated that the adaptive Lasso approach can be applied in talent
270   identification in other sports.
271
272

273 **References**
274 Abbott, A., & Collins, D. (2002). A Theoretical and Empirical Analysis of a 'State of the Art' Talent
275   Identification Model. *High Ability Studies*, *13*(2), 157-178.
276   https://doi.org/10.1080/1359813022000048798
277 Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian adaptive lasso regression. *Math Biosci*, *303*, 75-82.
278   https://doi.org/10.1016/j.mbs.2018.06.004
279 Avlonitou, E., Georgiou, E., Douskas, G., & Louizi, A. (1997). Estimation of body composition in
280   competitive swimmers by means of three different techniques. *Int J Sports Med*, *18*(5), 363-368.
281   https://doi.org/10.1055/s-2007-972647
282 Baker, J. (2003). Early Specialization in Youth Sport: A requirement for adult expertise? *High Ability
283   Studies - HIGH ABIL STUD*, *14*, 85-94. https://doi.org/10.1080/13598130304091
284 Barbosa, T. M., Costa, M., Marinho, D. A., Coelho, J., Moreira, M., & Silva, A. J. (2010). Modeling the links
285   between young swimmers' performance: energetic and biomechanic profiles. *Pediatr Exerc Sci*,
286   *22*(3), 379-391. https://doi.org/10.1123/pes.22.3.379
287 Bhattacharya, S., & McNicholas, P. D. (2014). An Adaptive LASSO-Penalized BIC.
288   https://arxiv.org/pdf/1406.1332.pdf
289 Born, D.-P., Stäcker, I., Romann, M., & Stöggl, T. (2022). Competition age: does it matter for swimmers?
290   *BMC Research Notes*, *15*. https://doi.org/10.1186/s13104-022-05969-6
291 Das, D., Gregory, K., & Lahiri, S. N. (2019). Perturbation bootstrap in adaptive Lasso. *The Annals of
292   Statistics*, *47*(4). https://doi.org/10.1214/18-aos1741
293 Geladas, N. D., Nassis, G. P., & Pavlicevic, S. (2005). Somatic and physical traits affecting sprint swimming
294   performance in young swimmers. *Int J Sports Med*, *26*(2), 139-144. https://doi.org/10.1055/s-
295   2004-817862
296 Hardy, J., Kuter, H., Campbell, M., & Canoy, D. (2018). Reliability of anthropometric measurements in
297   children with special needs. *Arch Dis Child*, *103*(8), 757-762.
298   https://doi.org/10.1136/archdischild-2017-314243
299 Jurimae, J., Haljaste, K., Cicchella, A., Latt, E., Purge, P., Leppik, A., & Jurimae, T. (2007). Analysis of
300   swimming performance from physical, physiological, and biomechanical parameters in young
301   swimmers. *Pediatr Exerc Sci*, *19*(1), 70-81. https://doi.org/10.1123/pes.19.1.70
302 Knechtle, B., Dalamitros, A. A., Barbosa, T. M., Sousa, C. V., Rosemann, T., & Nikolaidis, P. T. (2020). Sex
303   Differences in Swimming Disciplines-Can Women Outperform Men in Swimming? *Int J Environ
304   Res Public Health*, *17*(10). https://doi.org/10.3390/ijerph17103651
305 Koslovsky, M. D., Hébert, E. T., Businelle, M. S., & Vannucci, M. (2020). A BAYESIAN TIME-VARYING
306   EFFECT MODEL FOR BEHAVIORAL MHEALTH DATA. *Ann Appl Stat*, *14*(4), 1878-1902.
307   https://doi.org/10.1214/20-aoas1402
308 Latt, E., Jurimae, J., Haljaste, K., Cicchella, A., Purge, P., & Jurimae, T. (2009a). Longitudinal development
309   of physical and performance parameters during biological maturation of young male swimmers.
310   *Percept Mot Skills*, *108*(1), 297-307. https://doi.org/10.2466/PMS.108.1.297-307
311 Latt, E., Jurimae, J., Haljaste, K., Cicchella, A., Purge, P., & Jurimae, T. (2009b). Physical development and
312   swimming performance during biological maturation in young female swimmers. *Coll Antropol*,
313   *33*(1), 117-122. https://www.ncbi.nlm.nih.gov/pubmed/19408614
314 Latt, E., Jurimae, J., Maestu, J., Purge, P., Ramson, R., Haljaste, K., Keskinen, K. L., Rodriguez, F. A., &
315   Jurimae, T. (2010). Physiological, biomechanical and anthropometrical predictors of sprint
316   swimming performance in adolescent swimmers. *J Sports Sci Med*, *9*(3), 398-404.
317   https://www.ncbi.nlm.nih.gov/pubmed/24149633
318 Li, K., & Luo, S. (2019). Dynamic predictions in Bayesian functional joint models for longitudinal and
319   time-to-event data: An application to Alzheimer's disease. *Stat Methods Med Res*, *28*(2), 327-
320   342. https://doi.org/10.1177/0962280217722177

321 Morais, J. E., Costa, M. J., Mejias, E. J., Marinho, D. A., Silva, A. J., & Barbosa, T. M. (2011). Morphometric
322   study for estimation and validation of trunk transverse surface area to assess human drag force
323   on water. *J Hum Kinet*, *28*, 5-13. https://doi.org/10.2478/v10078-011-0017-x
324 Morais, J. E., Marques, M. C., Rodríguez-Rosell, D., Barbosa, T. M., & Marinho, D. A. (2020). Relationship
325   between thrust, anthropometrics, and dry-land strength in a national junior swimming team.
326   *Phys Sportsmed*, *48*(3), 304-311. https://doi.org/10.1080/00913847.2019.1693240
327 Morais, J. E., Sanders, R. H., Papic, C., Barbosa, T. M., & Marinho, D. A. (2020). The Influence of the
328   Frontal Surface Area and Swim Velocity Variation in Front Crawl Active Drag. *Med Sci Sports*
329   *Exerc*, *52*(11), 2357-2364. https://doi.org/10.1249/mss.0000000000002400
330 Nevill, A. M., Oxford, S. W., & Duncan, M. J. (2015). Optimal Body Size and Limb Length Ratios
331   Associated with 100-m Personal-Best Swim Speeds. *Med Sci Sports Exerc*, *47*(8), 1714-1718.
332   https://doi.org/10.1249/MSS.0000000000000586
333 Papic, C., McCabe, C., Gonjo, T., & Sanders, R. (2020). Effect of torso morphology on maximum
334   hydrodynamic resistance in front crawl swimming. *Sports Biomech*, 1-15.
335   https://doi.org/10.1080/14763141.2020.1773915
336 Papic, C., & Sanders, R. (2018). Towards an understanding of how torso shape influences human
337   swimming performance. In (pp. Chapter 5).
338 Putra, R., Rahayu, T., Kasmini, W., & Sulaiman. (2020). *The Importance of Paying Attention to*
339   *Anthropometrics in the Selection Athletes* International Conference on Science and Education
340   and Technology (ISET 2019),
341 Rizopoulos, D., Taylor, J. M. G., Rosmalen, J. V., Steyerberg, E. W., & Takkenberg, J. J. M. (2016).
342   Personalized screening intervals for biomarkers using joint models for longitudinal and survival
343   data. *Biostatistics*, *17*(1), 149-164.
344 Robertson, E., Pyne, D., Hopkins, W., & Anson, J. (2009). Analysis of lap times in international swimming
345   competitions. *Journal of sports sciences*, *27*, 387-395.
346   https://doi.org/10.1080/02640410802641400
347 Rumbo-Rodríguez, L., Sánchez-SanSegundo, M., Ferrer-Cascales, R., García-D'Urso, N., Hurtado-Sánchez,
348   J. A., & Zaragoza-Martí, A. (2021). Comparison of Body Scanner and Manual Anthropometric
349   Measurements of Body Shape: A Systematic Review. *International Journal of Environmental*
350   *Research and Public Health*, *18*(12), 6213. https://www.mdpi.com/1660-4601/18/12/6213
351 Saavedra, J. M., Escalante, Y., & Rodriguez, F. A. (2010). A multivariate analysis of performance in young
352   swimmers. *Pediatr Exerc Sci*, *22*(1), 135-151. https://doi.org/10.1123/pes.22.1.135
353 Sammoud, S., Nevill, A. M., Negra, Y., Bouguezzi, R., Chaabene, H., & Hachana, Y. (2018a). 100-m
354   Breaststroke Swimming Performance in Youth Swimmers: The Predictive Value of
355   Anthropometrics. *Pediatr Exerc Sci*, *30*(3), 393-401. https://doi.org/10.1123/pes.2017-0220
356 Sammoud, S., Nevill, A. M., Negra, Y., Bouguezzi, R., Chaabene, H., & Hachana, Y. (2018b). Allometric
357   associations between body size, shape, and 100-m butterfly speed performance. *J Sports Med*
358   *Phys Fitness*, *58*(5), 630-637. https://doi.org/10.23736/S0022-4707.17.07480-1
359 Sammoud, S., Nevill, A. M., Negra, Y., Bouguezzi, R., Helmi, C., & Hachana, Y. (2019). Key somatic
360   variables in young backstroke swimmers. *J Sports Sci*, *37*(10), 1162-1167.
361   https://doi.org/10.1080/02640414.2018.1546547
362 Schneider, U., & Wagner, M. (2012). Catching Growth Determinants with the Adaptive Lasso. *German*
363   *Economic Review*, *13*(1), 71-85. https://doi.org/10.1111/j.1468-0475.2011.00541.x
364 Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2).
365   https://doi.org/10.1214/aos/1176344136
366 Seifert, L., Barbosa, T. M., & Kjendlie, P.-L. (2011). Biophysical approach to swimming: Gender effect.
367   *Gender gap: Causes, experiences and effects*, 59-80.

368 Tan, X., Shiyko, M. P., Li, R., Li, Y., & Dierker, L. (2012). A time-varying effect model for intensive
369   longitudinal data. *Psychol Methods*, *17*(1), 61-77. https://doi.org/10.1037/a0025814
370 Ugarkovic, D., Matavulj, D., Kukolj, M., & Jaric, S. (2002). Standard Anthropometric, Body Composition,
371   and Strength Variables as Predictors of Jumping Performance in Elite Junior Athletes. *The*
372   *Journal of Strength and Conditioning Research*, *16*(2). https://doi.org/10.1519/1533-
373   4287(2002)016<0227:Sabcas>2.0.Co;2
374 Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2008). Talent Identification and
375   Development Programmes in Sport. *Sports Medicine*, *38*(9), 703-714.
376   https://doi.org/10.2165/00007256-200838090-00001
377 Wulfsohn, M. S., & Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with
378   error. *Biometrics*, *53*(1), 330-339. https://www.ncbi.nlm.nih.gov/pubmed/9147598
379 Zheng, Q., Gallagher, C., & Kulasekera, K. B. (2016). Robust adaptive Lasso for variable selection.
380   *Communications in Statistics - Theory and Methods*, *46*(9), 4642-4659.
381   https://doi.org/10.1080/03610926.2015.1019138

382

Figure 1

The coefficient solution path of adaptive Lasso

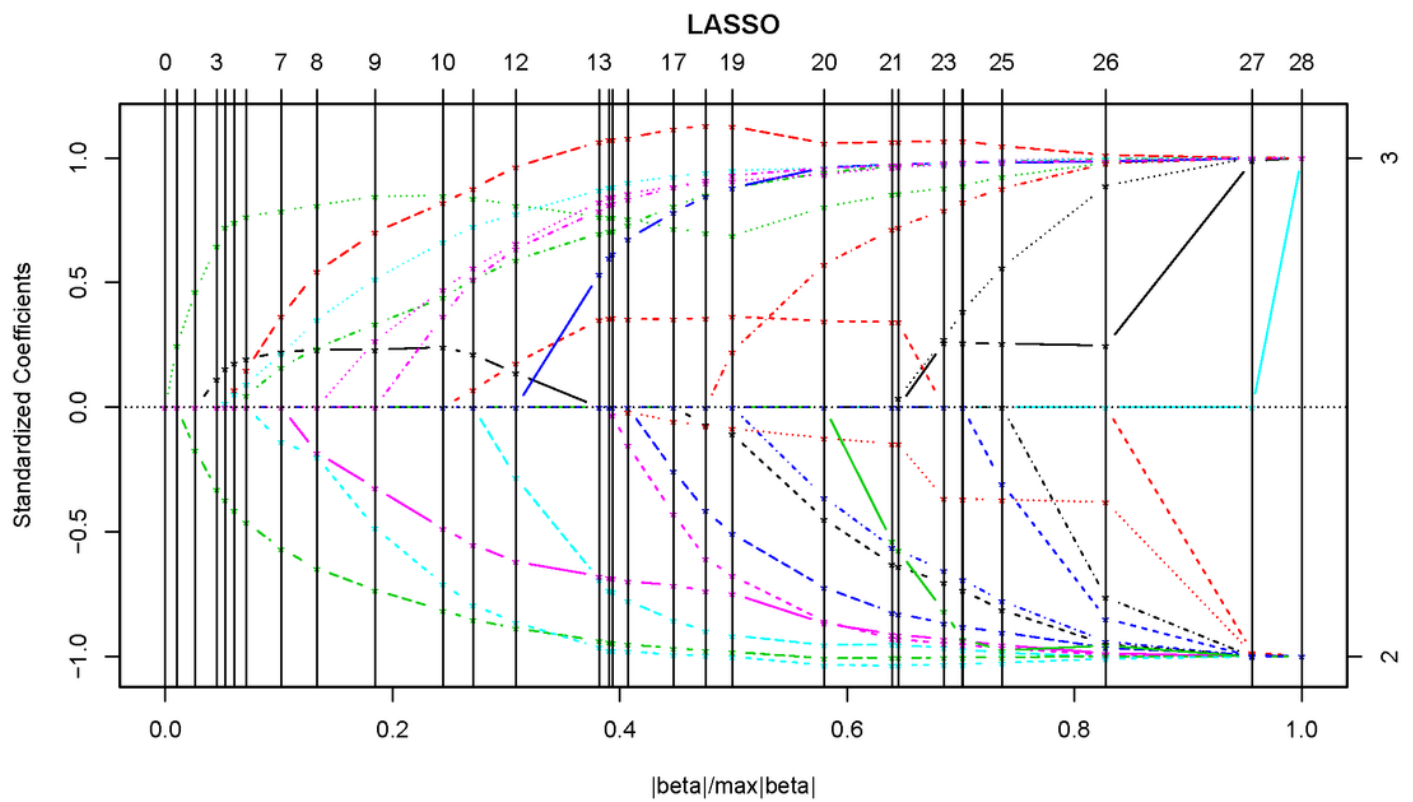



Figure 2

BIC variation with the number of entered variables.

BIC, the Bayesian Information Criterion. The minimum BIC was -465.387 when 15 variables entered.

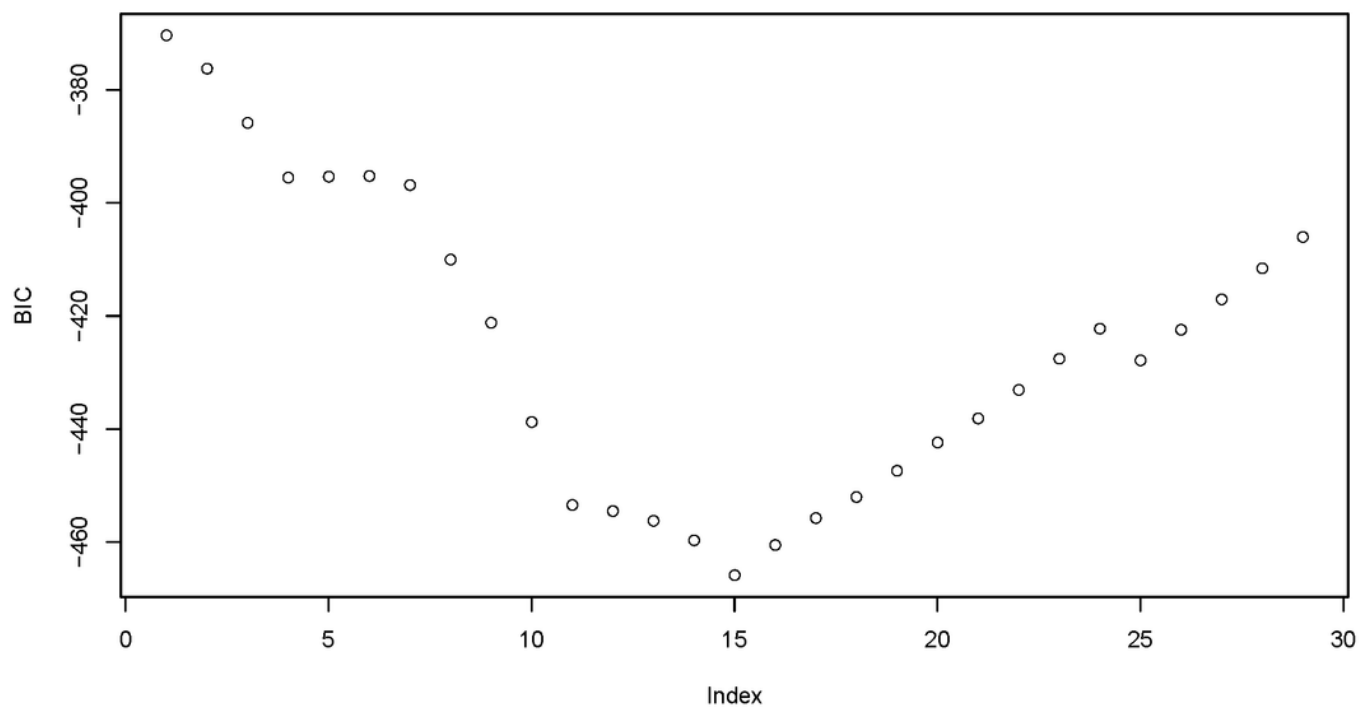



Figure 3

Fitted logistic graph of the probability of being an elite swimmer.

The predicted values were between 0 and 1, representing the probability of becoming an elite athlete.

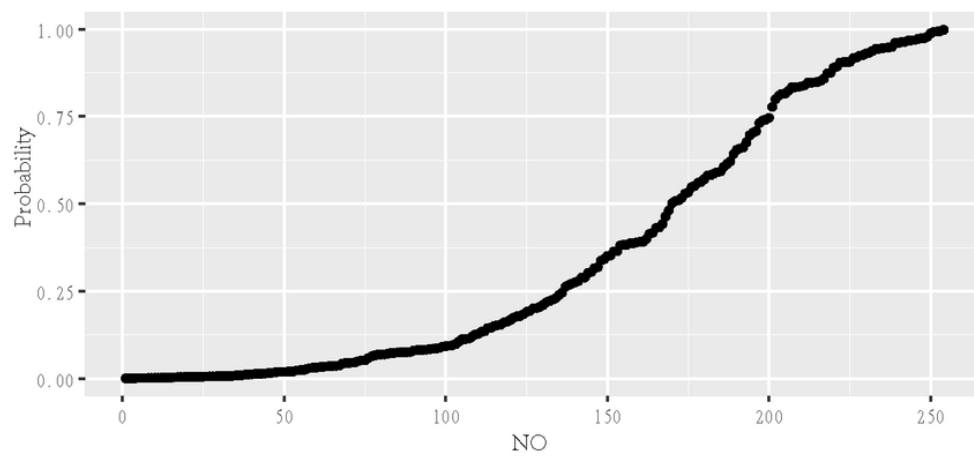



Figure 4

ROC curve of adaptive Lasso model using the whole sample.

This model explained 48.23% of the deviance, with an AUC (obtained by the bootstrapping) of 0.926 (95% CI: 0.895-0.956 ; P=2.42×10^{-29}) .

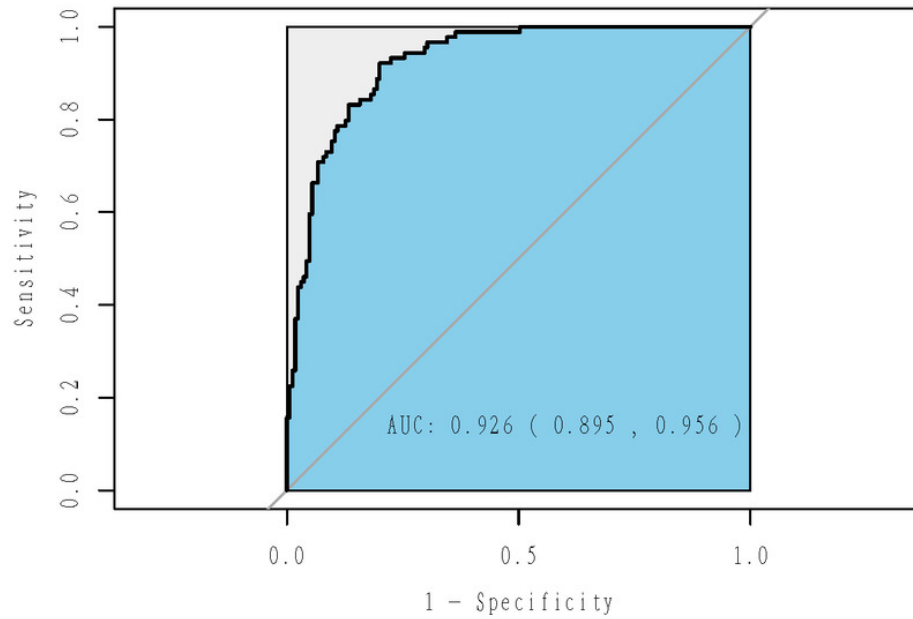



Figure 5

ROC curve of adaptive Lasso model for male swimmers.

The model for male explained 46.49% of the deviance, with an AUC of 0.921 (95% CI:

0.880-0.963 ; P=8.82×10^{-17}).

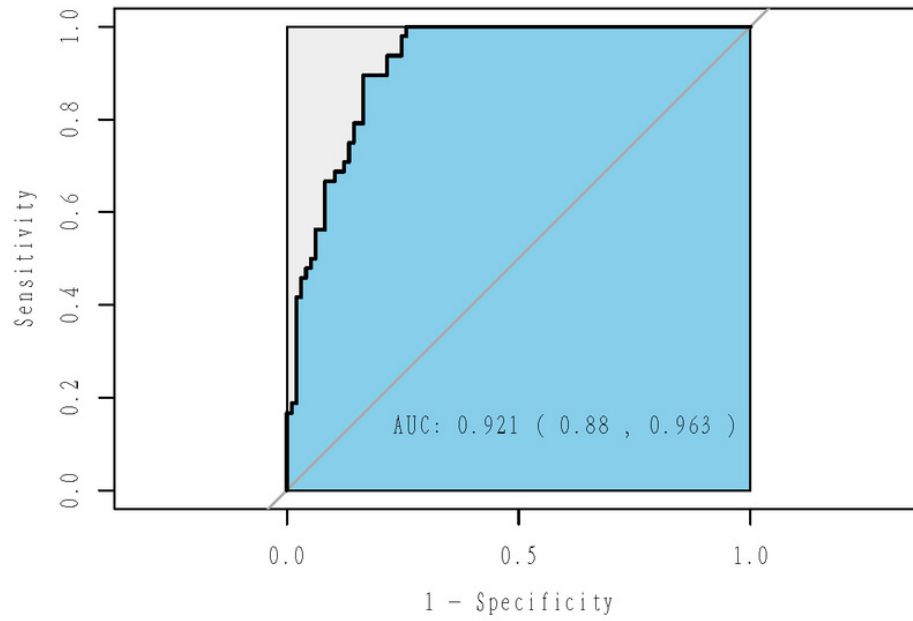



Figure 6

ROC curve of adaptive Lasso model for female swimmers

The model for female explained 54.85% of the deviance, with an AUC of 0.941 (95% CI:

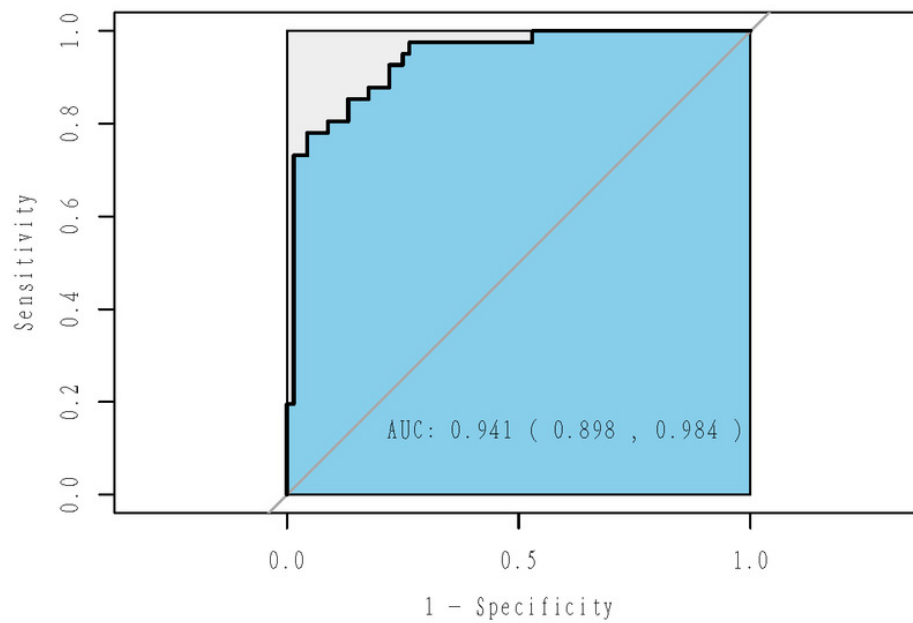
0.898-0.984 ; P=7.67×10^{-15}).

Basic information of the included swimmers.

Table 1 illustrates the basic information of the swimmers. Data were presented as mean±SD or n (%). The p-values were calculated by Welch two sample t-test and Pearson's Chi-squared test.

1 **Table 1. Basic information of the included swimmers.**

Variables	Elite swimmers (n=89)	Non-elite swimmers (n=165)	P
Age of attaining the highest level (years)	17.94±2.17	17.11±1.28	0.001
Gender, female (n, %)	41 (46%)	68 (41%)	0.5
FINA point	793.8±73.8	560.6±78.7	<0.001
Stature (cm)	179±8	176±7	<0.001
Weight (kg)	72±11	70±11	0.2
Back of foot height (cm)	7.58±0.53	7.50±0.58	0.3
Head circumference (cm)	56.47±1.85	57.14±1.60	0.004
Body fat percentage (%)	15±6	18±7	0.007
Skeletal muscle mass (kg)	34.5±6.4	32.3±6.3	0.009
Body mass index (BMI, kg/m2)	22.23±2.15	22.50±2.80	0.4
Ratio of arm to stature	44.18±0.84	43.80±0.96	0.001
Ratio of arm to leg length A	81.70±1.53	81.18±1.84	0.018
Torso length (cm)	54.6±3.2	52.4±3.0	<0.001
Torso length index	1.92±0.11	1.86±0.11	<0.001
Biceps circumference difference (cm)	2.15±0.85	1.84±0.80	0.004
Ratio of hand breadth to length	43.31±1.93	43.75±2.13	0.093
Ratio of ankle circumference to achillis tendon length	96±8	100±9	0.003
Ratio of upper to lower leg length	115.3±3.7	115.5±3.9	0.8
Ratio of leg length A to stature	54.09±1.04	53.97±1.11	0.4
Leg length A to H	93.09±1.00	92.67±0.95	0.001
Leg length B to H	87.47±0.88	87.37±1.12	0.4
Leg length C to H	76.50±1.26	75.95±1.26	0.001
Ratio of bitrochanteric to biacromial breadth	76.4±3.6	77.9±4.1	0.003
Ratio of biiliocristal to bitrochanteric breadth	86.9±2.9	86.6±3.3	0.5
Ratio of waist to chest	79.0±3.3	81.2±4.5	<0.001
Ratio of waist to hip	83.0±3.3	82.3±4.4	0.2
Ratio of foot breadth to length	36.08±1.75	36.29±1.68	0.3

2 Data were presented as mean±SD or n (%). The p-values were calculated by Welch two sample t-test and

3 Pearson's Chi-squared test.

4 SD, standard deviation.

5 BMI means weight / height^2 (kg/m^2); Arm means upper limb length (cm); Leg length A means iliospinale

6 posterior height (cm); Torso length means acromion height - trochanterion height (cm); Torso length

7   index means torso length / biiliocristal breadth; Biceps circumference difference means maximum biceps

8   circumference - minimum biceps circumference (cm); Upper leg length means thigh length (cm); Leg

9   length H means eristailiaca height (cm); Leg length B means trochanterion height (cm); Leg length C

10   means gluteal height (cm).

Logistic regression results using variables selected by adaptive Lasso (whole sample).

The penalty parameter was then chosen to shrink the coefficients of some of the candidate variables to zero (i.e., excluded from the model), and finally 12 out of the 24 variables were selected. OR, odds ratio; CI, confidence interval.

1 **Table 2. Logistic regression results using variables selected by adaptive Lasso (whole sample).**

	Coefficient	OR	95% CI		P
Intercept	-69.203	-	-		0.014
Age of attaining the highest level	0.304	1.355	1.052	1.789	0.024
Gender, male	5.301	200.591	32.352	1577.368	<0.0001
Ratio of arm to leg length A	0.190	1.209	0.942	1.566	0.141
Ratio of hand breadth to length	-0.153	0.859	0.683	1.064	0.174
Torso length	0.294	1.342	1.148	1.585	<0.0001
Leg length A to H	0.942	2.566	1.601	4.313	<0.0001
Ratio of bitrochanteric to biacromial breadth	-0.201	0.818	0.698	0.947	0.009
Head circumference	-0.787	0.455	0.32	0.624	<0.0001
Ratio of waist to chest	-0.264	0.768	0.647	0.897	0.001
Ratio of waist to hip	0.239	1.27	1.107	1.474	0.001
Biceps circumference difference	0.407	1.502	0.86	2.692	0.159
Skeletal muscle mass	0.287	1.332	1.165	1.542	<0.0001

2 OR, odds ratio; CI, confidence interval.

3 Arm means upper limb length (cm); Leg length A means iliospinale posterior height (cm); Torso length

4 means acromion height - trochanterion height (cm); Leg length H means eristailiaca height (cm); Biceps

5 circumference difference means maximum biceps circumference - minimum biceps circumference (cm).

Logistic regression results using variables selected by adaptive Lasso (male).

9 out of 23 variables were selected for male swimmers. OR, odds ratio; CI, confidence interval.

1   **Table 3. Logistic regression results using variables selected by adaptive Lasso (male).**

	Coefficient	OR	95% CI		P
Intercept	-129.538	-	**-**		0.003
Ratio of arm to stature	0.072	1.074	0.496	2.353	0.856
Ratio of arm to leg length A	0.063	1.065	0.71	1.596	0.760
Torso length	0.386	1.471	1.194	1.859	0.001
Leg length A to H	1.45	4.261	2.007	10.43	<0.0001
Head circumference	-0.812	0.444	0.256	0.708	0.002
Ratio of waist to chest	-0.426	0.653	0.488	0.836	0.002
Ratio of waist to hip	0.43	1.537	1.23	1.997	<0.0001
Biceps circumference difference	1.02	2.773	1.393	6.152	0.006
Skeletal muscle mass	0.176	1.192	1.001	1.443	0.056

2   OR, odds ratio; CI, confidence interval.

3   Arm means upper limb length (cm); Leg length A means iliospinale posterior height (cm); Torso length

4   means acromion height - trochanterion height (cm); Leg length H means eristailiaca height (cm); Biceps

5   circumference difference means maximum biceps circumference - minimum biceps circumference (cm).

Logistic regression results using variables selected by adaptive Lasso (female).

8 out of 23 variables were chosen for female athletes. OR, odds ratio; CI, confidence interval.

1 **Table 4. Logistic regression results using variables selected by adaptive Lasso (female).**

	Coefficient	OR	95% CI		P
Intercept	0.659	-	-		0.980
Ratio of arm to stature	-0.065	0.937	0.359	2.429	0.891
Ratio of arm to leg length A	0.357	1.429	0.892	2.373	0.146
Torso length	0.332	1.393	1.058	1.88	0.022
Head circumference	-0.918	0.399	0.201	0.694	0.003
Ratio of waist to chest	-0.391	0.676	0.51	0.853	0.002
Ratio of waist to hip	0.321	1.379	1.136	1.732	0.002
Skeletal muscle mass	0.512	1.669	1.239	2.434	0.002
Body fat percentage	-0.076	0.927	0.781	1.075	0.344

2 OR, odds ratio; CI, confidence interval.

3 Arm means upper limb length (cm); Leg length A means iliospinale posterior height (cm); Torso length

4 means acromion height - trochanterion height (cm).