

Bayesian Model Averaging for Predicting Maximal Oxygen Uptake in Athletes with Non-Exercise Data

Aldo Fabián Longo^{1,*}, Laura Pruzzo², Marcelo Luis Cardey¹,
Gustavo Daniel Aquilino¹, Enrique Oscar Prada¹,
and Rodolfo Juan Carlos Cantet³

ABSTRACT

Conventionally, non-exercise models to predict maximal oxygen uptake (VO_{2max}) have been built using the classical linear regression approach and frequentist techniques for model selection. However, uncertainty exists in the model selection process. The aim of this study was to develop a non-exercise model to predict VO_{2max} in athletes, considering model uncertainty by means of Bayesian Model Averaging (BMA). A further aim was to evaluate the predictive performance of the BMA in comparison to models derived from standard variable selection techniques. The data comprised 272 observations of the response variable, and records of Sex, Sport, Age, Weight, Height and Body mass index. A categorization of sports was also proposed for inclusion in the model-building process. BMA was applied based on two recognized methods: Occam's window and Markov Chain Monte Carlo Model Composition. Discordance was evident in variable selection among frequentist procedures. The two BMA strategies yielded comparable results. In agreement with the literature, the BMA showed better out-of-sample predictive performance than the models selected by standard techniques. The categorization of sports revealed consistent results.

Keywords: Data splitting, Model uncertainty, Predictive performance, Weighted mixture of Student's t-distributions.

Submitted: September 11, 2025

Published: November 03, 2025

 10.24018/ejsport.2025.4.6.254

¹Exercise Physiology Laboratory, National Center of High-Performance Athletics (CeNARD), Buenos Aires, Argentina.

²Department of Animal Production, College of Agriculture, University of Buenos Aires (UBA), Buenos Aires, Argentina.

³Research Institute for Animal Production (INPA), University of Buenos Aires-National Council of Science and Technology (UBA-CONICET), Buenos Aires, Argentina. National Academy of Agricultural and Veterinary Sciences, Buenos Aires, Argentina.

*Corresponding Author:
e-mail: alongo@deportesar.gob.ar

1. INTRODUCTION

The maximal oxygen uptake (VO_{2max}) is a key determinant of cardiorespiratory fitness. VO_{2max} , also referred to as maximal aerobic consumption, represents the highest rate at which oxygen can be taken in, distributed, and consumed by an individual's body during physical activity (Akalan *et al.*, 2004), and is a measure of the capability of transferring energy via the aerobic pathway (McArdle *et al.*, 2015). It is closely related to the physical ability known as Endurance, which is the physical and mental ability to resist fatigue in relatively long duration efforts, and the ability to quickly recover after the efforts (Grosser *et al.*, 1989; Zintl, 1991). Direct measurement under laboratory conditions is the gold standard for assessing VO_{2max} . However, it is complex and expensive because of the technological equipment and qualified human resources required. Consequently, a large variety of maximal and submaximal exercise tests have been designed for the indirect estimation of VO_{2max} , such as the Åstrand-Rhyming cycle ergometer protocol, Bruce treadmill test, timed run tests developed by Balke and Cooper, 1-mile steady-state jog of George *et al.*, and the 20-meter multistage shuttle run test of Léger *et al.* (Gibson *et al.*, 2019).

In contrast, statistical models have been proposed for predicting VO_{2max} using variables not related to exercise performance, which are also called non-exercise models. Maranhão Neto and Farinatti (2003) conducted a systematic historical review of the literature and reported 20 models built using non-exercise predictor variables. The predictors used were demographic data, anthropometric measures,

resting heart rate, smoking level, daily physical activity level, and perceived fitness. All the models were fitted using the classical linear regression approach. $\text{VO}_{2\text{max}}$ was expressed in absolute terms ($\text{L}\cdot\text{min}^{-1}$) and relative to body weight ($\text{ml}\cdot\text{kg}^{-1}\cdot\text{min}^{-1}$). Anthropometric and demographic predictors included age, sex, weight, height, body mass index, skinfold thicknesses, elbow diameter, leg volume, body surface, and percentage of body fat. The oldest models, published in 1971, were proposed by Shephard et al. (1971). A short time later, Bruce et al. (1973) were the first to use records of daily physical activity level in the model-building process. Among others, the models developed by Jackson et al. (1990) stand out because of the interest they have generated in the scientific community, covering an age range between 20 and 70 years. In addition, the review included the works of George et al. (1997) and Mathews et al. (1999), in which the values of the predictor variables were obtained by self-reporting. The models reported Adjusted R^2 values ranging from 0.22 to 0.87. Regardless of the variable selection technique used, the common denominator in all these studies is the fact that the uncertainty about the true model was not quantified. Subsequently, notwithstanding the machine learning algorithm implemented, non-exercise models for $\text{VO}_{2\text{max}}$ prediction were obtained following essentially the same statistical practice, i.e., without explicitly accounting for model uncertainty. More examples can be found in Malek et al. (2004), Bradshaw et al. (2005), Malek et al. (2005), Wier et al. (2006), Sanada et al. (2007), Duque et al. (2009), Nes et al. (2011), the overview of studies included in the work of Abut et al. (2016), and the review papers of Alzamer et al. (2021) and Ashfaq et al. (2022).

Conventionally, in a scenario with multiple candidate predictors, the final functional form of a linear regression model is the result of the implementation of standard variable selection methods. The emerging models are generally derived from selection criteria, such as Adjusted R^2 and Mallows' Cp, or from the implementation of selection variable algorithms, such as Forward, Backward or Stepwise (Clyde, 2003). It is well known that these methods may lead to different solutions (Weisberg, 2005). A main problem in the use of these selection strategies is that usually only one model is reported, virtually assuming that there is only one model to explain the variability of the data (Clyde, 2003; Raftery, 1995). Model uncertainty, which is inherent to the modeling process, is not formally considered for inference. Furthermore, the underestimation of model uncertainty involving the use of these procedures may result in overconfident inferences, either for the model parameters or for the prediction of future observations (Draper, 1995; Hodges, 1987; Hoeting et al., 1999; Raftery, 1996). The disadvantages of ignoring this uncertainty have been recognized by numerous authors (e.g., the collection of scientific articles edited by Dijkstra (1988)). Bayesian Model Averaging (BMA) (Leamer, 1978; Madigan & Raftery, 1994; Madigan & York, 1995) has been promoted in diverse disciplines as an alternative solution to incorporate model uncertainty into the analysis. According to the BMA approach, the competing models start with a prior probability and then obtain their posterior probabilities given the data sample. The resulting model is the average of the individual models weighted by their posterior probabilities (Hoeting et al., 1999).

In particular, the model uncertainty in the $\text{VO}_{2\text{max}}$ prediction with non-exercise data has not been explicitly considered. Conventionally, non-exercise models to predict $\text{VO}_{2\text{max}}$ have been built using the classical linear regression approach and frequentist techniques for model selection. Statistical analysis has generally been performed following the standard methodology; that is, once a model is chosen, the rest of the competing models are discarded, and the procedure continues as if the selected model has generated the data (Hoeting et al., 1999). Thus, only the uncertainty due to random errors is considered for inference, which is reflected in the confidence intervals for the model parameters and in the prediction intervals for future observations. Nonetheless, the model uncertainty has generally been underestimated in the statistical modeling of $\text{VO}_{2\text{max}}$. However, uncertainty regarding the functional form of the model in the field of linear regression may be substantial. More precisely, if k is the total number of potential predictors, the number of linear combinations between them is equal to 2^k (including the model with no predictors). For example, in the case of 15 predictors, the number of possible linear models reaches 32,768. On the other hand, BMA is a modern approach from a Bayesian perspective that provides a coherent mechanism to take into account model uncertainty in the analysis (Clyde, 2003). Comparative studies have shown that BMA has a higher predictive ability than any individual model selected using conventional procedures (Fernández et al., 2001a, 2001b; Hoeting et al., 1999; Madigan & Raftery, 1994; Raftery et al., 1996, 1997). Furthermore, 90% prediction intervals for future observations were constructed to compare the predictive performance of linear models obtained according to established criteria (Hoeting et al., 1999; Raftery et al., 1997, 2005; Wintle et al., 2003). The goal of this research was to develop a linear model for predicting $\text{VO}_{2\text{max}}$ (in $\text{L}\cdot\text{min}^{-1}$) in athletes from basic anthropometric and demographic data by means of BMA, as an alternative to the traditional frequentist techniques of model selection. A further goal was to compare the predictive performance of the BMA with those of models selected using standard procedures.

TABLE I: SUMMARY STATISTICS OF THE DATA

	Males (n = 187)	Females (n = 85)
VO _{2max} (L·min ⁻¹)	4.23 ± 0.72	2.98 ± 0.50
Age (years)	22.0 ± 4.7	22.2 ± 5.3
Weight (kg)	75.4 ± 11.5	61.3 ± 8.0
Height (m)	1.79 ± 0.09	1.67 ± 0.09
Body mass index (kg·m ⁻²)	23.5 ± 2.5	21.9 ± 1.7

Note: Data are reported as mean ± standard deviation.

2. MATERIALS AND METHODS

2.1. Subjects

Data used were records of 272 male and female athletes of the following sports disciplines: Athletics Races (Middle-distance and Long-distance Running), Boxing, Combined Winter Sports (Duathlon, Triathlon and Tetrathlon), Cross-country Skiing, Cycling, Kayaking, Field Hockey, Futsal, Handball, Judo, Karate, Rowing, Rugby, Speed Skating, Swimming, Taekwondo, Tennis, Volleyball and Wrestling. The database was provided by the Exercise Physiology Laboratory of the National Center of High-Performance Athletics (CeNARD) in Buenos Aires, Argentina. All procedures were conducted in accordance with the ethical principles of the Declaration of Helsinki of the World Medical Association ([World Medical Association \(WMA\), 2024](#)).

2.2. Study Design

This study used an observational cross-sectional design. Data were collected under laboratory conditions. VO_{2max} was assessed by maximal incremental exercise testing using either a treadmill, cycling ergometer, kayaking ergometer, or rowing ergometer. The VO₂ data were collected with the breath-by-breath method through a computerized open-circuit metabolic system (Medgraphics Cardiopulmonary Exercise System CPX/D, Breeze Ex v3.06 software; Medical Graphics Corporation, St. Paul, MN, USA). The VO₂ plateau was the primary criterion for the determination of VO_{2max} (VO₂ difference < 150 ml·min⁻¹ or 2.1 ml·kg⁻¹·min⁻¹ given an additional increment in work rate); secondary criteria were: exchange respiratory rate > 1.1 and heart rate ± 10 beats·min⁻¹ of the age-predicted maximal heart rate ([American College of Sports Medicine, 2009](#); [Howley et al., 1995](#); [O'Connor et al., 2009](#)). Age was computed in decimals. Weight and Height were measured using a height and weight scale (CAM 1001, Argentina) and expressed in kilograms and metres, respectively. And Body mass index was calculated as the ratio of weight to height squared (kg·m⁻²). [Table I](#) displays the sex-stratified summary statistics of the data.

2.3. Proposed Predictors

In order to predict VO_{2max} in athletes with non-exercise data, we evaluated six potential anthropometric and demographic explanatory variables. Sex and Sport were categorical variables, whereas the remaining four were continuous variables: Age, Weight, Height and Body mass index. Considering the large sample variability of sports, the small number of observations in some of them, and certain similarities, the following grouping strategy was considered. An elemental classification divides sports into two main groups: *Acyclic* and *Cyclic*. *Acyclic* sports involve varied and discontinuous motor actions that are typically performed at variable intensities, durations, and frequencies. Based on the available data, this group was subdivided into two categories: *Combat* sports (Boxing, Judo, Karate, Taekwondo and Wrestling) and *Game* sports (Field Hockey, Futsal, Handball, Rugby, Tennis and Volleyball). On the other hand, *Cyclic* sports, which are mostly classified as endurance sports, are disciplines such as Athletics Races, Cycling, Kayaking and Swimming. These sports are characterized by continuous and repetitive movement patterns and, in general, by a noteworthy contribution of the oxidative energy pathway. This grouping strategy is based on bioenergetic and biomechanical aspects and on competition characteristics, and is consistent with the works of [Neumann \(1988\)](#), [Platonov \(2001\)](#), and [Bompa and Haff \(2009\)](#). Furthermore, given the diversity of endurance sports included, a subdivision was proposed into two categories, taking into account the extent of development of aerobic power, which is strongly determined by the intrinsic characteristics of the discipline. A first level, denoted as *Endurance 1*, comprised Kayaking, Speed Skating and Swimming; a second level, referred to as *Endurance 2*, embraced Athletic Races (Middle-distance and Long-distance Running), Combined Winter Sports (Duathlon, Triathlon and Tetrathlon), Cross-country Skiing, Cycling and Rowing ([Åstrand et al., 2003](#); [Kenney et al., 2022](#)). Therefore, four categories for the factor Sport were defined: *Combat* (n = 48), *Game* (n = 89), *Endurance 1* (n = 51) and *Endurance 2* (n = 84).

2.4. Statistical Analysis

The records of VO_{2max} , Age, Weight, Height and Body mass index were initially summarized as the mean $\pm$ standard deviation. One dummy variable was generated for Sex, and three dummy variables were generated for Sport, which, together with Age, Weight, Height and Body mass index, totaled eight candidate predictors of VO_{2max} . First, we fitted the models using ordinary least squares (OLS) linear regression. Following [Raftery et al. \(1997\)](#) and [Hoeting et al. \(1999\)](#), the Maximum Adjusted R^2 and Minimum Mallows' Cp criteria and the Stepwise regression method were used to obtain the "best" subset of predictors. Stepwise regression was performed in two versions, according to the entry and stay significance level employed: $\alpha = 0.15$ and $\alpha = 0.05$. The Pearson correlation coefficient (r) was used to test the linear association between continuous variables, and multicollinearity was assessed via the variance inflation factor. Subsequently, the BMA method was applied. Due to the lack of literature on the relative plausibility of the different combinations of variables under the Bayesian approach, a neutral option was implemented, and BMA was carried out assuming equal prior probabilities for all variable combinations. In the first step, this was performed using the BIC approximation to compute the posterior model probabilities and the Occam's window procedure to select the models to be averaged ([Raftery, 1995](#)). For this purpose, we used the function *bicreg* of the BMA package ([Raftery et al., 2022](#)). The assumptions of the normal linear model and the study of advanced diagnostics for multiple regression in the selected models were also tested. In the second step, BMA was conducted based on the Markov Chain Monte Carlo Model Composition (MC^3) method, following the proposal of [Fernández et al. \(2001a\)](#), using the function *bms* in the BMS package ([Feldkircher et al., 2022](#)). However, given the moderate number of models to be averaged ($2^8 = 256$ possible linear combinations of predictor variables), the model averaging was computed by the complete enumeration of the model space instead of the approximation via the Markov Chain Monte Carlo sampling procedure. To compare the predictive performance of the models, 20 data splits were generated through repeated stratified random subsampling. Two-thirds of the data were assigned to the *Training* subset to build the models, and one-third of the data were assigned to the *Testing* subset to evaluate predictive performance ([Dobbin & Simon, 2011](#)). Subsequently, 90% prediction intervals were generated. In the models selected by the frequentist techniques, they were constructed following classical methodology ([Walpole et al., 2007](#)). To evaluate the predictive performance of the BMA via Occam's window, weighted mixtures of location-scale Student's t-distributions were computed with the function *rMit* in the AdMit package ([Ardia et al., 2022](#)), and the corresponding 90% prediction intervals were obtained with the function *quantile*. Given the implemented subsampling procedure, 90 weighted mixtures of location-scale Student's t-distributions were computed for each data split, generating a total of one thousand and eight hundred distributions. To evaluate the predictive performance of the BMA by MC^3 , the weighted mixtures of location-scale Student's t-distributions in the 20 data splits were made with the function *pred.density* in the BMS package, and the function *quantile* was used to obtain the corresponding 90% prediction intervals. All analyses were performed in R software environment version 4.4.0 ([R Core Team, 2024](#)).

3. RESULTS

3.1. Frequentist Analysis

The least squares linear regression on the full model yielded $R^2 = 0.8125$ and a residual standard error = $0.3848 \text{ L}\cdot\text{min}^{-1}$. It is worth noting the negligible contribution of the dummy variable corresponding to the sports category *Game*. On the other hand, the correlation structure among Weight, Height and Body mass index ($0.18 \leq r \leq 0.80$, $P < 0.01$) led to multicollinearity problems and over-parameterization of the model. The results of the linear regression analysis performed with all the candidate predictors are presented in [Table II](#).

As mentioned previously, there were two hundred and fifty-six possible linear regression models for fitting. Three popular techniques were used to select the "best" subset of predictors: Maximum Adjusted R^2 , Minimum Mallows' Cp and Stepwise regression. Stepwise regression was implemented in two ways, based on the entry and stay significance level employed: $\alpha = 0.15$ and $\alpha = 0.05$. The selected models are presented in [Table III](#). Model uncertainty was reflected in the discrepancies observed among the applied selection methods. According to the Maximum Adjusted R^2 and Minimum Mallows' Cp criteria, the best model included dummies for Sex *Male*, Sport *Endurance 1* and Sport *Endurance 2*, and Age, Height and Body mass index ($R^2 = 0.8124$; residual standard error = $0.3835 \text{ L}\cdot\text{min}^{-1}$). Instead, Stepwise regression selected the same dummy variables but not the same continuous variables. Moreover, the model derived from the Stepwise procedure with an entry and stay significance level of 0.15 included the continuous variables Age, Weight and Height ($R^2 = 0.8115$; residual standard error = $0.3844 \text{ L}\cdot\text{min}^{-1}$), while Weight was the only continuous variable retained when using an entry

TABLE II: LEAST SQUARES LINEAR REGRESSION ANALYSIS ON THE FULL MODEL

	Coefficient	SE	t value	p-value	VIF
<i>Intercept</i>	-5.7295	3.95	-1.45	0.15	
<i>X</i> ₁ : Sex Male	0.5274	0.07	7.63	< 0.001	1.89
<i>X</i> ₂ : Sport Game	0.0131	0.08	0.16	0.88	2.79
<i>X</i> ₃ : Sport	0.2920	0.08	3.57	< 0.001	1.88
<i>Endurance 1</i>					
<i>X</i> ₄ : Sport	0.6182	0.08	7.64	< 0.001	2.57
<i>Endurance 2</i>					
<i>X</i> ₅ : Age	0.0108	0.005	2.09	0.04	1.17
<i>X</i> ₆ : Weight	0.0097	0.03	0.35	0.72	211.70
<i>X</i> ₇ : Height	3.2426	2.24	1.45	0.15	105.74
<i>X</i> ₈ : Body mass index	0.1022	0.09	1.19	0.24	79.27

Note: SE: Standard error; VIF: Variance inflation factor.

TABLE III: MODELS SELECTED BY MAXIMUM ADJUSTED R^2 , MINIMUM MALLOW'S CP AND STEPWISE REGRESSION

	Coefficient	SE	t value	p-value	VIF
Adjusted R^2 and Mallow's Cp					
<i>Intercept</i>	-7.1327	0.50	-14.37	< 0.001	
<i>X</i> ₁ : Sex Male	0.5186	0.06	8.55	< 0.001	1.46
<i>X</i> ₃ : Sport	0.2854	0.06	4.52	< 0.001	1.12
<i>Endurance 1</i>					
<i>X</i> ₄ : Sport	0.6129	0.06	10.48	< 0.001	1.35
<i>Endurance 2</i>					
<i>X</i> ₅ : Age	0.0109	0.005	2.12	0.04	1.16
<i>X</i> ₇ : Height	4.0502	0.25	15.89	< 0.001	1.37
<i>X</i> ₈ : Body mass index	0.1322	0.01	12.43	< 0.001	1.21
Stepwise ($\alpha = 0.15$)					
<i>Intercept</i>	-1.1350	0.54	-2.10	0.04	
<i>X</i> ₁ : Sex Male	0.5384	0.06	8.92	< 0.001	1.44
<i>X</i> ₃ : Sport	0.2785	0.06	4.40	< 0.001	1.12
<i>Endurance 1</i>					
<i>X</i> ₄ : Sport	0.5961	0.06	10.22	< 0.001	1.34
<i>Endurance 2</i>					
<i>X</i> ₅ : Age	0.0106	0.005	2.06	0.04	1.16
<i>X</i> ₆ : Weight	0.0419	0.003	12.35	< 0.001	3.24
<i>X</i> ₇ : Height	0.6611	0.39	1.71	0.09	3.13
Stepwise ($\alpha = 0.05$)					
<i>Intercept</i>	-0.0722	0.15	-0.49	0.63	
<i>X</i> ₁ : Sex Male	0.5460	0.06	9.04	< 0.001	1.42
<i>X</i> ₃ : Sport	0.2783	0.06	4.37	< 0.001	1.12
<i>Endurance 1</i>					
<i>X</i> ₄ : Sport	0.6504	0.05	11.99	< 0.001	1.14
<i>Endurance 2</i>					
<i>X</i> ₆ : Weight	0.0463	0.002	20.46	< 0.001	1.42

Note: SE: Standard error; VIF: Variance inflation factor.

and stay significance level of 0.05 ($R^2 = 0.8073$; residual standard error = 0.3873 L·min⁻¹). However, the substantial decrease in the magnitude of the multicollinearity statistic in these three models is noteworthy.

3.2. Bayesian Model Averaging

First, BMA was performed using the BIC' approximation to compute the posterior model probabilities (PMP's) and the Occam's window procedure to select the models to be averaged. Table IV displays the location (post mean) and scale (post SD) measures for the posterior distributions of the regression coefficients of the model. Table IV also reports the posterior inclusion probability (PIP) of each of these coefficients, which is the probability that the coefficient value is other than zero given the data, and results from the sum of the PMP's of the models that contain that coefficient. Table V lists the nine selected individual models with their respective PMP's. As part of the analysis, the

TABLE IV: BMA MODELS DERIVED FROM THE OCCAM'S WINDOW AND MC³ METHODS

	Post mean	Post SD	PIP
Occam's window			
<i>Intercept</i>	-1.4531	2.67	1
X_1 : Sex <i>Male</i>	0.5405	0.06	1
X_2 : Sport <i>Game</i>	0.0019	0.02	0.04
X_3 : Sport <i>Endurance 1</i>	0.2811	0.06	1
X_4 : Sport <i>Endurance 2</i>	0.6406	0.06	1
X_5 : Age	0.0026	0.005	0.26
X_6 : Weight	0.0374	0.02	0.81
X_7 : Height	0.8066	1.55	0.29
X_8 : Body mass index	0.0237	0.05	0.26
MC ³			
<i>Intercept</i>	-1.4704	–	1
X_1 : Sex <i>Male</i>	0.5386	0.06	1
X_2 : Sport <i>Game</i>	0.0026	0.02	0.07
X_3 : Sport <i>Endurance 1</i>	0.2800	0.07	1
X_4 : Sport <i>Endurance 2</i>	0.6385	0.06	1
X_5 : Age	0.0026	0.005	0.26
X_6 : Weight	0.0370	0.02	0.82
X_7 : Height	0.8251	1.57	0.30
X_8 : Body mass index	0.0243	0.05	0.27

Note: post: posterior; SD: Standard deviation; PIP: Posterior inclusion probability.

TABLE V: POSTERIOR MODEL PROBABILITIES OF THE NINE MODELS SELECTED BY OCCAM'S WINDOW AND THE TEN BEST MODELS ACCORDING TO MC³

Model	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	PMP
Occam's window									
1	•		•	•		•			0.4624
2	•		•	•	•	•			0.1326
3	•		•	•			•	•	0.1175
4	•		•	•	•		•	•	0.0696
5	•		•	•		•	•		0.0684
6	•		•	•		•		•	0.0530
7	•		•	•	•	•	•		0.0360
8	•	•	•	•		•			0.0359
9	•		•	•	•	•		•	0.0246
MC ³									
1	•		•	•		•			0.4511
2	•		•	•	•	•			0.1244
3	•		•	•			•	•	0.1102
4	•		•	•		•	•		0.0652
5	•		•	•	•		•	•	0.0621
6	•		•	•		•		•	0.0506
7	•	•	•	•		•			0.0346
8	•		•	•	•	•	•		0.0326
9	•		•	•	•	•		•	0.0225
10	•	•	•	•	•	•			0.0110

Note: X_1 : Sex *Male*; X_2 : Sport *Game*; X_3 : Sport *Endurance 1*; X_4 : Sport *Endurance 2*; X_5 : Age; X_6 : Weight; X_7 : Height; X_8 : Body mass index; PMP: Posterior model probability.

assumptions of the normal linear model and the study of advanced diagnostics for multiple regression in the selected models were evaluated. Neither violations of the assumptions of the normal linear model nor influential observations were found. Furthermore, the VIF values did not indicate multicollinearity in the averaged models (maximum VIF = 3.24). The R^2 statistics in these models fluctuated between 0.8073 and 0.8124. As shown in Table V, the model chosen by Stepwise regression with $\alpha = 0.05$ showed the highest PMP. According to this criterion, the model selected by Maximum Adjusted R^2 and Minimum Mallows's Cp ranked fourth, while the model that emerged from the Stepwise regression with $\alpha = 0.15$ placed in the seventh position.

In the second step, BMA was conducted based on the MC³ method. The results obtained for the regression coefficients were similar to those achieved using the Occam's window method (see Table IV). Table V also reports the individual models with a PMP higher than 0.01 according to MC³. It is worth

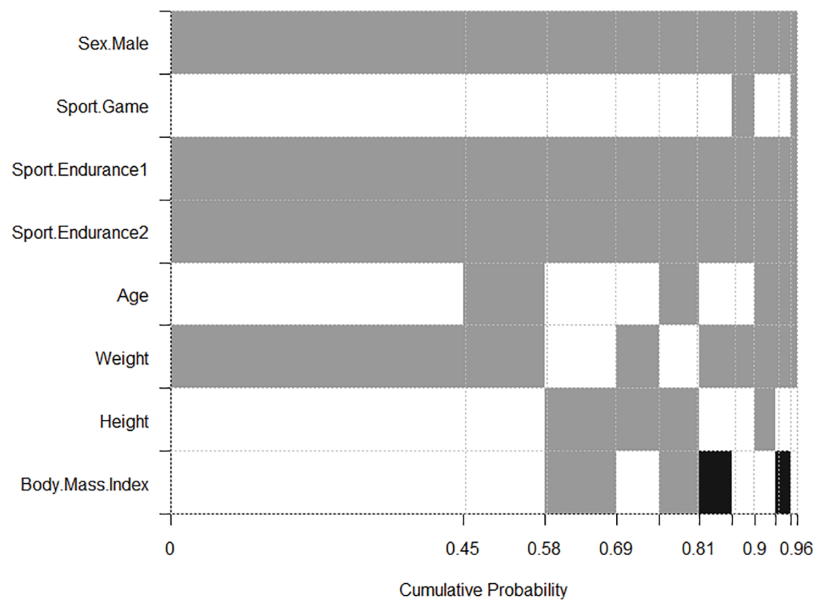


Fig. 1. Predictor variables and cumulative probability of the ten best models according to MC^3 ; the grey rectangles indicate a positive sign for the regression coefficient, the black rectangles a negative sign, and the white rectangles exclusion.

mentioning that, among the sets of predictors that included multicollinear variables (i.e., combinations including Weight, Height and Body mass index), the highest PMP was 0.0073; one of them was the full model, with a PMP equal to 0.0002. It can also be verified in Table V that the best nine models according to MC^3 are the same as those selected by Occam's window, and with very similar PMP's. Moreover, the ranking determined by the PMP for the models selected by frequentist techniques was virtually the same for the two BMA strategies applied. Additionally, Fig. 1 illustrates the contribution of the ten individual models with the highest weights in the BMA model obtained by MC^3 . The rows in the figure correspond to the variables, and the columns refer to the models. The models were located from left to right in descending order according to their PMP's. The predictors included in each model were identified on the vertical axis, whereas the horizontal axis displays the cumulative posterior model probabilities. The grey and black rectangles indicate that the predictor of that row is included in the model of the given column. The grey color indicates a positive sign for the regression coefficient, the black color indicates a negative sign for the regression coefficient, and the white rectangles indicate exclusion. The sum of the lengths of the grey and black rectangles corresponding to each predictor is approximately proportional to the PIP, as displayed in Table V.

3.3. Predictive Performance Comparison

The BMA models built using the Occam's window and MC^3 strategies showed, on average, a better predictive performance than the ones selected by Maximum Adjusted R^2 , Minimum Mallows C_p and Stepwise regression in the 20 data splits. Moreover, the two BMA models showed very similar abilities to predict future responses. They reached the best predictive coverage twice as often as the models chosen by the frequentist methods, and the worst one-third as often. Table VI presents a comparative summary of the predictive coverage of the models considered in the 20 data splits.

4. DISCUSSION

4.1. General Considerations

Several scientific papers have pointed out that underestimation of the uncertainty about the functional form of the statistical model may have negative consequences for inference. Thus, Raftery et al. (1997) and Hoeting et al. (1999) showed that ignoring model uncertainty leads to an overestimation of confidence in estimations. In the present study, a practical approach to firm theoretical grounds was used for inference with a normal linear regression model to predict VO_{2max} using non-exercise data, which explicitly takes into account model uncertainty in the modeling process. In this regard, BMA represents a coherent way to objectively consider model uncertainty in the analysis. Although this method has been applied in diverse research areas, no references have been found regarding its application for the statistical modeling of VO_{2max} with non-exercise data. The Bayesian methodology implemented provides a clear and accurate interpretation of the results and constitutes a direct

TABLE VI: BMA MODELS VERSUS MODELS SELECTED BY STANDARD FREQUENTIST METHODS: NUMBER OF TIMES WITH BEST AND WORST PREDICTIVE PERFORMANCE AND STATISTICAL SUMMARY OF THE PREDICTIVE COVERAGE OF THE 90% PREDICTION INTERVAL IN THE TWENTY DATA SPLITS

Method	Number of times		Predictive coverage (%)		
	Best	Worst	Minimum	Mean	Maximum
BMA (Occam's window)	14	3	81.1	88.9	96.7
Adjusted R ²	7	10	80.0	88.1	97.8
Mallow's Cp	8	9	82.2	88.2	97.8
Stepwise ($\alpha = 0.15$)	6	10	78.9	88.0	97.8
Stepwise ($\alpha = 0.05$) and BMA (MC ³)	9	9	78.9	88.2	96.7
BMA (MC ³)	14	3	80.0	88.9	96.7
Adjusted R ²	8	10	80.0	88.1	97.8
Mallow's Cp	8	9	82.2	88.2	97.8
Stepwise ($\alpha = 0.15$)	6	10	78.9	88.0	97.8
Stepwise ($\alpha = 0.05$)	8	9	78.9	88.2	96.7

instrument for posterior inference. Moreover, posterior model probabilities are a valuable formal means of weighting competing individual models. In addition, BMA preserves the essence of the Bayesian approach by allowing for an inferential interpretation of the model parameters. However, the literature employing frequentist solutions to incorporate model uncertainty in the analysis is far from extensive. A possible frequentist alternative cited by [Raftery \(1995\)](#) and [Hoeting et al. \(1999\)](#) is Bootstrap ([Efron, 1979](#)). However, [Freedman et al. \(1988\)](#) demonstrated that this technique does not necessarily yield satisfactory results.

In this research, BMA was performed following two strategies: on a reduced number of models (Occam's window), and exhaustively, taking into account all possible combinations of predictors (MC³). Common frequentist model selection techniques were also applied for comparison purposes. The first BMA strategy implemented, i.e., the BIC' approximation for the calculus of the posterior model probabilities and the Occam's window procedure for the identification of the models best supported by the data, has the advantage that it can be performed with the information provided in the output from the conventional statistical model-fitting software ([Raftery, 1995](#)). The second strategy applied was based on the proposal of [Fernández et al. \(2001a\)](#), which allows for the analytical computation of posterior model probabilities. It is worth mentioning that, even though the final functional form is in both cases a weighted average of models, the individual models (at least those with a substantial contribution to the final functional form) are not exempt from checking the assumptions of the normal linear model and from the analysis of the advanced diagnostics for multiple regression ([Raftery, 1995](#)).

4.2. Analysis and Interpretation of the Results

The output of the BMA analysis carried out using the Occam's window method was practically equivalent to that obtained by applying the MC³ method. Both procedures yielded comparable location and scale measures for the posterior distributions of the regression coefficients, as well as similar posterior model probabilities.

The BMA model confirmed the high explanatory power of Sex on VO_{2max}. Holding the values of all the other predictor variables constant, the estimated difference between males and females was 0.54 L·min⁻¹. This difference expressed relative to body weight represents 7.7 ml·kg⁻¹·min⁻¹ for a typical body weight of 70 kg. [George et al. \(1997\)](#) proposed a prediction model using non-exercise data in a population of physically active university students (18–29 years), which showed a similar difference between the sexes (7.0 ml·kg⁻¹·min⁻¹). In addition, [Wu and Wang \(2002\)](#) built a non-exercise model using data of 20- to 30-year-old workers that revealed a higher difference between males and females (1.27 L·min⁻¹). However, the latter study had a small sample size ($n = 24$). [Kenney et al. \(2022\)](#) published VO_{2max} normative data (in ml·kg⁻¹·min⁻¹) for athletes from diverse disciplines. The sex differences reported in these data were generally similar in direction and magnitude to the differences obtained from the BMA model predictions. Nonetheless, the grouping proposed for sports disciplines revealed reasonable results. The subclassification of *Acyclic* sports into *Combat* and *Game* sports found little support in the data. The BMA analysis assigned a very low PIP to this subdivision ($P < 0.1$), giving more weight to more parsimonious models resulting from combining both types of disciplines into the broader group of *Acyclic* sports. In contrast, the data strongly supported the subclassification proposed for endurance sports into the categories *Endurance 1* and *Endurance 2*; the BMA analysis revealed a PIP = 1 for the dummy variable corresponding to this partition. The difference in the values

of the linear parameters favored the category *Endurance 2* by $0.36 \text{ L} \cdot \text{min}^{-1}$. For a reference body weight of 70 kg and under equal values for the rest of the explanatory variables, this difference represents $5.1 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$, which also fits the normative values given by Kenney et al. (2022).

Regarding the continuous regressors, Weight was the most relevant predictor, with a PIP > 0.8 , while Age, Height and Body mass index showed lower predictive contributions in the BMA model, with PIP values between 0.26 and 0.30. On the other hand, as is evident from Fig. 1, when Body mass index and Height were entered into the same individual model, Body mass index was positively related to $\text{VO}_{2\text{max}}$, while this relationship became negative when Body mass index and Weight were entered into the same individual model. These results are congruent with the uncertainty about the statistical model that best explains the data-generation process. In addition, the exhaustive nature of the MC^3 strategy explains the small weights assigned to the models with a high level of multicollinearity. More specifically, the models including Weight, Height and Body mass index, barely accumulated a PMP of one percent (0.0121). The BMA via Occam's window did not include any of these models.

In terms of point estimation, the BMA model was contrasted with a widely known non-exercise model, that is, the body mass index-based model of Jackson et al. (1990). To attain a fair comparison, the highest value of the Physical Activity Rating (PAR) score was considered for the latter model, which corresponded to the highest level within the group of subjects who participated regularly in heavy physical exercise ($\text{PAR} = 7$). The continuous predictor variables were as follows: Age = 25 years, Weight = 68 kg, Height = 1.80 m, Body mass index = $21 \text{ kg} \cdot \text{m}^{-2}$. For this set of values, the body mass index-based model of Jackson et al. (1990) predicted a $\text{VO}_{2\text{max}}$ of $44.5 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ for females and $55.4 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ for males. In contrast, the $\text{VO}_{2\text{max}}$ predictions (related to body weight) produced by BMA for *Acyclic* sports (*Combat* and *Game*), *Endurance 1* sports and *Endurance 2* sports are, respectively, 45.7, 49.8 and $55.1 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ for females, and 53.6, 57.7 and $63.0 \text{ ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ for males.

It is worth mentioning that the results of either the frequentist or the Bayesian analyses resulted in fairly similar regression coefficients for the categorical variables, with the exception of the dummy corresponding to the sports category *Game*, which is absent in the models selected using the frequentist techniques. The difference in value of the coefficients from the two approaches were smaller than $0.05 \text{ L} \cdot \text{min}^{-1}$. Nevertheless, discrepancies were observed in the choice of continuous variables among the frequentist selection procedures (see Table IV). Interestingly, the model selected by the frequentist stepwise regression when $\alpha = 0.05$ was best supported by the data in terms of posterior model probability ($\text{PMP} \approx 0.5$). On the other hand, it is worth noting that the individual models with substantial weights in the BMA exhibited an appreciable fit, reaching R^2 values above 0.8.

One way to judge the validity of a model is to evaluate its ability to predict future responses (Raftery et al., 1996). Several scientific papers showed that BMA provides a higher predictive performance than any particular model that might reasonably be selected by a traditional technique (Madigan & Raftery, 1994; Raftery et al., 1996; Raftery et al., 1997; Hoeting et al., 1999; Fernández et al., 2001a; 2001b), and consistent results with this premise were found in the current research. The coverage of the 90% prediction interval of the BMA averaged 89% over the 20 data splits, against an average of 88% found in the models derived from the frequentist model selection strategies. Moreover, BMA generally reached the maximum predictive coverage recorded for each data split. Raftery et al. (1997) and Hoeting et al. (1999) assessed the out-of-sample predictive performance of linear regression modeling using the 90% prediction interval method. They found larger differences in the predictive coverage in favor of the BMA model in comparison with models selected by frequentist methods (between 2 and 22%). Raftery et al. (1997) also found differences as high as 6% in predictive coverage in favor of the MC^3 method over the Occam's window method. Nonetheless, the number of predictors in these studies was nearly twice the number of predictors considered in this study.

The underestimation of model uncertainty that entails the choice of a particular model to explain a determined phenomenon may affect the results of the statistical inference for the quantities of interest associated with the phenomenon (Hoeting et al., 1999). In the present study, this underestimation was reflected by a lower predictive coverage for new observations in the models selected by standard frequentist techniques, compared to the BMA models fitted either through Occam's window or the MC^3 strategy.

4.3. Consequences and Applications

Raftery (1995) pointed out that, given a wide set of candidate independent variables, the standard model selection techniques tend to find evidence for non-substantive effects; because of reasons related to statistical power, this trend becomes stronger with the increase in sample size. On the other hand, in BMA, all possible predictor combinations are weighted based on sample evidence. Simulation studies performed by Raftery (1995) and Raftery et al. (1997) showed that BMA tends to parsimony when there is no signal in the data suggesting a relationship between the predictors and the response variable.

With regard to the choice between the two BMA strategies we employed, the decision depends on the goal of the research, either parameter estimation or prediction. Occam's window tends to be computationally faster and more appropriate when the inference of the parameters in the model is the most important. However, the exhaustive nature of MC³ generates more accurate predictions with a higher computational demand. Nevertheless, these two approaches are sufficiently flexible to succeed in both situations (Raftery et al., 1997).

A non-exercise VO_{2max} prediction model represents a simple, practical and useful tool for sports evaluation. The current study developed a BMA model for predicting VO_{2max} in athletes using basic anthropometric and demographic data. Models obtained using frequentist variable selection techniques have also been reported. A categorization of sports was proposed for its inclusion in the model-building process, allowing it to cover a wide variety of disciplines. Moreover, no studies have been found in the literature on non-exercise models for athletic populations that include sports as an explanatory variable. In addition, the BMA framework offers a reasonable solution to the problem of adding more predictors to the modeling process: the larger the number of candidate variables, the larger the number of competing models, and thus, the greater the model uncertainty. Furthermore, considering the constant development in computational power, the BMA approach becomes natural. However, making use of prior information about the plausibility of the models to be averaged is a matter that deserves future investigation. It would also be advisable to collect more observations from different sports disciplines to evaluate more specific sports classifications, aiming to attain a higher explanatory power of VO_{2max} variability.

Overall, the implementation of BMA for the modeling of VO_{2max} with non-exercise data represents an original contribution, which is in line with the growth of the Bayesian approach in applied statistics.

5. CONCLUSIONS

Discordances were observed among frequentist techniques in the selection of available variables for predicting VO_{2max} in athletes. BMA provided a coherent and effective solution to the model uncertainty problem. By this method, all competing models were evaluated, taking into account the contributions of all variables. The combination of predictors with a high level of multicollinearity had very low posterior probabilities. The individual models that were best supported by the data displayed an appreciable fit. The BMA showed a higher predictive performance than the models derived from the least squares variable selection procedures. The frequentist and Bayesian approaches yielded similar VO_{2max} estimates for combat and game sports. Finally, the results obtained from both procedures support the proposed sub-classification for endurance sports.

ACKNOWLEDGMENT

The authors are particularly thankful to Néstor A. Lentini, Claudio A. Gillone, Enrique D. Balardini and Cristina Perez.

CONFLICT OF INTEREST

Authors declare that they do not have any conflict of interest.

REFERENCES

- Abut, F., Akay, M. F., & George, J. (2016). Developing new VO_{2max} prediction models from maximal, submaximal and questionnaire variables using support vector machines combined with feature selection. *Computer Biology Med*, 85, 182–192. <https://doi.org/10.1016/j.compbiomed.2016.10.018>.
- Akalan, C., Kravitz, L., & Robergs, R. R. (2004). VO_{2max}: Essentials of the most widely used test in exercise physiology. *ACSM's Health & Fitness Journal*, 8(3), 5–9. <https://doi.org/10.1097/00135124-200405000-00004>.
- Alzamer, H., Abuhmed, T., & Hamad, K. (2021). A short review on the machine learning-guided oxygen uptake prediction for sport science applications. *Electronics*, 10, 1956. <https://doi.org/10.3390/electronics10161956>.
- American College of Sports Medicine. (2009). *Guidelines for Graded Exercise Testing and Exercise Prescription*. 8th ed. Philadelphia, PA: Lippincott Williams & Wilkins.
- Ardia, D., Hoogerheide, L. F., & Van Dijk, H. K. (2022). *Adaptive Mixture of Student-t Distributions*. Version 2.1.9. <https://cran.r-project.org/package=AdMit>.
- Ashfaq, A., Cronin, N., & Müller, P. (2022). Recent advances in machine learning for maximal oxygen uptake (VO_{2max}) prediction: A review. *Informatics in Medicine Unlocked*, 28, 100863. <https://doi.org/10.1016/j.imu.2022.100863>.
- Åstrand, P. O., Rodahl, K., Dahl, H. A., & Strömme, S. B. (2003). *Textbook of Work Physiology: Physiological Bases of Exercise*. 4th ed. Champaign, IL: Human Kinetics.
- Bompa, T. O., & Haff, G. G. (2009). *Periodization: Theory and Methodology of Training*. 5th ed. Champaign, IL: Human Kinetics.

- Bradshaw, D. I., George, J. D., Hyde, A., LaMonte, M. J., Vehrs, P. R., Hager, R. L., & Yanowitz, F. G. (2005). An accurate $\text{VO}_{2\text{max}}$ nonexercise regression model for 18–65-year-old adults. *Res Q Exerc Sport*, 76(4), 426–432. <https://doi.org/10.1080/02701367.2005.10599315>.
- Bruce, R. A., Kusumi, F., & Hosmer, D. (1973). Maximal oxygen and nomographic assessment of functional aerobic impairment in cardiovascular disease. *American Heart Journal*, 85, 546–562. [https://doi.org/10.1016/0002-8703\(73\)90502-4](https://doi.org/10.1016/0002-8703(73)90502-4).
- Clyde, M. (2003). Model averaging. In S. J. Press (Ed.), *Subjective and objective Bayesian statistics: principles, models, and applications* (pp. 320–335). Hoboken, NJ: Wiley-Interscience.
- Dijkstra, T. K. (1988). *On Model Uncertainty and its Statistical Implications*. Berlin: Springer.
- Dobbin, K. K., & Simon, R. M. (2011). Optimally splitting cases for training and testing high dimensional classifiers. *BMC Med. Genomics*, 8, 4, 31. <https://doi.org/10.1186/1755-8794-4-31>.
- Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society: Series B*, 57, 45–97. <https://doi.org/10.1111/j.2517-6161.1995.tb02015.x>.
- Duque, I. L., Parra, J. H., & Duvallet, A. (2009). A new non exercise-based $\text{VO}_{2\text{max}}$ prediction equation for patients with chronic low back pain. *Journal of Occupational Rehabilitation*, 19(3), 293–299. <https://doi.org/10.1007/s10926-009-9180-5>.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1), 1–26. <https://doi.org/10.1214/aos/1176344552>.
- Feldkircher, M., Zeugner, S., & Hofmarcher, P. (2022). *Bayesian Model Sampling and Averaging*. Version 0.3.5. <https://cran.r-project.org/package=BMS>.
- Fernández, C., Ley, E., & Steel, M. F. J. (2001a). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100, 381–427. [https://doi.org/10.1016/S0304-4076\(00\)00076-2](https://doi.org/10.1016/S0304-4076(00)00076-2).
- Fernández, C., Ley, E., & Steel, M. F. J. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Econometrics*, 106, 563–576. <https://doi.org/10.1002/jae.623>.
- Freedman, D. A., Navidi, W., & Peters, S. C. (1988). On the impact of variable selection in fitting regression equations. In T. K. Dijkstra (Ed.), *On model uncertainty and its statistical implications* (pp. 1–16). Berlin: Springer.
- George, J. D., Stone, W. J., & Burkett, L. N. (1997). Non-exercise $\text{VO}_{2\text{max}}$ estimation for physically active college students. *Medicine and Science in Sports and Exercise*, 29, 415–423. <https://doi.org/10.1097/00005768-199703000-00019>.
- Gibson, A. L., Wagner, D. R., & Heyward, V. H. (2019). *Advanced Fitness Assessment and Exercise Prescription*. 8th ed. Champaign, IL: Human Kinetics.
- Grosser, M., Brüggemann, P., & Zintl, F. (1989). *Alto rendimiento deportivo: planificación y desarrollo*. Barcelona: Ediciones Martínez Roca.
- Hodges, J. S. (1987). Uncertainty, policy analysis and statistics. *Statistical Science*, 2, 259–275. <https://doi.org/10.1214/ss/1177013224>.
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14, 382–417. <https://www.jstor.org/stable/2676803>.
- Howley, E. T., Bassett, D. R. Jr. & Welch, H. G. (1995). Criteria for maximal oxygen uptake: Review and commentary. *Medicine and Science in Sports and Exercise*, 27, 1292–1301. <https://doi.org/10.1249/00005768-199509000-00009>.
- Jackson, A. S., Blair, S. N., Mahar, M. T., Weir, L. T., Ross, R. M., & Stuteville, J. E. (1990). Prediction of functional aerobic capacity without exercise testing. *Medicine & Science in Sports & Exercise*, 22, 863–870. <https://doi.org/10.1249/00005768-199012000-00021>.
- Kenney, W. L., Wilmore, J. H., & Costill, D. L. (2022). *Physiology of Sport and Exercise*. 8th ed. Champaign, IL: Human Kinetics.
- Leamer, E. E. (1978). *Specification Searches: Ad hoc Inference with Nonexperimental Data*. New York, NY: John Wiley & Sons.
- Madigan, D., & Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the American Statistical Association*, 89, 1535–1546. <https://doi.org/10.2307/2291017>.
- Madigan, D., & York, A. E. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, 63, 215–232. <https://doi.org/10.2307/1403615>.
- Malek, M. H., Housh, T. J., Berger, D. E., Coburn, J. W., & Beck, T. W. (2004). A new non-exercise-based $\text{VO}_{2\text{max}}$ prediction equation for aerobically trained females. *Medicine & Science in Sports & Exercise*, 36(10), 1804–1810. <https://doi.org/10.1249/01.MSS.0000142299.42797.83>.
- Malek, M. H., Housh, T. J., Berger, D. E., Coburn, J. W., & Beck, T. W. (2005). A new non-exercise-based $\text{VO}_{2\text{max}}$ prediction equation for aerobically trained men. *Journal of Strength and Conditioning Research*, 19(3), 559–565. <https://doi.org/10.1519/00124278-200508000-00013>.
- Maranhão Neto, G. de A., & Farinatti, P. de T. V. (2003). Non-exercise models for prediction of aerobic fitness and applicability on epidemiological studies: Descriptive review and analysis of the studies. *Revista Brasileira de Medicina do Esporte*, 9, 315–324. <https://www.scielo.br/rbme/a/WTH3WzPvQ7gbjMZbJTtLyNH/?lang=en&format=pdf>.
- Mathews, C. E., Heil, D. P., Freedson, P. S., & Pastides, H. (1999). Classification of cardiorespiratory fitness without exercise testing. *Medicine & Science in Sports & Exercise*, 31, 486–493. <https://www.ncbi.nlm.nih.gov/10188755>.
- McArdle, W., Katch, D., & Katch, V. L. (2015). *Exercise Physiology: Energy, Nutrition, and Human Performance*. 8th ed. Philadelphia, PA: Wolters Kluwer Health | Lippincott Williams & Wilkins.
- Nes, B. M., Janszky, I., Vatten, L. J., Nilsen, T. I., Aspenes, S. T., & Wisløff, U. (2011). Estimating $\text{VO}_{2\text{peak}}$ from a nonexercise prediction model: The HUNT study. *Norway Medicine & Science in Sports*, 43(11), 2024–2030. <https://doi.org/10.1249/MSS.0b013e31821d3f6f>.
- Neumann, G. (1988). Special performance capacity. In A. Dirix, H. G. Knuttgen, K. Tittel (Eds.), *The olympic book of sports medicine* (pp. 97–108). Oxford: Blackwell Scientific Publishing.
- O'Connor, F. G., Kunar, M. T., & Deuster, P. A. (2009). Exercise physiology for graded exercise testing: A primer for the primary care clinician. In C. H. Evans, R. D. White (Eds.), *Exercise testing for primary care and sports medicine physicians* (pp. 3–21). New York, NY: Springer.
- Platonov, V. M. (2001). *Teoría general del entrenamiento olímpico deportivo*. Barcelona: Editorial Paidotribo.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111–163. <https://doi.org/10.2307/271063>.
- Raftery, A. E. (1996). Approximate Bayes factor and accounting for model uncertainty in generalised linear models. *Biometrika*, 83, 251–266. <https://doi.org/10.1093/biomet/83.2.251>.
- Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133, 1155–1174. <https://doi.org/10.1175/MWR2906.1>.
- Raftery, A., Hoeting, J., Volinsky, C., Painter, I., & Yeung, K. Y. (2022). *Bayesian Model Averaging*. Version 3.18.17. <https://cran.r-project.org/package=BMA>.
- Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92, 179–191. <https://doi.org/10.1080/01621459.1997.10473615>.
- Raftery, A. E., Madigan, D., & Volinsky, C. T. (1996). Accounting for model uncertainty in survival analysis improves predictive performance (with discussion). In J. Bernardo, J. Berger, A. Dawid, A. Smith (Eds.), *Bayesian statistics*. 5 (pp. 323–349). Oxford: Oxford University Press.

- R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org>.
- Sanada, K., Midorikawa, T., Yasuda, T., Kearns, C. F., & Abe, T. (2007). Development of nonexercise prediction models of maximal oxygen uptake in healthy Japanese young men. *European Journal of Applied Physiology*, 99(2), 143–148. <https://doi.org/10.1007/s00421-006-0325-3>.
- Shephard, R. J., Weese, C. H., & Merriman, J. E. (1971). Prediction of maximal oxygen intake from anthropometric data. *Internationale Zeitschrift Fur Angewandte Physiologie, Einschliesslich Arbeitsphysiologie*, 29, 119–130. <https://doi.org/10.1007/BF00698022>.
- Walpole, R. E., Myers, R. H., & Myers, S. L. (2007). *Probability and Statistics for Engineers and Scientists*. 8th ed. London: Pearson Prentice Hall.
- Weisberg, S. (2005). *Applied Linear Regression*. 3rd ed. New York, NY: John Wiley & Sons.
- Wier, L. T., Jackson, A. S., Ayers, G. W., & Arenare, B. (2006). Nonexercise models for estimating VO₂max with waist girth, percent fat, or BMI. *Medicine and Science in Sports and Exercise*, 38(3), 555–561. <https://doi.org/10.1249/01.mss.0000193561.64152>.
- Wintle, B. A., McCarthy, M. A., Volinsky, C. T., & Kavanagh, R. P. (2003). The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, 17, 1579–1590. <https://doi.org/10.1111/j.1523-1739.2003.00614.x>.
- World Medical Association. (WMA). (2024). *WMA Declaration of Helsinki-Ethical Principles for Human Medical Research*. 75th WMA General Assembly, Helsinki, Finland. <https://www.wma.net/policies-post/wma-declaration-of-helsinki>.
- Wu, H. C., & Wang, M. J. J. (2002). Establishing a prediction model of maximal oxygen uptake for young adults. *Journal of the Chinese Institute of Industrial Engineers*, 19, 1–7. <https://doi.org/10.1080/10170660209509197>.
- Zintl, F. (1991). *Entrenamiento de la resistencia*. Barcelona: Ediciones Martínez Roca.