As accepted for publication in Pediatric Exercise Science, ©Human Kinetics
http://doi.org/10.1123/pes.2017-0249

Responding to the commentary on the article: "Improving the prediction of maturity from anthropometric variables using a maturity ratio".

Job Fransen, Adam Baxter-Jones and Stephen Woodcock

[1]Sport and Exercise Science, Faculty of Health, University of Technology Sydney, Sydney, Australia

[2]College of Kinesiology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

[3]School of Mathematical and Physical Sciences, University of Technology Sydney, Sydney Australia.

We appreciate the commentary of Nevill and Burton [1] on our recent manuscript "Improving the prediction of maturity from anthropometric variables using a maturity ratio" [10]. The constant interaction between academics, each with their own view and expertise, should be considered 'best practice' and we therefore thank these authors for their interesting insights into this multidisciplinary study.

We strongly believe the primary criticism uttered in Nevill and Burton's [1] commentary lacks validity and creates confusion about the difference between a statistical model for hypothesis testing and a statistical model for prediction. It is our contention, and one widely held in the statistical community to which some of the authors of our original manuscript belong, that these two categories of models can have very different properties and should be judged very differently [4, 5, 6]. For example, the statistician Galit Shmueli [6] summarises this issue as follows:

"As a discipline, we must acknowledge the difference between explanatory, predictive and descriptive modelling, and integrate it into statistics education of statisticians and non-statisticians".

More specifically, Nevill and Burton [1] indicated there were several issues with the novel model we proposed for predicting maturity in boys from an athletic and non-athletic population. They argued that 1) the model we created artificially inflates the explained variance because Chronological Age (CA) is included as both a predictor and response variable, 2) a multilevel modelling approach would have been more appropriate given the structure of our dataset, 3) multiplicative allometric models rather than polynomials would improve the model fit, and 4) a simplified model could be developed that appears to be more user-friendly. Therefore, we will use the following paragraphs to reply to these specific comments

1) Inflated explained variance caused by spurious correlations

Our modelling philosophy, while at odds with the opinions of Nevill and Burton [1], is in line with that of leading experts in the field of predictive modelling. For example, the classic text by Makridakis, Wheelwright and Hyndman [7] explicitly states that

"Multicollinearity is not a problem unless either (i) the individual regression coefficients are of interest, or (ii) attempts are made to isolate the contribution of one explanatory variable to Y, without the influence of the other explanatory variables. Multicollinearity will not affect the ability of the model to predict."

In our model, while Age at Peak Height Velocity (APHV) and Chronological Age are independent, our response variable that consists of their ratio is likely to be highly correlated with Chronological Age as correctly argued by Nevill and Burton [1]. Indeed, any model built on the assumption of independence between these variables would be badly flawed. Nevill and Burton [1] highlight their argument by using examples of studies in which spurious correlations might have influenced models to test a null hypothesis of no association between the resulting variables. However, this is not analogous to the work in our paper [6]. It is not detrimental to a predictive model to have a response variable that is highly correlated with a predictor variable. However, Nevill and Burton [1] point out that, in such cases, extremely high R-squared values are likely to be obtained but these are not "artificially high" as stated in their commentary. Moreover, their size is not artificial or incorrect, but rather a consequence of the model structure and its correlations. Additionally, while we quote R-squared values, we are not seeking to confuse explanatory power with predictive power and nor should our readers, as clearly indicated in the title of our manuscript. Therefore, although R squared values are cited, it is crucial that our readers understand that model selection was not carried out by looking at the explained variance of our model, but rather a likelihood based criterion such as the Aikaike Information Criterion (AIC)

Nevill and Burton [1] indicate that it may be worthwhile to exclude CA from one side of the equation. However, accommodating this request would not be feasible in the context of our model. However, this does not make any sense for the reasons stated below:

i) The model developed in this paper is intended as a tool to be implemented by researchers or practitioners conducting field-based research. The CA of a test subject would, in practice, always be recorded. Therefore, deliberately throwing away this information to satisfy the arbitrary requirement that the model would need a certain structure so it can be used for rather academic hypothesis testing as well as pure prediction, seems wasteful and highly impractical.

ii) There is, obviously, predictive power in the Chronological Age variable over and above its direct impact on the calculated Maturity Ratio. By excluding Chronological Age as a predictor, any resulting model would implicitly be allocating the same maturity ratio to a 12 and 15 year old boy, which would obviously contradict biological reality.

Finally, Nevill and Burton [1] attempt to strengthen their argument by using a numerical example (Figure 1 in [1]). However, we believe that this again creates confusion about the difference between models for hypothesis testing and models for prediction. These authors clearly show that the ratio of two independent variables can be correlated to one of those variables. This is rather obviously correct, but not analogous to the data in our study. Any model built for testing dependence between variables should rightly give the result that there is indeed now a correlation. However, this does not mean that this spurious result could be used to build a meaningful predictive model. Finding a correlation between the ratio of Chronological Age to APHV would not be of any assistance in predicting APHV given Chronological Age itself as there is no predictive power in knowing one value of a pair of independent variables. In our paper, we clearly demonstrate that there is significant predictive power by including terms dependent on Chronological Age as predictors, suggesting that the correlation structure in the model is much more complex than in the example used by Nevill and Burton [1], perhaps as Chronological Age could also be a proxy for other variables not measured in this study. Additionally, we believe the example used by Nevill and Burton [1] represents a 'straw man argument' as it is clearly very different and much simpler compared to our own. Therefore, the reasoning that their example reveals an error and deriving from that that our conclusions must thus be made in error, is obviously lacking logic.

2) Lacking a multilevel approach

Our model is further criticised by Nevill and Burton [1] for not taking a multilevel approach. These authors use their own [8] and Baxter-Jones et al. [10]'s example to demonstrate the need to use a multilevel approach in our study, yet fail to recognise that in these studies, the response variables that were modelled and the data used to build the models is obviously different. Therefore, this deductive argument is quite short sighted.

Our decision not to develop such a model was not an oversight or error on our part, rather a deliberate decision made in the context of the likely uses and applications of the resulting model. As the primary focus of this project was to refine and improve a currently used tool, the model was tailored to the data with which it would most likely be used. A multilevel model approach would likely provide better estimates of a given player's predicted APHV in the presence of multiple observations. This, however, is not the kind of dataset on which we expect the model to be employed. The tool is a refinement of the earlier work of Mirwald et al. [2], which is commonly employed to provide APHV predictions from potentially as little as a single observation (for example, when an athlete first enters an academy). Therefore, we were more interested in being able to make out of sample predictions (predictions in new data that were not used to build the current equation), rather than future within sample predictions (predictions in the same sample used to build the equation). If we would have developed a multilevel model, we would have had to build in a crude assumption, likely setting the player-level intercept simply to be the mean across the whole dataset. However, this would be an extremely blunt tool and of less practical value than the model presented in our manuscript. Furthermore, we were not convinced that such a model would prove to be very stable given the complexity and dimensionality of the dataset and the relatively few longitudinal measurements available.

3) Multiplicative allometric models would create a better model fit

Nevill and Burton [1] argue that the use of multiplicative allometric models would *almost certainly* create a better model fit than the additive polynomials used in our study. While we do not disagree these models *could potentially* provide a better fit, we do not endow the same level of certainty to these claims. Therefore, further research into the use of multiplicative allometric models is needed.

4) Developing a simplified model

Furthermore, Nevill and Burton [1] indicate that the original model in our study could be simplified. We would like to thank them for this suggestion. We believe that the practicality of our model could be greatly improved while minimising human error in using the prediction equation, if the equation could be modified to its most simple form. Therefore, we modified our prediction equation to the following equation as per Nevill and Burton's suggestions:

$$\textit{Maturity ratio} = 6.99 + (0.154 \times CA - 0.242) + (0.00452 \times \textit{Body Mass}) - (0.0000341 \times \textit{Body Mass}^2) - (0.152 \times \textit{Stature}) + (0.000933 \times \textit{Stature}^2) - (0.00000166 \times \textit{Stature}^3) + (0.0322 \times \textit{Leg Length}) - (0.000269 \times \textit{Leg Length}^2) - (0.000761 \times [\textit{Stature} \times CA])$$

We tested this new model on a new sample of 87 male soccer players (12.05 ± 0.60 years) and found the following:

The maturity ratio using our original equation [10] for this sample is: 0.87 ± 0.05 years. The maturity ratio for the modified equation is nearly identical at 0.86 ± 0.05 years. Resultantly, the APHV for this sample based on the original equation [10] is 13.9 years ± 0.4 compared to very similar at 14.0 ± 0.4 using the modified equation. The argumentation used by Nevill and Burton [1] as well as the similarity between the predictions obtained from both equations indicate that this modified and simplified equation could be used.

Conclusion

In conclusion, we believe we have provided a rationale and sufficient argumentation in support of our modelling method, mainly by emphasising the difference between an explanatory model for hypothesis testing and a predictive model. We would also like to stress the increasingly obvious need for adopting a multi-disciplinary approach to studies that require the modelling of complex and large data sets. Clearly, the role played by statisticians and applied mathematicians in thoroughly arguing mathematic modelling techniques is an invaluable asset to sport science research, especially with the availability of increasing amounts of data.

References

1. Commentary on the article “Improving the prediction of maturity from 1 anthropometric variables using a maturity ratio”. (2017*), Ped Exerc Sci* , *30*(1): xx:xx

2. Mirwald, R. L., Baxter-Jones, A. D., Bailey, D. A., & Beunen, G. P. (2002). An assessment of maturity from anthropometric measurements. *Medicine & Science in Sports & Exercise*, *34*(4), 689-694.

3. Moore, S. A., McKay, H. A., Macdonald, H., Nettlefold, L., Baxter-Jones, A. D., Cameron, N., & Brasher, P. M. (2015). Enhancing a somatic maturity prediction model. *Med Sci Sports Exerc*, *47*(8), 1755-64.

4. Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*(5): p. 815-824.

5. Greenberg, E. and R.P. Parks. (1997). A Predictive Approach to Model Selection and Multicollinearity. *Journal of Applied Econometrics*, *12*(1), 67-75.

6. Shmueli, G. (2010). To Explain or To Predict. *Statistical Science*, *84*(3), 289-310.

7. Makridakis, S., S.C. Wheelwright, and R.J. Hyndman. (1998). Forecasting: Methods and Applications, New York: Wiley.

8. Nevill, A. M., Holder, R. L., Baxter-Jones, A., Round, J. M., & Jones, D. A. (1998). Modelling developmental changes in strength and aerobic power in children. *Journal of Applied Physiology*, *84*(3), 963-970.

9. Baxter-Jones, A., Goldstein, H., & Helms, P. (1993). The development of aerobic power in young athletes. *Journal of Applied Physiology*, *75*(3), 1160-1167.

10. Fransen, J., Bush, S., Woodcock, S., Novak, A., Deprez, D., Baxter-Jones, A. D., ... & Lenoir, M. (2017). Improving the Prediction of Maturity From Anthropometric Variables Using a Maturity Ratio. *Pediatric Exercise Science*, 1-28.