## Scientific Letter

### Uncertainty Quantification in Medicine Science: The Next Big Step

*To the Director,*

Current technological progress opens the door to measuring ever more biological processes and identifying bio-mechanism patterns at an unprecedented level of resolution. Faced with the ensuing avalanche of information, the promises of digital, precision and personalized medicine are increasingly fulfilled,[1] and their success largely depends on the intensive use of data-driven approaches, particularly those stemming from the confluence of statistics and machine learning research. However, nowadays the application of statistics and machine learning in many modeling tasks, from specific risk injury prediction to general physiology,[2] remains limited and constitutes a topic of great controversy that has divided experts about their ability to obtain useful results.[3–7]

A critical point which has gone largely unnoticed in the medicine field is the need to quantify the predictive limits of the models through uncertainty analysis. Practitioners and researchers tend to focus on performing inferential analysis to measure the uncertainty regarding the conditional mean (the expected value for a particular characteristic given that a certain set of conditions is known to occur), which presents important limited practical interpretations as the sample size grows.[8] However, they usually ignore the conditional distribution, which is more informative in practice and helps to establish the reliability of the predictive results obtained from different sets of values for model inputs.

In our opinion, it would be advisable to provide, as the output of any predictive model, not only the point estimates, for example, of the conditional mean, but also a prediction band estimated by using the heteroscedastic signal of the predictive model. This prediction band is built as an estimate of the interval in which a future observation will fall, with a certain probability, given what has already been observed. We exemplify it from the results of our previous study on the submaximal prediction of maximal oxygen uptake in a general population.[9] Following previous functional regression models,[10] this time we quantify uncertainty using the conformal inference framework,[11,12] which ensures to construct valid prediction bands with respect to the coverage error.[13]

Fig. 1 (left) shows the prediction intervals for these individuals. In general, we observe a significant heterogeneity, indicating that the signal noise is heteroscedastic, and the uncertainty is high in the prediction results. This phenomenon happened despite the global predictive ability of the reported models, being reasonable in terms of the coefficient of determination ($R^2 = 0.80$). In a more specific analysis, we try to predict with generalized additive models (GAM) the factors on which the radius of the prediction intervals depends. Let us recall that GAM is a generalized linear model in which the response variable depends linearly on unknown smooth functions of some predictor variables.[14] After testing the linearity assumption, age is shown to exhibit a linear dependence, and with the increase of one year of age, the radius increases by 0.05, meaning that 40-year-old individuals versus 20-year-old individuals increase their radius in one point. In terms of peak oxygen (Fig. 1 (center)), we observe that when peak oxygen increases the uncertainty decreases, and a similar behavior can be observed when the weight increases (Fig. 1 (right)) in the range of 50–70 kg. Importantly, the $R^2$ of this model is equal to 0.16. According to this regression model, the uncertainty is heterogeneous and varies according to the level and physical conditions of the individuals. Ultimately, in those individuals with high-uncertainty values, the model's usefulness may be questioned.
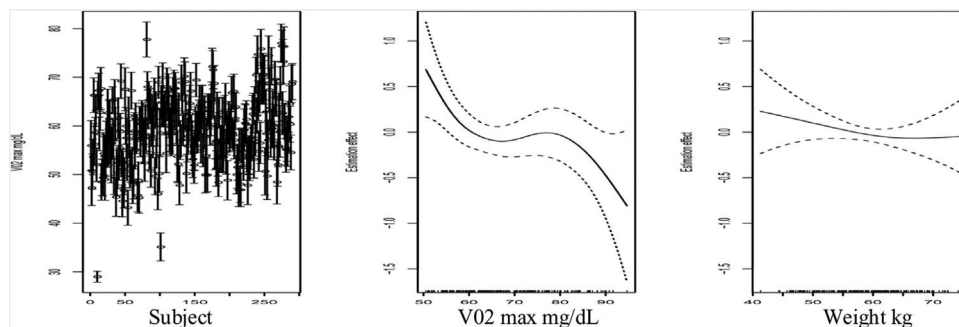


**Fig. 1.** (Left) Prediction interval of maximum peak oxygen with conformal inference techniques in all subjects. (Center) Additive effect of a GAM concerning oxygen consumption in the prediction of the radius of the prediction interval. (Right) Additive effect of a GAM concerning weight in the prediction of the radius of the prediction interval. Age, maximum peak oxygen and weight are statistically significant variables in the GAM model. Sex is not a statistically significant variable. The *R*-squared of the model is equal to 0.16. We use a linear link and a gaussian distribution as random error.

On the other hand, a closer inspection of the level of uncertainty in the prediction results can provide a crucial information for clinical management. As we have highlighted in a previous application in modeling five-year glucose changes in the general population, a phenotypically characterization of those subpopulations for which the model provides an unreliable prediction may be used to guide a specific approach, built on new assumptions, measurement procedures and interventions.[1,2,9] Particularly, we showed that long-term changes in glycated hemoglobin cannot be adequately predicted for individuals with elevated fast plasma glucose (FPG) levels, and the same holds true for individuals with FPG levels in the normoglycemic range and overweight. For these specific subpopulations we can then suggest a more personalized follow-up.

In conclusion, while the current era of statistics and machine learning holds great promise and excitement, it faces significant challenges in various scientific domains, particularly in the biomedical field, due to the inherent high variability in predictive tasks. However, it is crucial not to succumb to pessimism. We advocate for a stronger emphasis on uncertainty quantification as a crucial driver for improving predictive models.

A comprehensive evaluation of uncertainty levels in predictions and the subsequent characterization of the reliability conditions of these models should serve as a valuable guide to identify additional requirements for measurement, particularly in the context of personalized follow-up where the risk of diseases and uncertainty are both high. Recently, there has been ongoing debate surrounding the communication of uncertainty and its varying impacts on individuals and different formats.[15]

Ultimately, the next significant stride in predictive models within medicine relies not solely on a simplistic competition based on accuracy measurements,[16] but rather on the potential insights that lie within the realm of uncertainty. The COVID-19 pandemic confirms this fact, highlighting the limitations of simplistic utilization of systematic statistical models,[17,18] and presenting opportunities for uncertainty quantification in different contexts. For example, in the case of seroprevalence estimations in Spain, a new stochastic conformal simulation model provides a more comprehensive understanding of pandemic control.[19] Exciting new knowledge awaits discovery within this uncertain landscape, and the revolution of this statistical modeling is already underway.

## Authors' contributions

MM and MC drafted the article. All authors revised the article in depth and approved the submission to *Medical Education*.

## Ethics approval

No human subjects were included for the purpose of the present work; therefore, no ethics approval was needed.

## Funding

None.

## Conflict of interests

None declared.

## References

1. Matabuena M, Salgado FJ, Nieto-Fontarigo JJ, Álvarez-Puebla MJ, Arismendi E, Barranco P, et al. Identification of asthma phenotypes in the Spanish MEGA cohort study using cluster analysis. Arch Bronconeumol. 2023;59:223–31.
2. Matabuena M, Hayes PR, Puente-Maestu L. Prediction of maximal oxygen uptake from submaximal exercise testing in chronic respiratory patients. New perspectives. Arch Bronconeumol. 2019;55:507–8.
3. Bittencourt NF, Meeuwisse WH, Mendonça LD, et al. Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition—narrative review and new concept. Br J Sports Med. 2016;50:1309–14.
4. Bullock GS, Hughes T, Sergeant JC, et al. Methods matter: clinical prediction models will benefit sports medicine practice, but only if they are properly developed and validated. Br J Sports Med. 2021;55:1319–21.
5. Bullock GS, Hughes T, Arundale AH, et al. Black box prediction methods in sports medicine deserve a red card for reckless practice: a change of tactics is needed to advance athlete care. Sports Med. 2022;52:1729–35.
6. McCall A, Fanchini M, Coutts AJ. Prediction: the modern-day sport-science and sports-medicine "quest for the holy grail". Int J Sports Physiol Perform. 2017;12:704–6.
7. Nielsen RO, Shrier I, Casals M, et al. Statement on methods in sport injury research from the 1st methods matter meeting, Copenhagen, 2019. Br J Sports Med. 2020;54:941.
8. Altman N, Krzywinski M. Predicting with confidence and tolerance. Nat Methods. 2018;15:843–5.
9. Matabuena M, Félix P, García-Meixide C, et al. Kernel machine learning methods to handle missing responses with complex predictors. Application in modelling five-year glucose changes using distributional representations. Comput Methods Programs Biomed. 2022;221:106905.
10. Matabuena M, Vidal JC, Hayes PR, et al. A 6-minute sub-maximal run test to predict VO$_2$ max. J Sports Sci. 2018;36:2531–6.
11. Vovk V, Gammerman A, Shafer G. Algorithmic learning in a random world. Springer; 2005.
12. Vovk V, Nouretdinov I, Gammerman A. On-line predictive linear regression. Ann Stat. 2009;37:1566–90.
13. Lei J, G'Sell M, Rinaldo A, et al. Distribution-free predictive inference for regression. J Am Stat Assoc. 2018;113:1094–111.
14. Hastie TJ, Tibshirani RJ. Generalized additive models. Chapman & Hall/CRC; 1990.
15. van der Bles AM, van Der Linden S, Freeman ALJ, et al. Communicating uncertainty about facts, numbers and science. R Soc Open Sci. 2019;6:181870.
16. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? J Mach Learn Res. 2014;15:3133–81.
17. Collazos JA, Dias R, Medeiros MC. Modeling the evolution of deaths from infectious diseases with functional data models: the case of COVID-19 in Brazil. Stat Med. 2023;42:993–1012.
18. Quick C, Dey R, Lin X. Regression models for understanding COVID-19 epidemic dynamics with incomplete data. J Am Stat Assoc. 2021;116:1561–77.
19. Matabuena M, Rodriguez-Mier P, Garcia-Meixide C, Leboran V. COVID-19: estimation of the transmission dynamics in Spain using a stochastic simulator and black-box optimization techniques. Comput Methods Programs Biomed. 2021;211:106399.

Ziad Akram Ali Hammouri [a], Pablo Rodríguez Mier [b], Paulo Félix [a], Mohammad Ali Mansournia [c,d], Fernando Huelin [e], Martí Casals [f,g,h], Marcos Matabuena [a,*]

[a] Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Spain
[b] Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Institute for Computational Biomedicine, Bioquant, Heidelberg, Germany
[c] Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
[d] Sports Medicine Research Centre, Neuroscience Institute, Tehran University of Medical Sciences, Tehran, Iran
[e] Centro de Alto Rendimiento de Pontevedra, Spain
[f] Sport and Physical Activity Studies Centre (CEEAF), Faculty of Medicine, University of Vic-Central University of Catalonia (UVic-UCC), Spain
[g] Sport Performance Analysis Research Group, University of Vic-Central University of Catalonia (UVic-UCC), Barcelona, Spain
[h] National Institute of Physical Education of Catalonia (INEFC), University of Barcelona, Barcelona, Spain

*Corresponding author.
E-mail address: marcos.matabuena@usc.es (M. Matabuena).