DOI:
10.1093/cvr/cvac083

Document Version
Peer reviewed version

This is the Author's Accepted Manuscript version of the article: Lamata, P. (2022). Refining and simplifying decision models – tackling the "one size fits all" challenge. Accepted for publication on 6 May 2022

Refining and simplifying decision models – tackling the "one size fits all" challenge

Pablo Lamata

Dept of Biomedical Engineering, School of Biomedical Engineering and Imaging Sciences, King's College London

Risk prediction models are one pillar to inform decisions in the management of cardiovascular conditions. We need to predict the future, e.g. the occurrence of major adverse cardiovascular events (MACE), and we address this need by learning from past experience. One would think that the wider the experience, and the better the detail (i.e. number of variables) captured of each previous case, the better the predictions we can infer. But it is costly to build up that experience, and we face the risk of human errors during data collection and input. The quest is then to build risk prediction models with the maximal predictive ability and robustness at the minimal cost.

This is the problem addressed by Rios et al[1], specifically looking at the prognostic accuracy for MACE in patients undergoing single-photon emission computed tomography (SPECT). They conceptualise the problem as the search of the optimal combination between 58 automatic variables extracted from a SPECT image and 40 manually input variables from the electronic health record. They provide solid evidence that these 98 variables can be reduced to 23 without a loss of prognostic accuracy in a cross-validation experimental design with 20,414 subjects. It can be thus interpreted that the unnecessary redundancy in these variables has been identified and removed, and that the resulting risk prediction model is one step closer for an efficient clinical practice, as authors conclude.

There are nevertheless still important steps ahead beyond this initial piece of solid evidence. The first one is further testing the generality of this finding, where authors provide initial results with an external cohort of 2984 subjects: the 23 variable minimal model does experience a small drop of prognostic accuracy (from 0.739 of the 98 variable model to 0.723 in AUC - Area Under the Curve of the Receiver Operator Curve). Additionally, the most redundancy lays in the set of 47 image variables removed, since adding them only slightly improves performance (0.725 in AUC from 0.723). These findings highlight that variables initially found to be redundant can still add robustness and better generality of prediction performance, especially if those additional variables do not originate from the same experimental test (i.e. the same SPECT image).

An important methodological consideration is the potential gain of advanced machine learning (ML) choices instead of conventional options when building the risk prediction model. One would argue that the relationship between variables selected are complex and require functions that are beyond traditional linear or sigmoid choices. But a more complex model also has a larger risk of overfitting, and therefore of a poorer generalization. This study provides us with fresh evidence of this compromise: a ML choice clearly outperforms during training in a large cohort (AUC of 0.833 vs. 0.765 for ML-minimum and logistic^{+} respectively, Figure S8 in Rios et al[1]), but it does experience a

much larger drop of performance in the external cohort (drop of 0.110 vs. 0.047 for ML-minimum and logistic$^+$ respectively, comparison of Figure S8 to Figure 5 in Rios et al[1]) eventually leading to a similar performance (0.723 vs 0.718 for ML-minimum and logistic$^+$ respectively). The more complex the model is, the more important the need of an external cohort to back up any claim made from results[2].

These efforts and considerations are part of the "one size fits all" challenge that is intrinsic in current risk models. Studies with large sample sizes are required to generate the evidence that inform these models. The limit of statistical inference, of the inductive reasoning, is defined by these large studies: was the cohort representative of the next subject that needs to have their risk assessed? In fact, Rios et al[1] correctly argue that the small differences between the training and external cohorts may be one of the reasons that explains the drop in prognostic accuracy (from 0.789 to 0.739 for the complete 98 variable model, in Fig.S7 and Fig.5 respectively in Rios et al[1]). The key here is that the smaller the number of variables to include, that is the main objective in this work [1], the easier will it be to tailor the model to more representative sub-groups.

There is an ethical dimension of this limitation, stemming from the implicit biases towards the characteristics of the cohort used to build the model. Bad news is that the more complex the model is, the larger the chances of pick up on these biases, for example in simple tasks such as image segmentation[3]. Good news is that there are methodological solutions to alleviate the impact of these biases, by ensuring balance during training, by adding an AI classifier that helps the model to capture biases, or by eventually training different models for different subgroups[3].

The vision for the future is that of the precision cardiology, where decisions are tailored to the characteristics of individual patients. Getting better data, better biomarkers, is indeed one important direction, and equally important is the optimisation of the decision models discussed here. The final point to argue is that the most efficient way to get these optimal models is the combination of our inductive (predict from the past experience) and deductive (predict from mechanistic understanding) inference. The complementary strategy to make more robust inferences (i.e. predictions) is to use mechanistic models, not only statistical ones, to (i) find the coherent representation of the data we have of the patient, reducing the impact of errors; to (ii) predict variables that are not measurable and to (iii) predict the evolution of the system through forward simulations. This is the vision of the digital twin in cardiovascular medicine [4], and one of the recent examples of the synergies between statistical and mechanistic models is the prediction of the risk of conduction abnormalities induced by transcatheter aortic valve implantation [5].

References

1. Rios R, Miller RJH, Hu L-H, Otaki Y, Singh A, Diniz M, Sharir T, Einstein AJ, Fish MB, Ruddy TD, Kaufmann PA, Sinusas AJ, Miller EJ, Bateman TM, Dorbala S, Carli M Di, Kriekinge S Van, Kavanagh P, Parekh T, Liang JX, Dey D, Berman DS, Slomka P. Determining a minimum set of variables for machine learning cardiovascular event prediction: results from REFINE SPECT registry. *Cardiovasc Res* England; 2021;

2. Lamata P. Avoiding big data pitfalls. *Hear Metab* 2020;33.

3. Puyol-Antón E, Ruijsink B, Piechnik SK, Neubauer S, Petersen SE, Razavi R, King AP. Fairness in Cardiac MR Image Analysis: An Investigation of Bias Due to Data Imbalance in Deep Learning Based Segmentation BT - Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. In: Bruijne M de, Cattin PC, Cotin S, Padoy N, Speidel S, Zheng Y, Essert C, eds.

Cham: Springer International Publishing; 2021. p. 413–423.

4. Corral-Acero J, Margara F, Marciniak M, Rodero C, Loncaric F, Feng Y, Gilbert A, Fernandes JF, Bukhari HA, Wajdan A, Martinez MV, Santos MS, Shamohammdi M, Luo H, Westphal P, Leeson P, DiAchille P, Gurev V, Mayr M, Geris L, Pathmanathan P, Morrison T, Cornelussen R, Prinzen F, Delhaas T, Doltra A, Sitges M, Vigmond EJ, Zacur E, Grau V, et al. The 'Digital Twin' to enable the vision of precision cardiology. *Eur Heart J* 2020;

5. Galli V, Loncaric F, Rocatello G, Astudillo P, Sanchis L, Regueiro A, Backer O De, Swaans M, Bosmans J, Ribeiro JM, Lamata P, Sitges M, Jaegere P de, Mortier P. Towards patient-specific prediction of conduction abnormalities induced by transcatheter aortic valve implantation: a combined mechanistic modelling and machine learning approach. *Eur Hear J - Digit Heal* 2021;**2**:606–615.