we hope that our software exposé lowers the threshold to do so and convinces you that network analysis is worth a try. Finally, should you attend the International Biometric Conference 2022 in Riga, we will teach a pre-conference course that explores the possibilities of these packages more in-depth.

References

Babion, I., Miok, V., Jaspers, A., Huseinovic, A., Steenbergen, R. D., van Wieringen, W. N., and Wilting, S. M. (2020). Identification of deregulated pathways, key regulators, and novel miRNA-mRNA interactions in HPV-mediated transformation. *Cancers*, 12(3), 700.

de Leeuw, F. A., Peeters, C. F. W., Kester, M. I., Harms, et al. (2017). Blood-based metabolic signatures in Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 8, 196–207.

All code used in this article can be found here.

STRengthening Analytical Thinking for Observational Studies (STRATOS): Progress in the Topic Group on Evaluating Diagnostic Tests and Prediction Models (TG6)

Ewout W Steyerberg[1], Ben Van Calster[1 2], on behalf of STRATOS TG6

1 Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, the Netherlands

2 KU Leuven, Department of Development and Regeneration, Leuven, Belgium

In this note, we provide an update of the exciting activities of TG6 of the STRATOS initiative, which is concerned with Evaluating Diagnostic Tests and Prediction Models. TG6 has links with all other topic groups in the STRATOS initiative, and specifically with TG2 (Selection of Variables and Functional Forms in Multivariable Analysis) and TG8 (Survival Analysis).

Members of TG6 met in 2018 at the STRATOS workshop in Leiden (the Netherlands, https://www.lorentzcenter.nl/the-future-of-statistical-modeling-in-medical-data.html), and another great Banff meeting in 2019 (http://www.birs.ca/events/2019/5-day-workshops/19w5198).

Both workshops had two, closely interrelated, overarching objectives to (i) further boost and consolidate the research activities of the STRATOS Topic Groups, and to (ii) identify and initiate new interdisciplinary collaborations between experts in different areas of statistical methodology, regrouped in different TGs. The members of TG6 agreed on preparing an overview of sensible approaches to the evaluation of performance in survival models in collaboration with TG8.

In medical research, many data analyses are conducted by analysts with varying levels of statistical education, experience and interests. In the paper introducing the STRATOS initiative[1], statistical knowledge is categorized into three levels: (1) Low statistical knowledge (2) Experienced statistician (3) Expert in a specific area. We submitted 2 papers that are in category 1, providing guidance for applied researchers on acceptable methods that are easily implemented and highlighting weaknesses of common approaches.

The first paper was led by David McLernon (Aberdeen) on the evaluation of performance in standard survival models[2]. We all recognize that risk prediction models need thorough validation to assess their performance. Validation of models for survival outcomes poses challenges due to the censoring of observations and the varying time horizon at which predictions can be made. In this review paper, we give a description of measures to evaluate predictions and the potential improvement in decision making from survival models based on Cox proportional hazards regression. As a motivating case study, we consider the prediction of the combined outcome of recurrence and death (the 'event') in breast cancer patients following surgery. We develop a Cox regression model with three predictors as in the Nottingham Prognostic Index in 2982 women (1275 events within 5 years of follow-up) and externally validate this model in 686 women (285

events within 5 years). The improvement in performance was assessed following the addition of circulating progesterone as a prognostic biomarker. Model predictions were evaluated across the full range of observed follow up times or for the event occurring by a fixed time horizon of interest. We discuss recommended statistical measures that evaluate model performance in terms of discrimination, calibration, or overall performance. Furthermore, we evaluate the potential clinical utility of the model to support clinical decision making. SAS and R code is provided to illustrate apparent, internal, and external validation, both for the simple 3-predictor model and when adding progesterone.

The second paper was led by Nan van Geloven (Leiden) and will be published in the British Medical Journal[3]. In this paper, we consider competing risks and the evaluation of predictive performance. We recognize that for time-to-event outcomes such as breast cancer recurrence, death from other causes is a competing risk. Model performance measures must account for such competing events. We present a comprehensive yet accessible overview of performance measures for this competing event setting, including the calculation and interpretation of statistical measures for calibration, discrimination, overall prediction error, and clinical utility by decision curve analysis. All methods for model development and validation are illustrated for patients with breast cancer, with publicly available data and R code. Both papers have had a lengthy trajectory of exchanging different perspectives by e-mail and videoconferencing. A particular challenge was the issue on how to deal with censored observations in a validation context. Also, we had to remain within acceptable word count limits, which was solved by providing extensive supplementary material, including details on the data and R, and SAS code.

Future work will focus on the evaluation of performance in dynamic survival models (with Hein Putter) and models which use age as the time axis for prediction (with Terry Therneau).

1. Willi Sauerbrei, Michal Abrahamowicz, Douglas G. Altman, Saskia le Cessie, James Carpenter,on behalf of the STRATOS initiative. STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative. Stat Med 2014 Dec 30;33:5413-32

2. David McLernon, Daniele Giardiello, Ben Van Calster, Laure Wynants, Nan van Geloven, Maarten van Smeden, Terry Therneau, Ewout Steyerberg; on behalf of STRATOS TG6 and TG8. Assessing performance in prediction models with survival outcomes: practical guidance (submitted)

3. Nan van Geloven, Daniele Giardiello, Edouard Bonneville, Lucy Teece, Chava Ramspek, Maarten van Smeden, Kym Snell, Ben van Calster, Maja Pohar-Perme, Richard Riley, Hein Putter, Ewout Steyerberg on behalf of STRATOS TG6 and TG8. Validation of prediction models in presence of competing risks: a guide through modern methods (in press, BMJ 2022)