Thursday June 9, 14:00-15:15

Parallel session 1

Methods: Prediction & IPD meta-analyses

Chairs: Femke Atsma & Sander Groen in 't Woud

14:00 Low inter-rater reliability of the Prediction model Risk Of Bias ASsessment Tool (PROBAST) was difficult to improve (O1)
Junfeng Wang

14:15 Prediction of cancer-related fatigue using multiple machine learning models (O2)
Lian Beenhakker

14:30 Quantitative prediction error analysis to investigate predictive performance under predictor measurement heterogeneity at model implementation (O3)
Kim Luijken

14:45 Imputation of incomplete variables in an individual participant data meta-analysis: Relaxing the need for untestable assumptions (O4)
Johanna Munoz

15:00 Additive interaction in two-stage individual participant data meta-analysis: a proposed method and example from the PSY-CA consortium (O5)
Maartje Basten

O1. Low inter-rater reliability of the Prediction model Risk Of Bias ASsessment Tool (PROBAST) was difficult to improve.

Wang J., Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands
Ma J., Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

Background: The Prediction model Risk Of Bias ASsessment Tool (PROBAST) has been developed for assessing the methodological quality of individual prediction model studies included in a systematic review. We aim to assess the inter-rater reliability (IRR) and accuracy of using PROBAST in researchers with less experience, and whether training can improve their performance.

Methods: Twenty-four models in three model types were randomly selected from a systematic review and critical appraisal of COVID-19 prediction models. Eight reviewers were randomly assigned to two groups. An orthogonal design was used to assign publications to reviewers, to avoid the potential impact of publications and reviewers. Each publication was evaluated independently by a pair of reviewers before the training and by another pair after the training.
Cohen's quadratic weighted kappa coefficient (κ) was calculated for each publication as the measure of IRR. The inter-rater agreement between two reviewers and the accuracy of the evaluation were calculated for each signalling question (SQ) in PROBAST.

Results: The overall weighted kappa for all domains and all papers was 0.0259 (95% CI: -0.0595, 0.111) before training and 0.182 (0.0772, 0.287) after training. Kappa for each publication was improved in 15 out of 24 publications. Before training, inter-rater agreement for individual SQs ranged from 29.2% to 87.5% (median=62.5%) and accuracy ranged from 33.3% to 79.2% (median=55.2%). After training, the median of inter-rater agreement increased to 70.8%, however the accuracy decreased slightly to a median of 54.2%.

Conclusion: The relatively low inter-rater reliability and poor accuracy may jeopardize the validity of the assessment with PROBAST. For reviewers with less experience, training has little effect on improving the IRR but still not to the level desired.

O2. Prediction of cancer-related fatigue using multiple machine learning models.

Beenhakker L., Department of Biomedical Signals and Systems, University of Twente, Enschede, The Netherlands
Wijlens K.A.E., Department of Biomedical Signals and Systems, University of Twente
Witteveen A., Department of Biomedical Signals and Systems, University of Twente
Heins M., Department of Primary Care, Netherlands Institute for Health Services Research (NIVEL)
Bode C., Department of Psychology, Health and Technology, University of Twente
Siesling S., Department of Health Technology and Services Research, University of Twente, Department of Research and Development, Netherlands Comprehensive Cancer Organisation (IKNL)
Vollenbroek-Hutten M.M.R., Department of Biomedical Signals and Systems, University of Twente

Background: Early diagnosis and improved treatment has increased the number of breast cancer survivors. This increase results in more people struggling with long-term effects of cancer and its treatment. One of these effects is cancer-related fatigue (CRF). It is important to recognize CRF in time to prevent it from worsening and becoming chronic by starting a CRF intervention. Using machine learning models, we aimed to predict the individual risk of developing CRF.

Methods: Data from the Primary Secondary Cancer Care Registry (PSCCR) was used, in which information of the Netherlands Cancer Registry (NCR) was combined with data of General Practitioners (GPs) via Nivel Primary Care. We included 12.813 breast cancer patients of which 2.224 visited the GP with fatigue complaints. Predictors (n=64) were related to patient, tumour and treatment characteristics and GP visits before diagnosis. Missing data was imputed using Multiple Imputation by Chained Equations and risk was predicted using Random Forest Classifier, Logistic Regression, Gaussian Naïve Bayes, K-Nearest Neighbours and Multi-Layer Perceptron. A nested 5-fold cross validation was used to optimize hyperparameters and assess the performance of the models by comparing the area under the receiver operator characteristic curve (AUC-score).

Results: The performance of the models was poor to moderate, with AUC-scores ranging from 0.54-0.63. The Random Forest Classifier and the Logistic Regression model performed best, with AUC-scores of 0.63±0.014 and 0.62±0.09 respectively.

Conclusion: Using machine learning models on the PSCCR dataset, the individual risk for CRF cannot be predicted accurately. This can be due to the assessment of fatigue as outcome measure. Not all patients with fatigue complaints visit the GP and not all fatigue complaints might be related to the breast cancer diagnosis. In future studies, we hope to collect more detailed data and have a clearer differentiation between fatigued and non-fatigued patients.

O3. Quantitative prediction error analysis to investigate predictive performance under predictor measurement heterogeneity at model implementation.

Luijken K., Leiden University Medical Center, Leiden and University Medical Center Utrecht, Utrecht, the Netherlands
Song J., Leiden University Medical Center, Leiden, the Netherlands
Groenwold R.H.H., Leiden University Medical Center, Leiden, the Netherlands

Background: When a predictor variable is measured in similar ways at the derivation and validation setting of a prognostic prediction model, yet both differ from the intended use of the model in practice (i.e., 'predictor measurement heterogeneity'), performance of the model at implementation needs to be inferred. This study proposed an analysis to quantify the impact of anticipated predictor measurement heterogeneity.

Methods: A simulation study was conducted to assess the impact of predictor measurement heterogeneity across validation and implementation setting in time-to-event outcome data. The use of the quantitative prediction error analysis was illustrated using an example of predicting the 6-year risk of developing type-2 diabetes with heterogeneity in measurement of the predictor body mass index.

Results: In the simulation study, calibration-in-the-large of prediction models was poor and overall accuracy was reduced in all scenarios of predictor measurement heterogeneity. Model discrimination decreased with increasing random predictor measurement heterogeneity.

Conclusions: Heterogeneity of predictor measurements across settings of validation and implementation reduced predictive performance at implementation of prognostic models with a time-to-event outcome. When validating a prognostic model, the targeted clinical setting needs to be considered and analyses can be conducted to quantify the impact of anticipated predictor measurement heterogeneity on model performance at implementation.

O4. "Imputation of incomplete variables in an individual participant data meta-analysis" Relaxing the need for untestable assumptions.

Munoz J., Utrecht Medical Center, Utrecht, The Netherlands
De Jong V.M.T, Utrecht Medical Center, Utrecht, the Netherlands
Debray T., Utrecht Medical Center, Utrecht, the Netherlands

Background: Missing data is a common problem in medical research, and is commonly addressed using multiple imputation. Although traditional imputation methods allow for valid statistical inference when data are missing at random (MAR), their application is not justified when observations are clustered (e.g., within studies) or when the presence of missingness depends on unobserved information. This situation often occurs when individual participant data (IPD) from multiple studies are combined. Although several imputation methods have been proposed to address individual studies in which data are missing non-randomly (MNAR), their applicability and validity in large data sets with clustering remain unclear. Therefore, we propose a new imputation method for multilevel MNAR data. This method is based on the principles of Heckman selection models, and adopts a two-stage meta-analysis approach for the imputation of binary and continuous variables.

Methods: We simulated data from ten cohorts with a missing process compatible with the Heckman model. We considered different simulated scenarios by varying the MNAR specification, cohort sample size, number of cohorts, and error distribution. We evaluate the performance of our method and other previously proposed imputation methods in each scenario and finally illustrate the method in a multi-district study to estimate the prevalence of malaria in Ugandan children.

Results: Simulation analysis shows that our method provides comparable coefficient parameter estimates in terms of bias with a Heckman imputation method without clustering. Our method provides better coverage among the evaluated methods. However, it is sensitive to sample size and number of clusters, as well as to deviations from the normality distribution.

Conclusion: Under a correct specification of the imputation model and of the exclusion variables, the proposed method is reliable to impute binary and continuous variables following an M(N)AR mechanism according to the Heckman model, coming from multilevel data or IPD studies.

O5. Additive interaction in two-stage individual participant data meta-analysis: a proposed method and example from the PSY-CA consortium.

Basten M., Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht and Utrecht University, Utrecht, the Netherlands
Pan K-Y. Amsterdam UMC, Amsterdam, the Netherlands
van Tuijl L.A., University Medical Center Groningen, Groningen, the Netherlands
Knol M.J., National Institute for Public Health and the Environment
Ranchor A.V., University Medical Center Groningen, Groningen, the Netherlands
Voogd A.C., Maastricht University, Maastricht, the Netherlands
de Graeff A., University Medical Center Utrecht, Utrecht, the Netherlands
Hoogendoorn A.W., Amsterdam UMC, Amsterdam, the Netherlands
Vermeulen R., Utrecht University, Utrecht, the Netherlands
Portengen L., Utrecht University, Utrecht, the Netherlands
Lamers F., Amsterdam UMC, Amsterdam, the Netherlands
Dekker J., Amsterdam UMC on behalf of the PSY-CA consortium, Amsterdam, the Netherlands
Geerlings M.I., University Medical Center Utrecht, Utrecht, the Netherlands

Background: Additive interaction better reflects biological interaction and is more relevant for public health than multiplicative interaction. When studies have little power to study interaction, they can be combined using individual participant data (IPD) meta-analysis. Guidelines are available to study multiplicative interaction in IPD meta-analysis, but little is known on how to implement additive interaction measures, including relative excess risk due to interaction (RERI). We examined two procedures to study additive interaction in two-stage IPD meta-analysis within the Psychosocial factors and cancer (PSY-CA) consortium.

Methods: We used IPD from 14 cohorts to study interactions between several combinations of continuous measures of psychosocial factors (depression, anxiety) and health behaviors (smoking, alcohol use, physical activity, BMI) in relation to different cancer outcomes. Within each cohort, effects of the psychosocial factor, health behavior and their product term on cancer incidence were estimated using Cox regression (stage one). Procedure A: RERI estimates were calculated for each cohort (stage one) and entered into random-effects univariate meta-analysis (stage two). Procedure B: cohort-level effect estimates of the exposures and product term were entered into random-effects multivariate meta-analysis (stage two). These pooled estimates were used to calculate one overall RERI.

Results: In procedure A we faced two problems due to RERI's being calculated from relative effect estimates: 1) in several cohorts RERI estimates were invalid because one of the exposures was negatively associated with the outcome (preventive effect); 2) positive and negative RERI estimates could not be validly pooled together at stage two. These issues were overcome in procedure B.

Conclusion: To study additive interaction in two-stage IPD meta-analysis, we recommend to first meta-analyze cohort-level effect estimates of the exposures and their product term and use pooled estimates to calculate one overall RERI.