Thursday 25 July 2024

09:00-12:30 Mini Symposium 2 (Room 2)

STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative – recent progress and foci for the future

Organizers: Willi Sauerbrei and Els Goetghebeur in collaboration with the STRATOS Steering Group

Adjusting for covariate measurement error on functional form estimation: design and early results from a blinded, collaborative STRATOS project

Aris Perperoglou, Paul Gustafson, Michal Abrahamowicz, Victor Kipnis, Mohammed Sedki, Anne Thiébaut, Lawrence Freedman for TG2 - TG4

Introduction: The evaluation and estimation of relationships between outcome variables and covariates measured with error remains a challenge in observational studies. Challenges may be amplified when the true functional form between the covariate and the outcome is suspected to be non-linear. This work outlines the collaboration between two topic groups of the STRATOS initiative: TG2, specializing in the selection of variables and functional forms in multivariable analysis, and TG4, which focuses on addressing issues related to measurement error and misclassification. The project investigates the performance of methods for estimating complex functional relationships in observational data, where covariates are prone to measurement inaccuracies, specifically targeting the accurate estimation of non-linear relationships between outcome variables and covariates.

Project Design: The project adopts a blinded, multi-stage design to rigorously compare methodological approaches. A Data Generation and Evaluation team produces datasets simulating various functional relationships and measurement error scenarios, but withholds the true underlying model from the Methods teams. Three distinct Methods teams implement different analytic approaches, based on Bayesian methods, Imputation/Regression Calibration methods and SIMEX methods, respectively. For each method, estimation of the functional form based on (i) B-splines, (ii) P-splines and (iii) Fractional Polynomials are investigated, with pre-specified hyperparameters for each approach. The initial phase of the project involved generating 5 datasets, each comprising realizations of a binary outcome Y and a continuous covariate measured with independent error, alongside pairs of repeat covariate observations for validation purposes in a random subset. The relationship between Y and X was specified as a logistic regression, but the error structure and functional relationship between logit(P(Y=1) and X remained undisclosed to prevent bias in analysis and thus enhance the integrity of the study. Each methods team created their code and returned estimated functional relationships without knowledge of the generating model or the results from any other team's methods.

Early Results and Implications: The Evaluation Team assed the performance of the methods through a comparison of predicted values against undisclosed true values, using mean squared

error and other relevant metrics to gauge performance. Initial findings reveal performance disparities across methods (blinded for the study's integrity). In Phase 2 of the project a further 75 datasets have been simulated, varying sample size, functional form and size of measurement error in a systematic manner. These datasets are now being analyzed. The results will offer insights into how factors like sample size, validation study size and spline and covariate functions influence method accuracy This analysis will be important as it will guide the selection and refinement of statistical methods for the final phase of the project that will include several hundred more simulated datasets. This blinded, collaborative structure fosters an unbiased and efficient evaluation of statistical techniques. Results will contribute to the STRATOS Initiative's broader goal of providing guidance for analyzing observational data. Notably, this project design showcases a model for collaborative, transparent, and rigorous statistical research to address challenges in real-world settings.