Found 4 Records

CONTROL ID: 3408799
PRESENTER: Aris Perperoglou
PRESENTER (INSTITUTION ONLY): Newcastle University
TITLE: Outstanding issues in selection of variables and functional forms in multivariable analysis
ABSTRACT BODY:
Abstract Body: Background
When creating a descriptive multivariable model, variable selection and identification of functional forms for continuous variables are key concerns. Ad hoc ‘traditional’ approaches to variable selection have been developed at least 50 years ago and are still in use. Methods for determining functional forms for continuous variables were also first suggested many years ago but are still underused. More recently, alternative approaches to address these challenges have been proposed, but knowledge of their properties is scarce, probably because of a lack of meaningful comparisons between them. Therefore, there are many outstanding issues in multivariable modelling that prevent us to define a state of the art and to provide evidence-supported guidance.
Objective
Our main aims are to identify and illustrate such gaps in the literature.
Methods
We briefly discuss general issues in building descriptive regression models, strategies for variable selection, ways of choosing and modelling functional forms for continuous variables and methods for combining the selection of variables and functions. We review findings from Perperoglou et al (2019) and discuss issues when building descriptive regression models with and present open issues around the use and of splines in a multivariable setting, with a focus on interpretability, presentation of results and variable selection.
Results
Our overview revealed that there is not yet enough evidence on which to base recommendations for the selection of variables and functional forms in multivariable analysis. Such evidence may come from comparisons between alternative methods. In particular, we highlight seven important topics that require further investigation.
Conclusions
Selection of variables and of functional forms are important topics in multivariable analysis. To define a state of the art and to provide evidence-supported guidance further comparative research is required. Simulation studies and comparative analyses of real data sets will play a key role.
References
Perperoglou A, Sauerbrei W, Abrahamowicz M, Schmid M on behalf of TG2 of the STRATOS initiative. A review of spline function procedures in R. BMC Medical Research Methodology. 2019; 19:46.
Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M et al: State-of-the-art in selection of variables and functional forms in multivariable analysis--outstanding issues. Diagnostic and Prognostic Research (2020) 4:3. https://doi.org/10.1186/s41512-020-00074-3
AUTHORS/INSTITUTIONS: A. Perperoglou, Mathematics, Statistics and Physics, Newcastle University, Newcastle, UNITED KINGDOM|M. Schmid, Institut für Medizinische Biometrie, Informatik und Epidemiologie, Universität Bonn, Bonn, GERMANY|D. Dunkler, Center for Medical Statistics, Informatics and Intelligent Systems, Medizinische Universität Wien), Vienna, AUSTRIA|

CONTROL ID: 3467767

PRESENTER: Els Goetghebeur

PRESENTER (INSTITUTION ONLY): Ghent University, Karolinska Institutet

TITLE: Causal lessons from our Simulation Learner: SUTVA is fiction; randomisation fails as an instrument for many post randomisation exposures, and averaging causal effects over an observed (experimental) instrument may be irrelevant

ABSTRACT BODY:

Abstract Body: To support Causal reasoning from DAGS and a choice among available estimators, the STRATOS causal inference group developed a `simulation learner'. This engine generates per subject alongside observed exposure(s) and outcome a range of alternative exposures with their potential outcome. As in the Promotion of Breastfeeding Intervention Trial, we `randomize mother-infant pairs to standard of care or a breastfeeding encouragement (BFE) intervention. Main outcome is weight at 3 months. The path from randomization to outcome meets the intervention uptake (education program), followed by the start and a specific duration of breastfeeding. Simulated parallel worlds then enable visualization of various potential outcomes and causal estimands in specific populations. The necessary intermediate steps highlight that SUTVA must be context specific. We see randomisation act as an instrument for one exposure (e.g receiving an offer for the BFE programme or actually following the BFE programme), but not others (e.g. actually starting breastfeeding). We recognize that averaging causal effects over an observed (experimental) instrument may be irrelevant unless one conditions on the instrumental variable. We thus explore distinct estimation methods and compare results with the simulated population parameters. More recently it helped us to motivate an estimation approach when emulating a clinical trial for the estimation of treatment effect in patients hospitalized with COVID-19. R code is available on www.ofcaus.org, where SAS and Stata code for analysis is also provided. This is work on behalf of STRATOS TG 7 Causal Inference

AUTHORS/INSTITUTIONS: E. Goetghebeur, Applied Mathematics, Computer Science and Statistics, Ghent University, Gent, BELGIUM|E. Goetghebeur, Medical Epidemiology and biostatistics, Karolinska Institutet, Stockholm, SWEDEN|S. le Cessie, Leids Universitair Medisch Centrum, Leiden, NETHERLANDS|

CONTROL ID: 3467887

PRESENTER: Michal Abrahamowicz

PRESENTER (INSTITUTION ONLY): McGill University

TITLE: Designing accurate, generalizable and informative simulation studies: recent developments

ABSTRACT BODY:

Abstract Body: Proliferation of new, ever more sophisticated biostatistics methods permits addressing new analytical challenges of real-life research. However, for end-users, it also creates two important problems. First, the methods complexity implies that their consistency, accuracy and/or efficiency are difficult to be theoretically established, or can be proven only under very restrictive assumptions. Thus, the end-users are often uncertain if the proposed method will yield valid results in their specific real-life application. Secondly, data analysts need a solid evidence to decide which among the alternative methods addressing a particular methodological challenge should be applied in their application. To further complicate the matters, the choice of the preferable method may depend on the structure, quantity and/or quality of the data being analyzed. Both these generic issues can be addressed by well designed and carefully executed and interpreted simulation studies.

Whereas most statistical papers that propose new methods do include some simulations, their design and scope often raise concerns. Simulations, performed by the developers of the new method, often aim at providing evidence of its satisfactory performance and 'superiority' over the existing alternatives, under a limited range of 'favorable' assumptions. One of the over-arching goal of the STRATOS Simulation Panel is to develop objective, verifiable criteria for designing 'neutral', comprehensive, informative, and reproducible simulations [1]. Several 'generic' issues related to design, conduct and reporting of simulations comparing alternative statistical methods are addressed by Morris et al [2].

I will first summarize the recommendations from the recent paper by the Simulation Panel members[1]. Then, a brief discussion of some more advanced issues will focus on (i) making simulations plausible and practically relevant; (ii) need to assess the bias/variance trade-off, (iii) going beyond the 'average performance' and assessing the results of individual replications, (iv) watching for a 'tip of the iceberg'. Each issue will be illustrated using results of recent simulations and, together, they will demonstrate the advantages of 'learning from/through simulations' in modern methodological research.

[1] Boulesteix, Binder, Abrahamowicz, Sauerbrei, 2018. Biometrical Journal 60: 216-218.

[2] Morris, White, Crowther, 2019. Statistics in Medicine 38:2074-2102.

AUTHORS/INSTITUTIONS: M. Abrahamowicz, McGill University, Montreal, Quebec, CANADA|

CONTROL ID: 3469692
PRESENTER: James R Carpenter
PRESENTER (INSTITUTION ONLY): London School of Hygiene & Tropical Medicine
TITLE: Missing data: best practice and beyond in flexible modelling and causal inference
ABSTRACT BODY:
Abstract Body: Missing data are ubiquitous in medical and social research; however a number of tools (not least multiple imputation) are now well
established for many common settings[1].

In this talk, we review recent developments for missing data in models with splines and interactions and non-linear functions, and then present recent work on missing data in Marginal Structural Models (MSMs) [2].

These are commonly used to estimate causal intervention effects in longitudinal non-randomised studies. However, a common challenge when using MSMs to analyse observational studies is incomplete confounder data, where a poorly informed analysis method will lead to biased intervention effect estimates. Despite a number of approaches described in the literature to handle missing data in MSMs, there is little guidance on what works in practice and why.

We discuss the assumptions that commonly used methods (complete cases, last observation carried forward, missingness pattern, inverse probability weighting and multiple imputation) make, and compare them using a carefully constructed simulation study.

Our findings highlight that careful consideration of the scientific setting is needed to choose the appropriate approach.

References:

[1] Carpenter, J. R. (2020) Missing data: a framework for practice. Biometric Journal (revision under review)

[2] Leyrat, C., Carpenter, J. R., Bailly, S. and Williamson E. J. (2020) Common methods for missing data in marginal structural models: what works and why. American Journal of Epidemiology, in press.

AUTHORS/INSTITUTIONS: J.R. Carpenter, C. Leyrat, E. Williamson, Medical Statistics, London School of Hygiene & Tropical Medicine, London, UNITED KINGDOM|