

DAGSTAT 2022

SESSION

STRengthening Analytical Thinking for Observational Studies (STRATOS) –
Research, Guidance and Education is needed to enhance data analyses

Organizers and Chairmen: Heiko Becher (Hamburg) and Willi Sauerbrei (Freiburg)

h.becher@uke.de, wfs@imbi.uni-freiburg.de

In 2013 the STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative was created with the aim to systematically evaluate existing methodologies, identify unresolved issues, stimulate research in these areas, develop guidance and education material. The overarching aim is to enhance methodological accuracy of real-life data analyses.

Program:

Willi Sauerbrei (Freiburg) and Heiko Becher (Hamburg)

- STRATOS – aims, tasks, support of the initiative (10 min)

Georg Heinze (Vienna) for TG2 (22+3)

- Selection of variables and functional forms for multivariable models

Veronika Deffner (Munich) for TG4 (22+3)

- Measurement error and misclassification of covariates: Should we worry?

James Carpenter (London) for TG1 (22+3)

- Handling missing data in the analysis: practical guidance for structuring the analysis, choosing the tools, and reporting the results.

General Discussion (15)

Abstracts:

Selection of variables and functional forms for multivariable models

Georg Heinze (Vienna) for STRATOS TG2

georg.heinze@meduniwien.ac.at

In hardly any statistical topic the gap between cutting-edge methodology and daily practice of data analysis is larger than in multivariable model building. For example, according to recent systematic reviews, univariate screening of variables considered for multivariable models is still common practice and sometimes even believed to be a precondition for inclusion. Furthermore, while there are sophisticated and well-investigated methods for considering nonlinear functional forms of continuous variables in such models, dichotomization of such variables is still popular. If continuous variables are not dichotomized, the linearity assumption is hardly ever questioned. We give an overview of methods for selection of variables and functional forms, discuss recommendations for practitioners and point at open gaps in research. Furthermore, we present some activities of our initiative to guide and educate researchers with different levels of statistical education, including our workshops, short videos and interactive shiny apps.

Measurement error and misclassification of covariates: Should we worry?

Veronika Deffner (Federal Office for Radiation Protection, Germany) for STRATOS TG4

vdeffner@bfs.de

In order to quantitatively investigate various phenomena in the real world using statistical models, researchers rely on measurements of the variables that are of interest. But what if these measurements have errors? Should we worry about their impact on the validity of the results? Topic group 4 of the STRATOS initiative developed theoretical and practical guidance to help evaluate the impact of the problem and deal with measurement error and misclassification in statistical modeling. The guidance document and the Shiny application “MEM-Explorer” will be presented in the talk with a focus on multiple linear regression with error-prone covariates.

Covariate measurement error may affect the analysis of the association of the covariate with the outcome variable, with regard to three aspects: the estimate of the regression coefficient, the test of the null hypothesis of no association and the power to detect the association. The magnitude of the impact depends on the magnitude and type of the error, as well as on the characteristics of the regression model and the data. In particular, the role of the type of error is often misjudged. Various methods are available to account for covariate error in a regression analysis, but they all require information about the type and magnitude of the covariate error. “MEM-Explorer” allows interactive exploration of the impact of covariate measurement error: data, based on user-specified settings, are generated and the resulting estimates from the regression models are visualized, with and without the application of methods to adjust these estimates for covariate error.

Handling missing data in the analysis: practical guidance for structuring the analysis, choosing the tools, and reporting the results.

James Carpenter (Department of Medical Statistics, LSHTM and MRC Clinical Trials Unit, UCL) for STRATOS TG1

James.Carpenter@lshtm.ac.uk

Missing data are a pervasive problem in data analysis. Although there is increasing guidance on how to handle missing data, practice is changing slowly and misapprehensions abound, particularly in observational research. Further, the lack of transparency around methodological decisions is threatening the validity and reproducibility of modern research.

Drawing on the recent TARMOS 'Treatment And Reporting of Missing data in Observational Studies' [1] framework from the STRATOS initiative, and using data from the Youth Cohort (Time) Series (YCS) for England, Wales and Scotland, 1984-2002, we first discuss how to structure the analysis.

Then, we consider three common 'tools' for handling missing data[2]: (a) complete-case analysis, where only units that are complete on the variables in an analysis are included; (b) weighting, where the complete cases are weighted by the inverse of an estimate of the probability of being complete; and (c) multiple imputation (MI), where missing values of the variables in the analysis are imputed as draws from their predictive distribution under an implicit or explicit statistical model, the imputation process is repeated to create multiple filled-in data sets, and analysis is carried out using simple MI combining rules. We give a non-technical discussion of the strengths and weakness of these approaches.

Finally, we discuss the arguments for systematic reporting of analysis which make use of these techniques, and sketch how this might work.

References:

[1] Lee, K. J., Tilling, K. M., Cornish, R. P., Little, R. J. A., Bell, M. L., Goetghebeur, E., Hogan, J. W. and Carpenter J. R. on behalf of the STRATOS initiative (2021). Framework for the treatment and reporting of missing data in observational studies: The Treatment and Reporting of Missing data in Observational Studies (TARMOS) framework. *Journal of Clinical Epidemiology*, **134**, 79-88.

[2] Little, R. J. A., Lee, K. J. and Carpenter J. R. on behalf of the STRATOS initiative (2022). A comparison of three popular methods for handling missing data: complete case analysis, weighting and multiple imputation. *Revision submitted to Sociological Methods and Research*.