MS 2: Design and analysis of observational studies in medical research

MS 2.0

Initiative *Guidance for key issues of design and analysis of observational studies*

Sauerbrei W.[1]
[1]*University Medical Center Freiburg, Institute of Medical Biometry and Informatics, Freiburg, Germany*

In 2012 a new international initiative entitled 'Guidance for key issues of design and analysis of observational studies' was started. By the end of May 2013, more than twenty researchers from 8 countries are involved, many more will be necessary for this ambitious initiative. This will be the first presentation at a meeting. Rationale, aims, general strategies and results of the starting phase will be presented.
There will be three general talks about the rationale, aims and intended strategies of the initiative. For the beginning the Steering Group has chosen seven topics of general methodological interest. The current status of our work, and the related challenges, will be presented in the second part of the Mini-Symposium, with some talks more detailed than other.
Beside of presenting the current state of the initiative to a broader audience, we also aim to interest further colleagues to work on guidelines for both the seven topics that will be presented and on some new topics, which can be proposed to the Steering Group.
In the afternoon (starting at 14.15, room tba) there will be a session of the members of the initiative. Interested colleagues are very welcome. For further information about the initiative please contact Willi Sauerbrei (IMBI, University Medical Center Freiburg, Germany, wfs@imbi.uni-freiburg.de).

MS 2.1

Guidance for key issues of design and analysis of observational studies

Sauerbrei W.[1], for the steering group of the initiative
[1]*University Medical Center Freiburg, Institute of Medical Biometry and Informatics, Freiburg, Germany*

In the last decades, statistical methodology has seen substantial developments. The improvement of computer facilities can be viewed as the cornerstone. Unfortunately, many sensible improvements are ignored in practice and often design and analysis of observational studies exhibit severe weaknesses. The overwhelming concern with theoretical aspects in the literature, often with very limited guidance on key issues that are vital in practice, frequently discourages researchers from utilizing more sophisticated and possibly more appropriate methods. It is also obvious that many analyses are conducted by people with limited statistical knowledge. To improve the situation, at least two tasks must be addressed.
First, experts in specific methodological areas have to work towards developing guidance for practically relevant issues. This requires international cooperation. Second, as a result of the expansion in statistical methodology and software, there is an ever-increasing need for continuing education of analysts.
It is important that the knowledge gained through research on statistical methodology is transferred to the broader community of researchers with different levels of statistical knowledge. We classify them as: level 1 - low; level 2- experienced statistician; level 3 - expert in a specific area. Many of them would be most grateful for an overview on the current state of the art and for guidance documents helpful for practice.
For a small number of highly relevant topics assessment of the current state of practise, identification of current documents trying to provide some guidance and 'agreement' what to recommend within each of the three levels are short term aims of the new initiative. The overarching long term aim is the improvement of statistical analyses. Development of guidance documents for most relevant topics and their broad acceptance by analysts would be a cornerstone. This may imply that the percentage of analysts with 'level 1' knowledge only will decrease substantially.
Obviously, guidance documents have to be regularly improved. It is the aim that there is sufficient evidence that some approaches can be moved from level 3 into a level 2 recommendation. This also requires that 'easy-to-use' software is generally available.

MS 2.2

Why we need guidance documents for the design and analysis of observational studies

Altman D.[1]
[1]*Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom*

Many medical research projects do not include statisticians or epidemiologist among the team, and there is wide access to cheap software to enable statistical analyses. Hundreds of reviews of published articles, especially those relating to randomised controlled trials (RCTs), have consistently shown that methodological problems are common and key information is frequently missing from trial reports. Similar evidence is accumulating for observational research. Common errors include inappropriate analysis, failing to account for clustering in data analysis, improperly addressing missing data, and a host of problems associated with hypothesis testing. In addition, reporting of much observational research has been shown to be deficient.
There is a clear need for higher standards of design, analysis, and reporting of observational research. Several reporting guidelines, including STROBE and REMARK, have outlined the essential elements of reporting observational studies of different designs (see equator-network.org). There is a clear need to companion guidelines for research conduct. These would be of particular benefit to those who lack formal training, but who nevertheless conduct a large proportion of the research of this type.

MS 2.3

Guidance initiative: a route map, and the journeying of the missing data topic group (TG1)

Carpenter J.R.[1]
[1]*London School of Hygiene & Tropical Medicine, Department of Medical Statistics, London, United Kingdom*

In this talk I will elaborate on the route the steering group envisages the Guidance Initiative following. Broadly speaking, this entails identifying a list of topic areas (including those in today's programme), and topic chairs. These are then responsible for convening their topic group, then scoping and in due course drafting the guidelines.
The steering group will seek to ensure that the initiative's guidelines take the form of strategies for thinking about the issues, rather than a recipe book approach.
I will use the missing data topic group (TG1) to illustrate, and commend, the journey!

MS 2.4

TG 3: Descriptive and initial data analysis

le Cessie S.[1]
[1]*Leiden University Medical Center, Department of Clinical Epidemiology, Clinical Epidemiology/Medical Statistics and bioinformatics, Leiden, Netherlands*

The initial steps of the data analysis of a study consist of checking and cleaning of the data, examining the collected data, describing the study population and preparing the data for further analyses. In each of these steps several decisions have to be made. In this lecture we provide guidance to these initial steps.
The effort to reach high quality data starts fare before data collection and includes the construction of the questionnaire, right tools for data entry (double entry, plausibility checks etc.) and a well-planed system for checking data for errors and inconsistencies. After this, the next step is to “get to know” the collected data and examine it for any particularities: skewness of continuous variables, limited variation, number and patterns of missing values, distributions of categorical variables. The inclusion and flow of the study should be explored, with an overview of missing measurements and follow-up data, however this step should not influence the statistical analysis plan. We will give some orientation for these steps and consider how the distribution of the different variables best can be examined.
An important aim of these initial steps is to provide a clear description of the study population in tables and figures. This can be done in many different ways: summary statistics can be reported for the total population or for subgroups; continuous variables can be summarized by means and standard deviations, by medians and percentiles or by categorizing them. Categorical variables with many small categories can be heaped together in many different ways. We will discuss the advantages and

limitations of the different approaches. In particular we discuss the problem of reporting when the variables of interest contain missing values.
Finally we consider the preparation of data for more advanced analyses. Here decisions have to be made about the way variables are used in further analyses, and how outliers and missing data are handled. We will discuss ways to handle (partially) missing data and outliers and guide decisions regarding dropping certain subjects or certain variables from future analyses. We consider transformations of skewed variables and discuss the pros and cons of categorizing continuous variables, which is often done in medical research.

MS 2.5

TG 2: Selection of variables and functional forms; flexible approaches improve estimation and inference

Abrahamowicz M.[1]
[1]*Epidemiology & Biostatistics, McGill University, Montreal, Canada*

Analyses of observational studies typically rely on multivariable statistical models. Topic group 2 focuses on two inter-related questions associated with the specification of a multivariable model: (1) selection of 'relevant' variables, and (2) choice of the functional form for the effect of each continuous variable. Because prior knowledge is often limited, both issues usually involve data-dependent decisions, at least for a subset of variables. Yet, in applied research, these important decisions are often made on *ad hoc* basis, with strong tendency to replicate 'conventional' approaches, and without providing solid (*a priori*) theoretical justification or (*a posteriori)* empirical evidence that the chosen approach is 'optimal' or even 'appropriate'. Often models with (too) many variables are proposed and for modeling the effects of continuous variables, most real-life studies either *a priori* assume a linear dose-response function or categorize the continuous variable using arbitrary cut-offs. Yet, both these conventional approaches have been shown to induce serious risks of biased estimates and incorrect inference. On the other hand, several flexible approaches such as fractional polynomials, splines or generalized additive models have been proposed, but are often ignored in practice. The limitations of conventional approaches and the advantages of flexible modeling will be illustrated using both simulations and real-life examples.

MS 2.6

TG 7: Causal inference - at work

Goetghebeur E.[1]
[1]*Universiteit Gent, Ghent, Belgium*

The past two decades have seen enormous progress on causal inference with a range of methodological approaches being developed. Tools range from Causal DAGS and marginal modeling and principle strata application over G-estimation, to mediation analysis and optimal dynamic treatment regimes dealing with time-varying confounders. The latter can be avoided when instrumental variables approaches are used instead. The range of options and opportunities that may lead to success - or failure - is quite enormous. Assumptions involved are intricate and to an important degree untestable. The methodology itself is demanding since it is working at the added abstract level brought by outcomes under different potential exposures than the one observed. The desire to bring causal answers and the seeming ease with which some of the methodology can be applied through custom made software bring great opportunities but also harbor a risk of producing misleading results. With Big Data knocking at our door the demands and risks are only likely to grow. In this talk we propose a road map and a compass to navigate the causal inference landscape in the busy environment of the practicing statistician. How to approach this from design to reporting and benefit from the available choices without falling in the traps is the challenge we plan to address in topic group 7.

MS 2.7

TG 6: Evaluating diagnostic tests and prediction models

Macaskill P.[1], Steyerberg E.[2], for the members of the TG6 group
[1]*School of Public Health, University of Sydney, Sydney, Australia,* [2]*Erasmus MC, Rotterdam, Netherlands*

The evaluation of diagnostic tests has a long history. Various performance measures are available for the cross-tabulation of presence of disease versus test result, including sensitivity, specificity, positive and negative predictive value (PPV, NPV) and likelihood ratios. Assessment of test performance must consider the potential role of the test which may include as a replacement for an existing test, an add-on test, or a triage test. The evaluation of prediction models typically focuses on measures for relative risk such as the odds ratio (OR) for the predictors in the model. The quality of predictions from a model for binary outcomes may be quantified by various measures, including those related to overall predictive performance (such as R^2), discrimination (such as the c statistic), and calibration (such as calibration-in-the-large). The increasing importance and relevance of quantifying the incremental value of a diagnostic test or marker to a prediction model has led to recent methodological work on measures that include the Net Reclassification Improvement (NRI) and Net Benefit (NB). The final impact attributed to a diagnostic test, marker, or prediction model requires a cost-effectiveness analysis supported by strong evidence, either from observational studies or from RCTs.

MS 2.8

TG 4: Measurement error

Küchenhoff H.[1]
[1]*Statistical Consulting Unit, LMU München, München, Germany*

In many cases, some variables in observational studies cannot be measured exactly. Relevant examples are blood pressure, nutrition variables, air pollution exposure and diagnosis of certain diseases. For categorical variables, measurement error is usually called misclassification.
The presence of measurement error can lead to bias in the estimation of parameters in the analysis of the data at hand. This holds in particular for multiple regression models (linear, generalized linear, Cox, etc.) with error in predictor variables. Bias depends on the type and the extent of measurement error. For example, the distinction between the so called classical additive measurement error and a Berkson type error is one essential aspect. Many methods have been developed to correct for measurement error and to include the measurement process into the statistical model. Regression calibration, full likelihood or Bayesian modeling, SIMEX are examples. Most of these methods rely on more information on measurement error, e.g. validation studies, replicate or further assumptions. The lack of this information and knowledge of the methods makes the application of measurement error methods difficult.
In the talk we will present some examples for the effect of measurement error and for correction methods. The aim of the group is to develop criteria and strategies for including measurement error into the analysis.