WBS Sommer Seminar 2014
jointly organized with CeMSIIS-ReUSe

RE-USE OF HEALTH DATA
FOR MEDICAL RESEARCH

Datum&Zeit: Donnerstag 12. Juni 2014, 14-18 Uhr
Ort: Jugendstilhörsaal der Medizinischen Universität Wien,
Bauteil 88 – Ebene 3, Spitalgasse 23, 1090 Wien
Plan siehe http://cemsiis.meduniwien.ac.at/allgemeines/anschrift/

AGENDA

Session 1		Chair: Georg Heinze
14.00-14.40	Susanne Rässler	Combining Data from Different Sources: Different Techniques - Different Worlds
14.40-15.00	Georg Heinze	Use of health data and innovative models for estimating prevalences: chronic kidney disease in Austria
15.00-15.40	Stefanie Rinderle-Ma	Re-use of Health Data for Clinical Process Analysis: The Evidence-based Medical Compliance Cluster (EBMC2)
15.40-16.10	Break	
Session 2		Chair: Martin Posch
16.10-16.50	Judit Simon	Using patient-reported outcome measures in health economics research: The issue of missing data
16.50-17.30	Hanno Ulmer	The use of routinely collected health examination data for medical research
17.30-17.50	Gerhard Svolba	When the patient becomes a customer. Re-using my experience gained in biometry for customer data mining
18.00 –	Informal debriefing in Altes AKH	

Registration of attendance (free): per e-mail to franz.koenig@meduniwien.ac.at until 9 June 2014.
Please feel free to distribute the announcement to colleagues. The WBS runs a mailing list for announcing talks in the field of biostatistics. For subscription to the mailing list, send an e-mail to franz.koenig@meduniwien.ac.at as well.

Speakers

Prof. Dr. Susanne Rässler

Lehrstuhl für Statistik und Ökonometrie
Otto-Friedrich Universität Bamberg
Feldkirchenstraße 21
D-96045 Bamberg

Ao Univ.-Prof. Dr. Georg Heinze

Section for Clinical Biometrics
Center for Medical Statistics, Informatics and
Intelligent Systems
Medical University of Vienna
Spitalgasse 23
A-1090 Vienna

Univ.-Prof. Dr. Stefanie Rinderle-Ma

Research Group Workflow Systems and Technology
Faculty of Computer Science
University of Vienna
Währingerstrasse 29/4.39
A-1090 Vienna

Univ.-Prof. Dr. Judit Simon

Department of Health Economics
Center for Public Health
Medical University of Vienna
Kinderspitalgasse 15/I
A-1090 Vienna

Ao Univ.-Prof. Dr. Hanno Ulmer

Department of Medical Statistics, Informatics and
Health Economics
Innsbruck Medical University
Schöpfstraße 41/1
A-6020 Innsbruck

Dr. Gerhard Svolba

SAS Austria
Mariahilfer Straße 116
A-1070 Vienna

Abstracts

Combining Data from Different Sources: Different Techniques - Different Worlds

Susanne Rässler

University of Bamberg, Germany

In this talk we will first give an overview about some main data integration techniques like record linkage and statistical matching. The latter is sometimes used synonymously for data fusion as well as propensity score matching. Record linkage aims at combining data from different sources that contain information from the same statistical units usually by using some identifiers, e.g. name and address or an identification number or other descriptors. Propensity score matching is typically used in observational studies to find statistical twins for a treatment group in another group that has not been treated. Thus, a control group is created which is comparable to the treatment group to, e.g., estimate average treatment effects. Although the statistical matching situation for creating a control group and fusing data seems to be very similar, we will note that propensity score matching should not be used in the case of data fusion.

Since record linkage as well as propensity score matching techniques are well known and understood, the talk will focus on the identification problem inherent in data fusion. Data fusion techniques typically aim to achieve a complete data file from different sources which do not contain the same units. Traditionally, data fusion is done on the basis of variables common to all files. It is well known that those approaches establish conditional independence of the (specific) variables not jointly observed given the common variables, although they may be conditionally dependent in reality. However, if the common variables are (carefully) chosen in a way that already establishes conditional independence, then inference about the actually unobserved association is valid. In terms of regression analysis, this implies that the explanatory power of the common variables is high concerning the specific variables. Unfortunately, this assumption is not testable yet but we are able to give so called uncertainty bounds. Using these bounds, we treat the data fusion situation as a problem of missing data by design and suggest imputation approaches to multiply impute the specific variables. Thus, sensitivity analyses are possible to account for violations of the conditional independence assumption.

Use of health data and innovative models for estimating prevalences: chronic kidney disease in Austria

Georg Heinze
Medical University of Vienna

Chronic kidney disease (CKD) is a progressive chronic disease, usually leading to the necessity of renal replacement therapy. CKD causes considerable and steadily increasing burden to health care providers. Therefore, estimation of the prevalence of CKD, in particular of its early stages, is of particular interest. We exemplify how several routine health data bases can be re-used and combined to estimate the prevalence of CKD among diabetic patients in Austria.

In particular, we made use of a) pseudonymized data of the Austrian Dialysis and Transplant Registry, b) the pseudonymized data base of health insurance claims ‚GAP-DRG' run by the Main Association of Austrian Social Security Institutions, c) the anonymized data base of hospital discharge diagnoses run by the Austrian Ministry of Health, and d) pseudonymized laboratory records routinely collected over several years in the Vienna General Hospital.

We show how these data sources were combined, using exact linkage with pseudonymized identifiers and probabilistic linkage based on six descriptors, strictly observing general rules of data protection. Furthermore, we discuss patient selection, high-dimensional modeling, prevalence estimation, calibration and resampling steps that were performed to obtain reliable point and interval estimates of prevalence of CKD stages. This work was funded by the European Commission, seventh framework programme, grant agreement number 241533 (SysKID).

Re-use of Health Data for Clinical Process Analysis: The Evidence-based Medical Compliance Cluster (EBMC2)

Stefanie Rinderle-Ma
University of Vienna

Analyzing health data from a process-oriented perspective has become more and more important during the last years. One goal can be to compare actual treatment processes with guidelines in order to analyze possible deviations. The Evidence-based Medical Compliance Cluster (EBMC2) aims at such analysis in the context of skin cancer treatment. The talk will present analysis goals in more detail, introduce the relevant guideline, and describe the data sources. Specifically, the presentation will focus on the extraction and integration of data from the sources in a process-oriented format for later analysis.

Using patient-reported outcome measures in health economics research: The issue of missing data

Judit Simon

Medical University of Vienna

Missing data is a known and considerable problem with patient-reported health outcome measures (PROMs) collected for medical/health economics research. EQ-5D-3L is a multi-attribute utility scale and it is one of the most widely used PROM in the economic evaluation of health care technologies. So far researchers have used very different methods to deal with missing EQ-5D-3L data. Current evidence recommends multiple imputation (MI), but there is little guidance on whether MI should be carried out at domain level or index values. This talk will present a research project aimed to evaluate the impact of imputing individual EQ-5D-3L domains versus index values to deal with missing information using a simulation study based on data from a large multinational clinical trial in the area of subarachnoid haemorrhage.

The use of routinely collected health examination data for medical research

Hanno Ulmer

Medical University of Innsbruck

In 1976, the general practitioner Leopold Bischof wrote an article in Methods of Information in Medicine entitled "Die Datenverarbeitung für die Gesundheitsvorsorge in Vorarlberg". In this work, Bischof laid the visionary foundation of a system of population-based, data-driven health examinations in Vorarlberg, the westernmost province of Austria. The Agency for Preventive- and Social Medicine began to routinely document health examinations, from 1985 onwards with an IT system that functions to this day. In the beginning, contrarily to Bischof's intent, the data-recording was mainly used for accounting purposes. It was only in 2003, that the data started to be applied for medical research under the name of "Vorarlberg Health Monitoring & Promotion Programme" (VHM&PP).

During the last decade, the research output from analyses of the VHM&PP database was considerable. Original publications on how risk factors track over time, on the way patterns of heart diseases vary by season, and on gender differences and secular trends in chronic disease have all made important contributions to medical literature. The research has provided a novel understanding of how gamma-glutamyltransferase and uric acid are associated with both cardiovascular and cancer outcomes. Recently VHM&PP contributed to several international pooled analyses which enhanced the understanding on how metabolic factors are involved in the risk of chronic diseases.

When the patient becomes a customer. Re-using my experience gained in biometry for customer data mining

Gerhard Svolba

SAS Austria

„Sample Size Planning, pre/post comparisons and p-values corrections in multiple testing" on the one side; "Campaign-response probabilities, risk weighted assets and lift-values" on the other side. - In 1999, I moved from medical statistics in the academic field to a private software company to increase the profitability and efficiency of our customers with topics like data mining, statistical forecasting and customer optimisation. This talk shows, which biometric experiences I could re-use in my new job; what was completely new for me; and which similarities exist in both areas.