Data Integration for Integrated Research and Care

Alfred Winter¹; Ralf-Dieter Hilgers²; Ralf Hofestäd³; Petra Knaup-Gregori⁴; Claudia Ose⁵; Antje Timmer⁶

¹Leipzig University, Institute for Medical Informatics, Statistics and Epidemiology, Leipzig, Germany;
²RWTH Aachen University, Institute for Medical Statistics, Aachen, Germany;
³Bielefeld University, Bioinformatics Department, Bielefeld, Germany;
⁴Heidelberg University, Institute for Medical Biometry and Informatics, Heidelberg, Germany;
⁵Duisburg-Essen University, Faculty of Medicine, Centre for Clinical Trials, Essen, Germany;
⁶Carl von Ossietzky University of Oldenburg, Department for Epidemiology and Biometry, Oldenburg, Germany

Keywords
Medical informatics, medical statistics, epidemiology

Summary
A national German funding initiative for Medical Informatics focusing at data integration for medicine gives an opportunity to reopen a window to Germany. In the open window appears a best paper selection of the 2015 annual conference of the German Society of Medical Informatics, Biometry and Epidemiology and papers of the German journal *GMS Medical Informatics, Biometry and Epidemiology* (MIBE). The publications in focus deal with data integration by transferring clinical routine data into an electronic data capture (EDC) system, using natural language processing to make unstructured data processable, measuring quality of record linkage, and by using a unified metadata scheme for integrated documentation in laboratories. Two additional papers present methods for data analysis especially for change point detection in binary sequences and for analyzing categorical data.

Correspondence to:
Prof. Dr. Alfred Winter
Leipzig University
Institute for Medical Informatics, Statistics and Epidemiology
Haertelstr. 16–18
04107 Leipzig
Germany
E-mail: alfred.winter@imise.uni-leipzig.de

© Schattauer 2016

Medical Informatics seemed to become an old-fashioned term in Germany, but since 8 months we see an exciting revival. German federal government launched a national funding initiative for Medical Informatics aiming at establishing data integration centers for better medical care and research. Although no money has been spent so far first integration effects are already visible: We have never before experienced such close cooperation in our field between information management departments of university hospitals and medical informatics research departments, between doctors in care and IT specialists, between hospital management and academia and between different sites across the country. And cooperation of people is the very prerequisite for integration of systems.

This inspiring and encouraging developments give opportunity to open again a window to Germany in *Methods of Information in Medicine* (MIM) and to focus on data integration.

In the window first appear three papers of this MIM issue. They are part of a best paper selection based on presentations having been given at the 2015 annual conference of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS). Two of them are dealing with the data integration challenge.
Kaspar et al. address data integration by developing a solution for transferring clinical routine data into an electronic data capture (EDC) system supporting clinical studies. As an intermediary they use an R-based data warehouse [1]. Of course you can only integrate data you have at hand. In our hospitals we actually have a plethora of data available. However there is still a considerable amount of paper based data and in (too) many cases even computer based data is not ready for data processing because of its unstructured representation. Löpprich et al. introduce their natural language processing solution for automatically extracting and classifying of diagnoses and state of disease from free text diagnostic reports [2].

Looking through this window to Germany you can also see GMDS’ e-journal *GMS Medical Informatics, Biometry and Epidemiology (MIBE)* and you can find another paper examining the data integration challenge. Cancer and other registries have to link, i.e. integrate, records from different data sources and sites pertaining to the same patient. If unique identifiers are not available record linkage errors unavoidably occur. Schmidtmann et al. quantified and measured the quality of record linkage in a highly automated cancer registry that relies on encrypted identity data. GMS Med Inform Biom Epidemiol. 2016; 12(1). Available from: http://www.egms.de/static/de/journals/mibe/2016–12/mibe000164.shtml.

The third paper from GMDS 2015 in this MIM issue deals with analysis of (hopefully) integrated data. If fixators are used in traumatic orthopedic surgery the skeletal pins of the fixators cut through the skin and frequently infections occur. Ellenberger et al. show how to find out if a change in treatment strategy led to a decrease in the infection rate. They use an unconditional test for change point detection in the binary sequence describing the courses of treatment [5]. Finally, I would like to draw the attention of users of statistical methods for data analysis to a more tutorial overview on methods for analyzing categorical outcome data. Kersten explains how to use statistical software in order to solve this problem without loosing information by dichotomizing the categorical variables [6].

**References**