

Please join the Biometric Colloquium

CARSTEN OLIVER SCHMIDT

Institute for Community Medicine, SHIP-KEF,
University Medicine of Greifswald, Greifswald, Germany

WHAT DO I NEED FOR AUTOMATED DATA QUALITY ASSESSMENTS?

CONCEPT AND APPLICATION EXAMPLE BASED ON THE STUDY OF HEALTH IN POMERANIA (SHIP).

May 27th, 2026 at 9:00 am

Seminarraum Center for Medical Data Science,
Spitalgasse 23, Room 88.03.513
Medical University of Vienna, 1090 Wien
Host: Georg Heinze

Abstract:

Data quality assessment is a prerequisite for reliable statistical analysis. It comprises a broad range of procedures, from simple rule-based checks, such as detecting range violations or invalid data types, to statistical methods for identifying outliers, unexpected distributions, associations, or cluster effects that may indicate measurement error or measurement heterogeneity.

Automating these assessments in complex datasets requires balancing two competing needs: flexibility, to adapt methods to specific data properties, and standardization, to ensure reproducible and scalable workflows. This talk presents a metadata-driven approach to this challenge using data quality assessment tools in R and Stata, specifically `dataquieR` and `dqrep`.

The central role of metadata will be highlighted. Metadata encode expectations about variables, permissible values, measurement properties, and study-specific requirements, thereby guiding automated checks and supporting transparent reporting. The approach will be illustrated using data from the population-based Study of Health in Pomerania (SHIP).

The talk's key message is that systematic organization and use of study knowledge through metadata can make the scientific data analysis workflow more efficient, reproducible, and transparent.