

## **Kolloquium „Statistische Methoden in der empirischen Forschung“**

Wann: 17. November 2015, 17:00 – 18:30 Uhr

Wo: Robert Koch-Institut | Nordufer 20 | 13353 Berlin (Wedding),  
S41, S42, U9 Westhafen | U9, Bus 142 Amrumer Str

**Simone Wahl (Helmholtz-Zentrum München)**

### **Prädiktive Modellgüte bei fehlenden Werten: Multiple Imputation kombiniert mit Resampling**

Fehlende Werte stellen ein häufiges Problem bei Studien am Menschen dar. Sie treten beispielsweise aufgrund von unvollständig verfügbaren biologischen Proben auf, aufgrund von Nachweisgrenzen bei technischen Geräten oder aufgrund von unvollständig ausgefüllten Fragebögen. Häufig stellt multiple Imputation (MI) eine angemessene Strategie dar, mit den fehlenden Werten umzugehen. Sie umfasst drei Schritte: (1) die eigentliche Imputation, also das mehrfache Ersetzen der fehlenden Werte mit plausiblen Werten, (2) die statistische Auswertung basierend auf jedem der imputierten Datensätze, und (3) die Kombination der daraus resultierenden multiplen Ergebnisse.

Für viele Schätzer gibt es einfache Regeln, Schritt (3) umzusetzen. Ist das Zielkriterium jedoch die prädiktive Güte eines Modells (alleine oder im Vergleich zu einem Basismodell), also möchte man beispielsweise Maße wie die AUC (area under the receiver operating characteristic curve) oder Pseudo-R<sup>2</sup>-Maße schätzen, so ist bislang weitaus weniger etabliert, wie man diese im Rahmen der MI zusammenfassen und Konfidenzintervalle für diese erhalten kann.

Im Vortrag werden verschiedene Ansätze vorgestellt und in einer umfangreichen Simulationsstudie verglichen, wie man durch die Kombination von MI und Resampling-Strategien unverzerrte Schätzer für die prädiktive Modellgüte von logistischen Regressionsmodellen erhalten kann und wie valide Konfidenzintervalle für diese konstruiert werden können. Dabei wird der Einfluss von Studiencharakteristiken (Stichprobengröße, Anzahl der Variablen, Anteil und Mustern von fehlenden Werten, tatsächliche Modellgüte etc.) auf die Resultate untersucht. In einem Anwendungsbeispiel wird die Methodik auf die Prädiktion der Gesamtmortalität mithilfe von Entzündungsmarkern im Blut in einer bevölkerungsbasierten Studie angewandt.