

## Analysen mit der DRG-Statistik - Herausforderungen und Lösungsansätze

**Lösch, Tobias; Hochgürtel, Tim**

Statistisches Bundesamt

tobias.loesch@destatis.de; tim.hochguertel@destatis.de

Mit der DRG-Statistik (Fallpauschalenbezogene Krankenhausstatistik) steht eine Vollerhebung aller vollstationären Krankenhausfälle in allgemeinen Krankenhäusern für statistische Analysen zur Verfügung. Über die Forschungsdatenzentren (FDZ) der Statistischen Ämter des Bundes und der Länder haben Wissenschaftler/-innen die Möglichkeit die Mikrodaten der DRG-Statistik zu nutzen. Aus Gründen des Datenschutzes können diese nicht direkt für wissenschaftliche Forschungsvorhaben an den jeweiligen Datennutzer übermittelt werden. Über die kontrollierte Datenfernverarbeitung besteht jedoch die Möglichkeit Auswertungsprogramme in der Programmiersprache SAS an das Forschungsdatenzentrum des Statistischen Bundesamtes zu übermitteln. Dort werden die SAS-Programme ausgeführt und die Ergebnisse nach einer Geheimhaltungsprüfung an den Datennutzer übermittelt.

Die Analyse der DRG-Statistik zeichnet sich in der Regel durch lange Rechenzeiten aus. Dies ist in erster Linie auf den Umfang der DRG-Statistik zurückzuführen, die seit dem Berichtsjahr 2005 zur Verfügung steht und pro Berichtsjahr nahezu 20 Millionen Datensätze beinhaltet. Dennoch kann der Nutzer durch eine effiziente Programmierung die Dauer der Analysen entscheidend reduzieren.

Häufig identifizieren die Datennutzer besondere Teilpopulationen unter den Behandlungsfällen, welche über ICD-Haupt- und Nebendiagnosen oder OPS-Kodes abgegrenzt werden. Daneben sind oft die Häufigkeiten von bestimmten ICD-Haupt- und Nebendiagnosen oder OPS-Kodes von Interesse. Dies hat sich in der Vergangenheit als äußerst rechenzeitintensiv erweisen.

Die DRG-Statistik umfasst 89 Merkmale mit ICD-Nebendiagnosen sowie 100 Merkmale zu den OPS-Kodes. In der Regel weist ein Krankenhausfall nur wenige ICD-Nebendiagnosen und OPS-Kodes auf, so dass die meisten Merkmale mit Bezug zu den ICD-Nebendiagnosen und OPS-Kodes mit Missings besetzt sind. Um die Laufzeiten der Analysen zu reduzieren, wurden verschiedene Methoden zur Identifikation der relevanten Teilgruppen und die Auszählung von Häufigkeiten getestet. Bisher wurden die Laufzeiten eines array-basierten Verfahrens, eines proc transpose- und proc sql-basierten Vorgehen sowie einer proc transpose- und proc summary- basierten Methode ermittelt und mit einander verglichen. Weitere Tests, unter anderem in einer proc iml Umgebung, werden derzeit geplant.

Für die Datennutzer wurde ein Makro entwickelt, welches in Abhängigkeit von nutzerspezifischen Parametern die nach den bisherigen Erkenntnissen effizienteste Methoden zur Erstellung neuer Dummy- und Häufigkeitsvariablen wählt.

Literatur:

Hochgürtel, Tim / Lösch, Tobias: Das SAS-Makro newvar. Entwicklung und Anwendung eines Hilfsinstruments zur effizienten Erstellung neuer Variablen in der DRG-Statistik, FDZ-Arbeitspapier 44.