

WF STUPA: Data Management der Studie zur Wirksamkeit und Verträglichkeit von VERUM auf die peridurale Narbenbildung und rekurrente Schmerzen bei Bandscheibenoperationen

Bezler, Irina; Haag, Ulrich

Hochschule Ulm

bezler@mail.hs-ulm.de; uhaag@mail.hs-ulm.de

Ziele:

Ziel dieser Arbeitsgruppe ist zum einen das Erstellen von auswertbaren SAS-Dateien mit allen benötigten Variablen, die für weitere Arbeitsschritte zur Verfügung gestellt werden und zum anderen eine Dokumentation, der Kollegen alles wichtige zu den Daten, den Variablen und - falls vorhanden - die Formeln zur Berechnung dieser entnehmen können.

Vorgehen:

Grundlage für das Data Management sind eine Access-Datenbank und zwei ASCII-Dateien, die Patientendaten (Stammdatensätze, Operationszeitpunkte, Komplikationen, Gruppeneinteilung u. ä.) enthalten. Die Bearbeitung findet mit SAS statt.

Um ein späteres „Mergen“ zu ermöglichen, werden die alphanumerischen Patientennummern in rein numerische Werte umgewandelt, denen man noch das jeweilige Studienzentrum entnehmen kann.

Anschließend werden die neun einzeln vorliegenden Tabellen mit einem Makro in SAS eingelesen, wobei jeder Tabelle die neu generierte Patienten-ID angefügt wird. Diese Tabellen werden nun zu größeren logischen Tabellen zusammengefügt: es liegen jetzt die Stammdatentabelle, die Komplikationstabelle und die Ausschlusstabelle vor. Zum Schluss wird an jede dieser Tabellen eine weitere Tabelle angefügt, die diejenigen Variablen enthält, die immer schnell einsehbar sein sollten, z. B. ID, Gruppe, Geschlecht, Alter und Zentrum.

Als nächster Schritt werden für die weitere Auswertung relevante Variablen erstellt. Hier werden nach bekannter Formel der BMI berechnet. Des Weiteren wird die Hauptzielgröße, der FFbH-Score, nach vorgegebener Formel errechnet. Weitere neue Variablen sind: Alter der Patienten und Schnitt-Naht-Zeiten.

Um die vorliegenden Formate in SAS einzulesen und um sie permanent anzulegen wird ein weiteres Makro verwendet. In diesem Makro wird auch gleichzeitig die Dokumentation mit Hilfe der Option „fmtlib“ vorgenommen. Nun können die Formate wie beispielsweise „Geschlecht“ oder „ja/nein“ auf die Variablen der Tabellen angewendet werden. In diesem Schritt werden auch eventuell vorkommende Missing-Werte, die vorher mit „9“ codiert waren, durch „.“ ersetzt, um eine spätere numerische Auswertung zu erleichtern.

Im folgenden Arbeitsschritt wird eine Tabellentransformation vorgenommen: die Angaben aus dem Schmerzfragebogen liegen pro Patient dreimal vor, da die Daten zu drei Zeitpunkten erhoben wurden. Diese vertikale Datenstruktur ist für grafische Zeitverläufe geeignet, nicht aber für Differenzberechnungen. Die Datenstruktur hierfür erreicht man durch Horizontalisierung der Tabelle mittels „PROC TRANSPOSE“.

Ein weiteres Teilprojekt ist die Query-Generierung als Folge von Plausibilitätsprüfungen durch „IF“-Abfragen. Die Tabellen werden beispielsweise auf Missing-Werte, Altersangaben außerhalb des zugelassenen Intervalls und unrealistische BMI-Werte geprüft. Im Falle von entdeckten Unstimmigkeiten werden die entsprechenden Werte nach Studienzentrum gruppiert und mittels „ODS“ in jeweils eine PDF-Datei pro Zentrum geschrieben. So erhält jedes Studienzentrum eine Übersicht zu allen seinen fehlerhaften Angaben.

Die abschließende Dokumentation besteht zum Einen aus der Aufstellung der Formate durch „fmtlib“ und zum Anderen aus einer Übersicht über alle vorkommenden Variablen. Letzteres wird auf der Basis der „proc contents“ erstellt und durch die Formeln zur Berechnung der neuen Variablen ergänzt.

Ergebnisse:

Nach den Maßnahmen des Data Managements liegt nun die große Patientenstammdatentabelle mit passenden Formaten und den benötigten neuen Variablen vor. Durch das Zusammenfügen aller relevanten Tabellen ist nun schnellerer Zugriff gewährleistet und späteres „mergen“ erübrigt sich.

Die mit vorher mit „9“ codierten Missing-Werte hätten Auswertungen von numerischen Variablen erschwert. Die vorgenommene Umwandlung in „.“ beseitigt diese Fehlerquelle.

Die Angaben des Schmerzbogens, die pro Patienten dreimal erfragt wurden, liegen nun sowohl horizontal für Differenzberechnungen als auch vertikal angeordnet für grafische Zeitverläufe vor.

Ausblick:

Beim Erstellen der Formate kann noch eine weitere Automatisierung vorgenommen werden. Hierzu wird ein Makro benötigt, das automatisch pro Format die Werte und Bezeichnungen einliest. Dies wird bisher bei jedem Format von Hand eingegeben.