

Auswertung stratifizierter Studien mittels des Cox-Modells im Falle von wenigen Events

Fink, Simon

Universität Ulm

simon-fink@gmx.net

Für die folgende Fragestellung wurden 3 verschiedene Auswertungsmethoden für stratifizierte Überlebenszeitstudien im Hinblick auf die Schätzung des Hazard Ratios betrachtet:

1. Stratifizierte Auswertung
2. Nichtstratifizierte Auswertung mit Stratum als Kovariable ohne Wechselwirkung
3. Nichtstratifizierte Auswertung mit Stratum als Kovariable mit Wechselwirkung.

Im Zuge von Überlebenszeitstudien kann es, je nach Zielgröße, vorkommen, dass aufgrund einer sehr niedrigen Eventrate nur wenige Events beobachtet werden können. Eine solche Datensituation wirkt sich negativ auf die Größe der Konfidenzintervalle aus und kann zusätzlich die Schätzer, insbesondere für den Behandlungseffekt, verzerren. Zusätzlich verstärkt wird dieser Effekt durch eine Stratifizierung der Daten. So kann es sogar dazu führen, dass in einem oder mehreren Strata gar kein Event auftritt.

Würde in einem solchen Fall die Stratifizierung vernachlässigt und stattdessen eine nichtstratifizierte Auswertung vorgenommen, bei der die Stratifizierungsvariablen als Kovariablen ins Modell aufgenommen werden, kann die Situation, dass in einer Gruppe keine Events eintreten, meist vermieden werden.

Um eventuelle Unterschiede dieser nichtstratifizierten Methode gegenüber einer üblichen stratifizierten Auswertung identifizieren zu können, wurden die Schätzer für das Hazard Ratio mittels verschiedener Simulationen miteinander verglichen. Dafür wurden im Rahmen eines SAS Macros exponentialverteilte Überlebenszeiten generiert und anschließend die 3 anfangs aufgeführten Modelle mit der PROC PHREG an die erzeugten Daten angepasst.

Um sich nun langsam an die eigentlich zu untersuchende Situation von wenigen Events heranzuarbeiten, wurden in der Simulation erst vollständige Datensätze produziert, also Daten ohne jegliche Zensierungen, anschließend wurde die Eventrate dann schrittweise erst auf 70-80% und dann auf unter 10% herabgesenkt.

Für alle drei Fälle wurde bei variierenden Parametern die Power der unterschiedlichen Modelle geschätzt und unter anderem grafisch miteinander verglichen. Anhand der Power war feststellbar, ob zwischen den betrachteten Modellen grundlegende Unterschiede zu beobachten sind.