

Beurteilung der Anpassungsgüte multivariater Regressionsmodelle in Überlebenszeitmodellen unter besonderer Berücksichtigung des C-Index

Katrin Kupas¹, Hendrik Schmidt²

¹Statistical Consultant, Deutschland; ²Boehringer Ingelheim Pharma GmbH & Co KG; katrin.kupas.ext@boehringer-ingelheim.com

In der Arzneimittelentwicklung haben Subgruppenanalysen und die Identifizierung bestimmter Prädiktoren für Krankheiten und besonders Krankheitsverläufe in den letzten Jahren an Bedeutung gewonnen. Verstärkt wird diese Bedeutung zudem durch die zunehmend frühe Nutzenbewertung von Arzneimitteln sowie durch den Trend zu personalisierter Medizin, bei der anhand von bestimmten Patientencharakteristika der Erfolg einer Therapie eingeschätzt wird.

In vielen Therapiegebieten ist die Zeit bis zum Eintritt eines Ereignisses (Tod, Verschlechterung des Krankheitszustandes, etc.) ein wichtiger und patientenrelevanter Endpunkt zur Bestimmung des Krankheitsverlaufes. Häufig wird zu Beginn einer Studie von jedem Patienten eine Vielzahl an Charakteristika erhoben, die als mögliche Prädiktoren für diese Endpunkte in Frage kommen. Multivariate Regressionsmodelle sind eine Möglichkeit, aus diesen Kovariaten signifikante Prädiktoren und damit entsprechende Patientensubgruppen zu identifizieren und auszuwerten.

SAS® bietet mit PROC PHREG verschiedene Algorithmen an, zum Beispiel eine vorwärts- oder rückwärtsgerichtete oder auch eine bidirektionale Selektion, die aus diesem Pool an Kovariaten automatisch das optimale Modell auswählt. Dabei werden anhand der einzelnen Chi-Quadrat-Statistiken und einer vom Benutzer zu setzenden Signifikanzgrenze nacheinander die signifikantesten Kovariaten ausgewählt oder die am wenigsten signifikanten Kovariaten eliminiert.

Insbesondere beim Vorliegen vieler Kovariaten und/oder kleiner Patientenzahlen stellt sich die Frage nach der Anpassungsgüte der erhaltenen „optimalen“ Modelle. Der C-Index als absolute Maßzahl bietet eine Möglichkeit, die Anpassungsgüte des Modells an die Daten zu bewerten und verschiedene Modelle zu vergleichen. Wir stellen den mathematischen Hintergrund zum C-Index dar und präsentieren einen Algorithmus zu dessen Berechnung, der mit SAS® implementiert wurde. Zudem präsentieren wir verschiedene Beispiele, wie die Variablenselektion mit Hilfe des C-Index bewertet werden kann.

Literatur:

- Pencina MJ, D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine* 2004; 23:2109-2123
- Steyerberg EW. *Clinical Prediction Models: A practical Approach to Development, validation and Updating*. Springer, 2009. *Epidemiology* 2010; 21:128-138
- Uno H, Cai T, Pencina MJ, D'Agostino RB and Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine* 2011; 30:1105-1117
- Wiegand RE. Performance of using multiple stepwise algorithms for variable selection. *Statistics in Medicine* 2010; 29:1647-1659