

Rouven Behnisch
Dr. sc. hum

Comparison of methods to analyze time-to-event endpoints when treatment effect is delayed

Fach/Einrichtung: Medizinische Biometrie
Doktorvater: Prof. Dr. Meinhard Kieser

Die jüngsten Fortschritte in der Onkologie zu Therapien, die auf der Aktivierung und Verstärkung des Immunsystems Tumorzellen zu erkennen und zu bekämpfen beruhen haben diese Arbeit motiviert. Eine häufig beobachtete Eigenschaft dieser immuntherapeutischen Ansätze ist, dass der beobachtbare Behandlungseffekt um die benötigte Zeit eine Immunantwort zu trainieren verzögert wird. Insbesondere bei Ereigniszeitendpunkten wie dem Gesamtüberleben oder dem progressionsfreien Überleben muss dies bei der statistischen Auswertung sorgfältig berücksichtigt werden, da die üblicherweise verwendeten Methoden auf der Annahme der proportionalen Hazards beruhen. Diese Annahme ist verletzt, wenn der Behandlungseffekt verzögert eintritt und die herkömmlichen Methoden haben eine geringere statistische Power einen Unterschied zwischen den Therapien zu erkennen. Ziel dieser Arbeit war es daher, die Performance verschiedener Alternativen zum häufig verwendeten Logrank-Test in Bezug auf den Fehler 1. Art und die Power in dieser Situation zu untersuchen.

In einer systematischen Literaturrecherche wurden zuerst statistische Methoden identifiziert, die für die Analyse von Ereigniszeitdaten vorgeschlagen wurden, und ob diese insbesondere entwickelt wurden, um nicht-proportionale Hazards umzugehen. Die Methoden wurden dann in einer umfangreichen Simulationsstudie unter Berücksichtigung der folgenden Parameter verglichen: die Gesamtstudiendauer τ , der Anteil des Rekrutierungszeitraums acc , die Form des Weibull-Hazards k_C und das mediane Überleben med_C im Kontrollarm und der maximale Behandlungseffekt θ sowie die Verzögerung t_2^* und der Zeitpunkt der Änderung t_1^* des generalisierten linearen Lag-Modells. Als Gütemaß wurde der Fehler 1. Art und die Power der Methoden und der Einfluss der verschiedenen Simulationsparameter auf diese Gütemaße untersucht. Bei Methoden, bei denen Analyseparameter gewählt werden mussten, wurde auch die Auswirkung einer falschen Spezifikation dieser Parameter untersucht.

Die meisten Methoden kontrollierten den Fehler 1. Art, und die dies nicht taten, waren bereits aus früheren Veröffentlichungen bekannt. In Bezug auf die Power schnitten dem datengenerierenden Prozess nahekommende Methoden in der Simulationsstudie am besten ab und waren zumindest in Szenarien mit nichtproportionalen Hazards nicht anfällig gegenüber einer falschen Spezifikation der Analyseparameter. Die Spezifikation dieser Parameter erfordert jedoch die Kenntnis von Form und Ausmaß der zu erwartenden Verzögerung im Voraus. Ist dies nicht der Fall, können andere Alternativen wie der Fleming-Harrington Test für späte Unterschiede oder Kombinationen von gewichteten Logrank-Statistiken verwendet werden. Ihre Power hängt jedoch von der zugrundeliegenden Verteilung der Daten ab. Für Szenarien mit proportionalen Hazards hat der Logrank-Test die größte Power unter allen Tests, die den Typ 1 Fehler kontrollieren.

Durch einen Verweis auf das R-Paket, in dem die Methode implementiert ist, oder durch die Bereitstellung eigenen Codes im Anhang wird die Verwendung der vorgestellten Methoden erleichtert. Die Interpretation der Ergebnisse dieser Methoden ist jedoch durch das Fehlen interpretierbarer Effektmaße oft schwierig. Einige Aspekte dieser Arbeit, wie z.B. die

Berechnung des Stichprobenumfangs oder die Anwendung der Methoden in adaptiven oder gruppensequentiellen Designs könnten zukünftige Forschung wert sein.

Diese Arbeit liefert eine detaillierte und gut strukturierte Übersicht alternativer Analysemethoden zum Logrank-Test, wenn Ereigniszeitdaten mit nichtproportionalen Hazards vorliegen. Darüber hinaus kann die umfangreiche Simulationsstudie zusammen mit dem Platzierungssystem helfen geeignete Methode zur Analyse zukünftiger Studien fundiert auszuwählen. Im Falle proportionaler Hazards bleibt der Logrank-Test die Methode der Wahl, und sofern der Datenerzeugungsprozess bei nichtproportionalen Hazards vollständig verstanden ist, sollten Methoden gewählt werden, die dieses zusätzliche Wissen nutzen. Falls jedoch die genaue Struktur der Verzögerung nicht bekannt ist, sollte auf vielseitige Methoden zurückgegriffen werden, die gewichtete Logrank-Statistiken kombinieren.