

Julia Krzykalla
Dr. sc. hum.

Identification of predictive factors for treatment efficacy in randomized and non-randomized settings

Fach/Einrichtung: DKFZ (Deutsches Krebsforschungszentrum)
Doktormutter: Prof. Dr. Annette Kopp-Schneider

Im Rahmen der Präzisionsmedizin entstehen aktuell viele neue Forschungsprojekte, sowohl in der Medizin, als auch in angrenzenden Forschungsgebieten wie der Statistik. Um eine solide wissenschaftliche Basis für individualisierte Behandlungsstrategien zu schaffen, sind neue, effizientere Auswertungsmethoden gefragt. Neben der Vorhersage des Behandlungseffekts für einen einzelnen Patienten ist es wichtig, Faktoren zu identifizieren, die gegebenenfalls über den Nutzen einer Behandlung entscheiden. Die meisten statistischen Methoden konzentrieren sich primär auf eine möglichst genaue Vorhersage des Behandlungseffekts. Welche Faktoren im Einzelnen für das Vorhersagemodell verwendet werden und ob diese Faktoren auch klinisch relevant sind, ist dabei zweitrangig. Eine inhaltliche Interpretierbarkeit des Modells ist folglich nicht immer gegeben und somit sind bestehende Methoden meist nicht geeignet, um Indikatoren für eine effektive Behandlung ('prädiktive Faktoren') zu identifizieren.

Der model-based recursive partitioning (MOB)-Ansatz von Seibold et al. (2016) bildet mithilfe von Entscheidungsbäumen Subgruppen von Patienten mit ähnlicher Prognose und homogenem Behandlungseffekt. Methodisch basiert der Ansatz auf einem Modell bestehend aus einem Intercept und einem Term für den Effekt der Behandlung auf die Zielvariable ('Basismodell'). Wird die Schätzung der Modellkoeffizienten von einem Biomarker beeinflusst, empfiehlt sich eine Teilung des Patientenkollektivs entsprechend dieses Biomarkers. Ist die Schätzung des Intercepts betroffen, kommt dies einem prognostischen Effekt gleich, eine Instabilität des geschätzten Behandlungseffekts kann als prädiktiver Effekt interpretiert werden. Welche Art von Effekt für die jeweilige Aufteilung verantwortlich ist, bleibt jedoch unbekannt und somit lassen sich prädiktive und prognostische Faktoren nicht unterscheiden. In der vorliegenden Arbeit wird eine Modifizierung des MOB-Ansatzes, genannt predMOB, vorgeschlagen. Durch eine Reparametrisierung des Basismodells erfolgt eine Teilung ausschließlich aufgrund prädiktiver Effekte. Beide Algorithmen, MOB und predMOB, können mit Ensembleverfahren kombiniert werden, um robustere Ergebnisse zu erzielen. Dazu wird eine Vielzahl von Bäumen ('Wald') anhand zufälliger Teilstichproben aus den Daten erzeugt. Welche der untersuchten Variablen für die Konstruktion des Waldes entscheidend sind, kann anschließend mittels Kenngrößen der variable importance beurteilt werden. PredMOB-Wälder erfassen die wahren prädiktiven Faktoren zuverlässig, und im Gegensatz zum MOB-Ansatz bleiben Variablen, die nur prognostische Effekte zeigen, unbeachtet. Insbesondere werden auch mehrere prädiktive Faktoren erkannt, die entweder einzeln oder gemeinsam auf den Behandlungseffekt einwirken. Darüber hinaus erweist sich der predMOB Ansatz als robust gegenüber Korrelationen des prädiktiven Faktors mit anderen prognostischen Variablen. Vorhersagen zum individuellen

Behandlungseffekt eines neuen Patienten sowie verlässliche Schätzungen für die prädiktiven Effekte sind ebenfalls möglich.

In der Praxis sind randomisierte Daten oft nicht verfügbar, weshalb im zweiten Teil dieser Arbeit die Anwendung des predMOB auf nicht-randomisierte Daten, z.B. aus Beobachtungsstudien, untersucht wird. Dazu werden gängige Korrekturmethode (Kovariablenadjustierung, inverse probability of treatment weighting (IPTW), Matching und eine Kombination aus Kovariablenadjustierung und IPTW) mit predMOB-Wäldern kombiniert und hinsichtlich ihrer Fähigkeit verglichen, eine Verzerrung aufgrund Confoundings zu verhindern. Kovariablenadjustierung schneidet hinsichtlich der Identifizierung prädiktiver Faktoren, aber auch für die Vorhersage des individuellen Behandlungseffekts und die Schätzung der prädiktiven Effekte am besten ab. Die Kombination mit (IPTW), welche vergleichbare Ergebnisse liefert, ist aufgrund potenzieller Vorteile in komplexeren Szenarien dennoch empfehlenswert.

Insgesamt zeigte sich, dass predMOB einen geeigneten Ansatz für die spezifische Suche nach prädiktiven Faktoren darstellt. Er bietet die Flexibilität baumbasierter Methoden, beliebige funktionale Zusammenhänge und Interaktionen höherer Ordnung zu finden, ohne diese vorab spezifizieren zu müssen. Darüber hinaus ist er leicht auf höher-dimensionale Daten anwendbar. Um ein vollständiges Bild zu erhalten, ist es in der Praxis ratsam, predMOB im Anschluss an MOB anzuwenden. Kovariablenadjustierung oder ein mit IPTW kombinierter Ansatz können verwendet werden, um Verzerrungen in nicht-randomisierten Daten zu vermeiden. Der predMOB Ansatz, so wie hier vorgestellt, ist auf stetige und binäre Endpunkte anwendbar. Eine Erweiterung für Überlebenszeitendpunkte und konkurrierende Risiken wäre eine attraktive Option, insbesondere für die onkologische Forschung.