

Transfer von ML-Modellen zur Vorhersage des Transplantatüberlebens aus einer US-Kohorte auf Daten des deutschen Organtransplantationsregisters: Machbarkeit, Vorhersagegüte und Rekalibrierung

24.01.2023 | Kandidatin: Franziska Schmidt | Referentin: Prof. Dr. Antje Jahn | Korreferent: Prof. Dr. Gunter Grieser

Zusammenfassung:

Die Kalibrierung ist neben der Diskriminierungsfähigkeit ein wichtiges Merkmal von klinischen Prognosemodellen. Verwendet man ein bereits vorhandenes Prognosemodell für eine neue Kohorte, muss stets untersucht werden, wie gut die Kalibrierung dieses Modells für die neue Kohorte ist. Schlecht kalibrierte Risikoschätzungen können zu falschen Erwartungen bei Patienten führen und sind somit aus ethischen Gründen problematisch. In bestimmten Fällen kann es dennoch von Vorteil sein, ein bereits vorhandenes Modell zu verwenden, anstatt ein neues zu trainieren. Insbesondere wenn die Stichprobenzahl in der neuen Kohorte gering ist, kann so dem Risiko einer Überanpassung entgegengewirkt werden. Dazu muss das alte Modell zunächst rekaliert werden, um eine zufriedenstellende Kalibrierung zu erreichen. In der Überlebenszeitanalyse gibt es für die Cox-Regression und andere semi-parametrische und parametrische Modelle gut erforschte Rekalibrierungsmethoden. Bei der Entwicklung von Prognosemodellen werden zunehmend auch speziell auf zensierte Daten angepasste Methoden des maschinellen Lernens eingesetzt. Für diese Methoden gibt es jedoch bislang keine bekannten Rekalibrierungsmethoden. In dieser Arbeit wird daher eine Methode zur Rekalibrierung des Random Survival Forests entwickelt und evaluiert. Dazu sollen Prognosemodelle zu Nierentransplantationen, welche mit US-amerikanischen Daten entwickelt wurden, auf eine deutsche Kohorte angewendet werden. Hierbei werden ein Cox-Regressionsmodell und ein Random Survival Forest betrachtet. Es wird zunächst untersucht, welche Voraussetzungen erfüllt sein müssen, um die US-Modelle auf die deutsche Kohorte anwenden zu können. Anschließend wird die Kalibrierung der US Modelle auf den deutschen Daten untersucht und eine Rekalibrierung durchgeführt. Für den Random Survival Forest wird hierfür die in dieser Arbeit entwickelte Methode verwendet. Abschließend wird die Performance der nicht rekalierten, rekalierten und auf den deutschen Daten neu trainierten Modelle verglichen.

Methoden:

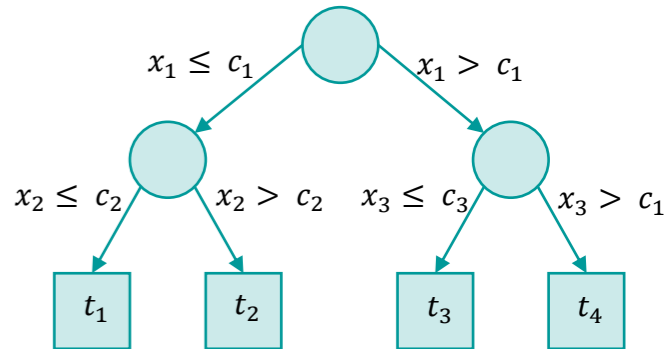
- Das deutsche Transplantationstegister enthält Daten aus drei verschiedenen Datenquellen, deutsche und US-Datensätze unterscheiden sich zudem stark in ihrer Struktur und den erhobenen Kovariaten. Daher ist der Feature Engineering Prozess aufwändig und erfordert Domänenwissen.
- Systematische Suche nach publizierten Prognosemodellen, die auf Cox-Regression oder dem Random Survival Forest beruhen und welche mit den von UNOS bereitgestellten US-amerikanischen Daten erstellt wurden
- Trainieren von eigenen Modellen auf den US-Daten mit Kovariatenauswahl angelehnt an Rao et.al 2009 [1]
- Für Cox-Regression [2] und den RSF [3] wird mit den deutschen Daten jeweils ein neues Modell trainiert, das auf den US-Daten trainierte Modell wird ohne Anpassung verwendet, oder das auf den US-Daten trainierte Modell wird rekaliert.
- Für die Rekalibrierung der Cox-Regression wird, wie von von Van Houwelingen 2000 [4] vorgeschlagen, der Baseline-Hazard neu geschätzt und ein eindimensionaler Skalierungsfaktor für den Slope berechnet.
- Für die Rekalibrierung des RSF wird eine neue Methode vorgestellt. Diese ist eine Erweiterung einer Methode von Dankowski und Ziegler 2016 [5] für die Rekalibrierung von RF zur Wahrscheinlichkeitsschätzung. Die neue Methode ist auf rechtszensierte Daten ausgelegt.
- Für alle Modelle werden Calibration-Slope und Calibration-Intercept bestimmt. Außerdem wird die Kalibrierung durch eine Kalibrierungskurve graphisch dargestellt.
- Zur Bewertung der Prognosegüte werden zusätzlich der Brier-Score und der Concordance-Index betrachtet

Ergebnisse:

- Es kann weder bei der Cox-Regression noch beim RSF ein bereits existierendes Modell verwendet werden, beide werden auch auf den US-Daten neu trainiert.
- Die auf den US-Daten trainierten Modelle haben für die deutschen Daten eine schlechte Kalibrierung, es besteht also der Bedarf zur Rekalibrierung
- Durch die Rekalibrierung der Cox-Regression konnte die Kalibrierung der US-Modelle auf der deutschen Kohorte für alle betrachteten Zeitpunkte verbessert werden. Für die Jahre 5 bis 7 nach der Transplantation wies das rekalierte Modell sogar eine bessere Kalibrierung auf als das neu trainierte Modell.
- Mit der neu entwickelten Methode für den Random Survival Forest konnte eine Rekalibrierung des US-Modells auf die deutsche Kohorte erzielt werden. Allerdings konnte hier für die Jahre 1 und 2 nach der Transplantation nur der Calibration-Intercept verbessert werden, während sich der Calibration-Slope für diese Zeitpunkte leicht verschlechterte.
- Der auf den deutschen Daten neu trainierte Random Survival Forest war für die deutsche Kohorte für jeden Zeitpunkt besser kalibriert als das nicht rekalierte und das rekalierte US-Modell

Ausblick:

- Andere Schätzer für den Baseline-Hazard oder verschiedene Skalierungsfaktoren für die einzelnen Kovariaten verwenden
- Pruning-Algorithmus mit in die neue Rekalibrierungsmethode integrieren
- IPWC-Gewichtung in Kombination mit Rekalibrierungsmethoden, die nicht auf Daten mit Rechtszensierung ausgelegt sind, verwenden

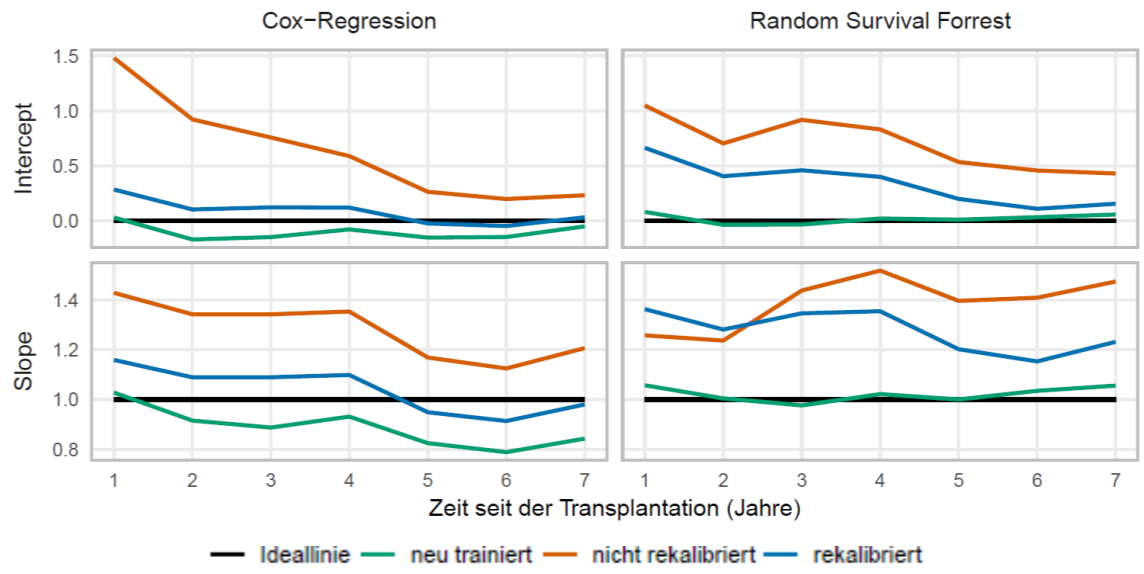
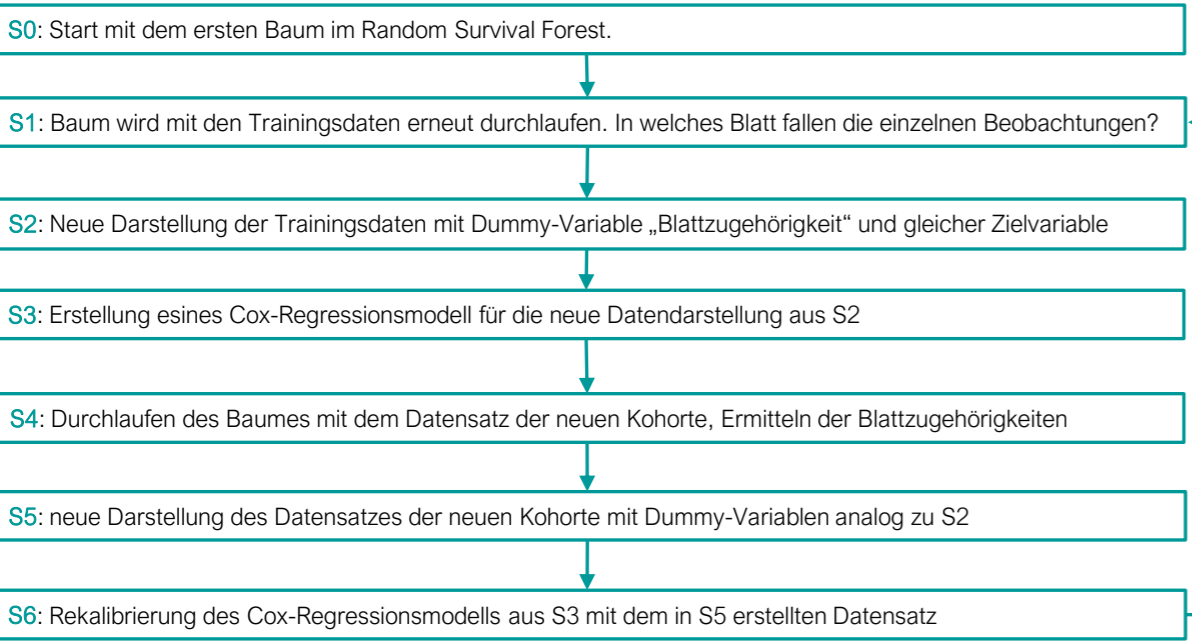


$$d_1 = \begin{cases} 1, & x_1 \leq c_1, x_2 \leq c_2 \\ 0, & \text{sonst} \end{cases}$$

$$d_2 = \begin{cases} 1, & x_1 \leq c_1, x_2 > c_2 \\ 0, & \text{sonst} \end{cases}$$

$$d_3 = \begin{cases} 1, & x_1 > c_1, x_3 \leq c_3 \\ 0, & \text{sonst} \end{cases}$$

$$d_4 = \begin{cases} 1, & x_1 > c_1, x_3 > c_3 \\ 0, & \text{sonst} \end{cases}$$



Referenzen:

[1] Panduranga Rao and Douglas Schaubel. A comprehensive risk quantification score for deceased donor kidneys: The kidney donor risk index. *Transplantation*, 88:231–236, 2009.

[2] D R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society*, 34:187–220, 1972.

[3] Hemant Ishwaran, Udaya B. Kogalur, Eugene H. Blackstone, and Michael S. Lauer. Random survival forests. *Annals of Applied Statistics*, 2:841–860, 9 2008.

[4] Hans C Van Houwelingen. Validation, calibration, revision and combination of prognostic survival models. *Statistics in Medicine*, 19:3401–3415, 2000.

[5] Theresa Dankowski and Andreas Ziegler. Calibrating random forests for probability estimation. *Statistics in Medicine*, 35:3949–3960, 9 2016.