

## **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 rekali­briert 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 rekali­brierten, rekali­brierten und auf den deutschen Daten neu trainierten Modelle verglichen.

Schlagworte: Klinische Vorhersagemodelle, Rekalibrierung, Kalibrierung, Diskriminierung, externe Validierung, Cox Proportional Hazards Model, Random Survival Forest

## **Abstract**

Calibration, along with the ability to discriminate, is an important aspect of clinical prediction models. If one uses an existing prediction model for a new cohort, it is always necessary to investigate how well this model is calibrated for the new cohort. Poorly calibrated risk estimates can lead to false expectations among patients and are therefore problematic for ethical reasons. In certain cases, it may however be advantageous to use an existing model rather than train a new one. Especially if the sample size in the new cohort is small, this can counteract the risk of overfitting. To do this, the old model must first be re-calibrated to achieve a satisfactory calibration. In survival analysis, well-studied re-calibration methods exist for Cox regression and other semi-parametric and parametric models. Machine learning methods specifically adapted to censored data are also increasingly used in the development of clinical prediction models. However, there are no known re-calibration methods for these methods so far. In this thesis, a method for re-calibrating the random survival forest is therefore developed and evaluated. For this purpose, prediction models for kidney transplantation, which were developed with data from the USA, are to be applied to a German cohort. A Cox regression model and a random survival forest are considered. First, the requirements for applying the US-models to the German cohort are investigated. Subsequently, the calibration of the US-models on the German data is examined and a re-calibration is performed. For the random survival forest, the method developed in this thesis is used for this purpose. Finally, the performance of the non-re-calibrated, re-calibrated and newly trained model on the German data is compared.

**Keywords:** clinical prediction models, re-calibration, calibration, discrimination, external validation, Cox proportional hazards model, random survival forest