Predicting time-to-event outcomes under different intervention strategies: methods and applications

Prosepe, I.

Summary

This dissertation focuses on developing and applying estimation methods to make "what if" predictions in clinical research. The primary aim of these methods is to estimate risks for time-to-event outcomes using observational data in a manner that supports valid prediction under different intervention strategies, potentially conditional on patient and treatment characteristics. To this end, we formulate the identifiability assumptions required to ensure the validity of these estimates.

Chapter 2 reviews Covid-19 prediction models to assess whether their analysis strategies align with their intended purposes. This chapter motivated the remainder of the thesis and provided context for the estimation methods developed later. The review revealed frequent mismatches: 64% of the papers recommended their model for decision-making, yet employed methods that do not support that aim. Moreover, 21% of the models lacked a clearly defined prediction estimand, leaving the meaning of their estimated risks ambiguous. These findings indicate that the use of estimation strategies inappropriate for the targeted prediction estimand is common. Using nationwide Dutch data on SARS-CoV-2-positive hospitalized patients in 2020, we illustrated how such misalignment can be misleading by under- or overestimating the intended risks.

Chapters 3 and **4** examine the concept of conditional survival benefit, defined as the difference in survival between two intervention strategies (with versus without treatment) over a specified time horizon for patients with given characteristics. This concept is explored in the context of the allocation of livers from deceased donors, where a ranking system based on expected benefit may be of particular interest.

Chapter 3 showcases the use of marginal structural models accommodating multiple baseline times and time scales to estimate survival benefit among *treated* patients over calendar time. This estimation method adjusts for time-varying confounding while enabling repeated prediction at different time points. Applying it to liver transplantation data from the United Network for Organ Sharing (UNOS) waiting list, we compared patients with and without hepatocellular carcinoma (HCC) and found the largest life-year gain was among those without HCC, highlighting potential disparities in access to transplantation between the two groups.

Chapter 4 introduces a novel estimation strategy for the repeated estimation of the conditional expected survival benefit for *all* patients on the transplant waiting list. Building on the existing methodology showcased in the previous chapter, we derived a modified reweighting strategy to target the entire waiting list rather than only treated patients. This extension required careful formulation of the estimands and identification assumptions as well as addressing additional complexities that do not arise when estimating survival benefit for the treated. Through a simulation study, we demonstrated that the proposed method outperforms simpler approaches that fail to account for time-dependent confounding or the need to integrate multiple time scales. Applying this methodology to a European dataset of patients on the liver transplant

waiting list, we showed how our predictions facilitate comparison between the current MELD-based allocation system and a hypothetical benefit-based prioritization rule.

Chapter 5 presents the development of a decision-support algorithm to aid decision making in the NICU (Neonatal Intensive Care Unit). For preterm infants with severe thrombocytopenia, we estimated the conditional 3-day risk of major bleeding or death based on whether a prophylactic platelet transfusion is (i) administered within 6 hours or (ii) withheld for 3 days. Model development used an international multicenter cohort (2017–2021) and combined landmarking with clone-censor-reweighting to address time-varying confounding and properly target the two intervention strategies. The model was then validated in a national multicenter cohort (2010–2014) using an extension of classic performance measures adapted for models targeting specific intervention strategies, showing good calibration.

Chapter 6 presents a novel estimation approach that combines multistate modeling with g-computation to estimate the causal effect of treatment delay in observational data with baseline confounding. This modeling approach relies on formal causal assumptions for identification and on modeling assumptions for estimation, which we outlined. Through a simulation study, we compared our proposed method with the clone–censor–reweight approach. While both are expected to yield asymptotically unbiased estimates of recovery probabilities under their respective assumptions, our model used the data more efficiently (smaller variance). We applied the method to estimate the effect of treatment delay in a cohort of couples with unexplained subfertility seeking intrauterine insemination.

Chapter 7 introduces the concept of interventional updating, an approach that incorporates external evidence (e.g., from clinical trials) to update predictions under changing intervention strategies. This method addresses the challenge of keeping clinical prediction models current as clinical guidelines evolve. Using electronic health record data from Covid-19 patients, we developed a 28-day mortality risk model with data from March–July 2020 and applied both standard and interventional updates to data from subsequent months. In our case study, interventional updating did not improve discrimination but yielded better calibration compared to standard model updating techniques.