Essays on Predicting Hospital Readmissions: Personalized Risk Estimation Using Multi-State Models and Machine Learning

Hospital readmission has long been recognized as a key indicator of healthcare quality and system efficiency. A hospital readmission often signals unresolved medical conditions, insufficient discharge planning, or inadequate post-discharge support, making it a vital measure of care coordination and continuity (e.g. (Chang & Tung, 2024; Kripalani et al., 2019). Within the healthcare industry, tracking and managing readmissions serves not only to assess clinical performance but also to identify systemic gaps requiring structural or process-level improvements including better capacity planning and resource allocation. High readmissions may indicate quality gaps that warrant structural or process improvements, reinforcing Donabedian model's emphasis on outcomes as the ultimate gauge of healthcare quality.

Given the economic burden of hospital readmissions, the Patient Protection and Affordable Care Act (ACA) instituted the Readmission Reduction Program that required the Centers for Medicare and Medicaid (CMS) to reduce reimbursements for hospitals with higher than predicted readmission rates (CMS; Kruger et al., 2019). In this context, penalties dictated by CMS motivated hospitals to identify better clinical and operations decisions for admitted patients to improve operational and financial outcomes and reduce 30-day unplanned readmissions.

Having established patient readmissions as a phenomenon that parallels core principles of operations and supply chain management, it is equally important to examine how these events can be systematically measured and predicted. Just as forecasting and demand planning are central to resilient supply chains, estimating readmissions is critical for healthcare systems seeking to anticipate patient flows, optimize resource allocation, and reduce avoidable costs. The literature on readmission estimation spans a wide range of approaches, from traditional statistical

risk models to more advanced machine learning and systems-based methods, each aiming to identify high-risk patients before they reenter the hospital. This transition from conceptualizing readmissions as a systems-level challenge to exploring the methods for quantifying and predicting them provides a necessary bridge between theory and actionable decision-making in healthcare operations.

In light of the literature and practice, this study aims to develop personalized prediction models that estimate the probability of 30-day unplanned hospital readmission through continuous-time models and machine learning (ML) models using patient-level and hospital-level data from the Nationwide Readmissions Database (NRD). To achieve this goal, our study will consist of three parts because this study will be based on "manuscript format." The first part focuses on conducting a systematic literature review that analyzes and synthesize 30-day unplanned readmission estimation models at acute care hospitals (ACHs). The second and third parts aim to develop personalized readmission estimation models. The second part includes a two-state continuous-time model that includes Cox Proportional Hazard model to incorporate patient-level covariates. The third part focuses on implementing concept drift, online learning algorithms, and online feature selection in readmission estimation using patient-level and hospital-level covariates.