Balancing the scale: Readmission risk at discharge vs inpatient unit congestion

Since the introduction of Hospital Readmissions Reduction Program (HRRP) and other incentives for reducing readmissions, hospitals have focused much of their efforts on reducing avoidable readmissions through post-discharge methods such as phone calls, home visits, and tele-medicine. As the post-discharge methods are approaching the limit of their capabilities, hospitals have begun to consider interventions during the hospital stay that can reduce a patient’s readmission risk. Chief among these interventions is the timing of discharge, which has been identified as important both in the academic literature and in industry, including several hospitals that we surveyed. While discharge timing has been clearly demonstrated as impactful, fewer analytical solutions exist in this space to help hospitals balance inpatient unit congestions with the inherent risks/benefits of shortening/extending length of stay.

**Background.** Frequent overloading of inpatient units contributes to emergency department (ED) overcrowding, denial of ICU admission, cancellation of elective surgeries, and higher risk of readmissions, among other consequences. When hospital units become congested, doctors frequently discharge existing patients early, as suggested by empirical work showing that discharge rates increase when units are at high occupancy (KC and Terwiesch 2017). This practice mitigates the impact of overcrowding in the ward by shifting the burden to the early discharge patients, who may experience increased risk of readmission, mortality, and other adverse outcomes (Oh et al. 2017). On the other hand, when occupancies are low,
hospitals may keep patients longer, which can have a positive impact on patient outcomes (Bartel et al. 2014). The balancing act between individual discharge risk and ward congestion has grown into a major stress point in the face of recent pressures to dramatically reduce readmissions (e.g. HRRP, Accountable Care Organization). Hospitals manage this tradeoff through ad-hoc early discharge practices that lack analytical decision support.

**Contributions.** In this work, we develop a practical decision support tool that can aid hospitals to make analytical discharge decisions. To ground this work in practice, we worked with a readmission-focused data-analytics company and their client hospital to develop and implement our research. We begin by addressing the gaps in the existing readmission prediction methodologies, which currently lack key capabilities to support our discharge decision framework, specifically: (1) excess zero count (most patients are not readmitted), (2) endogeneity of LOS (most existing models treat LOS as an exogenous variable, while LOS correlates with both patient severity and readmission risk), (3) lack of heterogeneity (the commonly used Cox proportional hazard model assumes the same baseline function for all patients). We develop a prediction model that integrates several methods to overcome these three main barriers.

We then integrate this prediction model into a discharge optimization that balances the tradeoff between inpatient unit congestion and individual discharge risk. This optimization takes the form of a large-scale Markov decision process that is intractable except in very small problem instances. We overcome the curse of dimensionality using a series of methods to reduce the impact of large action space and state space. First, we analyze the structure of the problem to show the
existence of discharge priorities among patient classes and that the optimal policy takes the form of a state-dependent threshold policy. Next, we approximate the cost-to-go function using the closed form solution of an approximating linear quadratic control problem. Finally, we leverage the structure of the optimal solution and the approximated cost-to-go function to design an effective decomposition method for iteratively calculating the discharge actions for each patient class.

From the analysis of the decomposition method, we are able to develop a ranking criterion of patients, which compares the full discharge risk trajectory as a function of length of stay. We show that this ranking criterion is not only strictly followed in the optimal policy, but also leads to an interpretable and easily implementable discharge policy for our partner hospitals. At the end of this talk, we discuss in details the ongoing implementation efforts of both the readmission prediction model and the discharge optimization tools.

References:
