Estimating the Utility of Using a Predictive Model for Optimizing Advanced Care Planning

Sehj Kashyap, BS1, Ken Jung1, PhD, Stephanie Harman2, MD, Margaret Smith2, MBA, Nigam H. Shah1, MBBS PhD, Ron Li2, MD
1Center for Biomedical Informatics Research, Stanford University; 2Dept of Medicine, Stanford University School of Medicine

Introduction
Predictive models created via Machine learning (ML) are being explored in healthcare to risk stratify patient populations and efficiently allocate clinical resources to the right patient at the right time. While researchers typically evaluate whether the model predictions are accurate, few evaluations assess whether a model delivers its intended clinical utility. We describe a pre-deployment assessment of clinical utility for a mortality prediction model that is intended to improve advanced care planning (ACP) for hospitalized patients.

Advanced care planning is integral to aligning patient care with patient wishes, but studies show that high-need patients often do not receive it or in a timely manner. Recently, predictive models created via machine learning have been proposed for improving advanced care planning and advanced illness management. A model that predicts a surrogate outcome of death within a given timeframe could be used for ensuring that those patients are prioritized for access to advanced care planning. Such a model may improve care by several mechanisms, including being more accurate, more consistent, and scalable than existing processes for identifying patients in need of advance care planning. We are working towards improving ACP delivery enabled by such predictive models at our institution. As part of the pre-deployment assessment, we completed a stakeholder and current state analysis of ACP delivery and identified that the number of hospitalizations in the prior year is a common heuristic that clinicians use for identifying patients for ACP. In order to quantify the potential clinical utility of our model, we compared the proposed machine learning model to the existing heuristic by asking if the model would detect a larger proportion of patients in need of ACP with more lead time than the heuristic.

Methods
We used a dataset of electronic health record (EHR) data from Stanford Hospital from January 2010 to December 2018, consisting of 116,067 admissions of 58,016 patients. We define admissions to be in high need for ACP if the patient dies within 12 months of the admission date. All admissions that occurred after 2017 and were seen by the medicine department were isolated for comparison between the machine learning model and the clinical heuristic for flagging patients for advanced care planning. A gradient boosted tree model using 73,085 features (representing patient treatments, diagnosis codes, and clinical encounters within the year preceding admission) was used to predict the patient’s risk for death within 1 year of the admission date. The clinical heuristic was the number admissions in the year prior to the current admission. In total there were 3,058 admissions (representing 2,152 patients) that were analysed, of which 1139 (37.2%) ended in death within the selected time-frame and were therefore considered high need for ACP. To explore how much time before death the model and heuristic afforded clinicians for ACP, we analysed a subset of patients (N = 458) who died in 2018 and had at least one medicine admission in the preceding one year.

Results
The model has significantly higher precision at all levels of recall than all possible cutoffs for the number of admissions clinical heuristic. At a threshold of three or more admissions (which was reported by our clinicians as their trigger for ACP referral), the heuristic detects only 21% of admissions that are high-need for ACP at a cost of screening 2.46 admissions for correctly finding every 1 admission in needs. In contrast, at the same recall, the model only prompts for screening of 1.08 admissions. For further analysis, we chose three or more admissions as the threshold of operation.
for the clinical heuristic. For the model, we selected the threshold at which the model would yield the highest recall while prompting fewer than 2 admissions to screen (denoted by the markers in Fig 1a).

![Figure 1](image.png)

**Figure 1. a)** Precision and recall at different thresholds (for the admission heuristic this translates to numbers of prior admissions ranging from 1 to 12). The red dot marks the threshold (>=3 admissions in prior year) that clinicians report using for initiating ACP, while the blue star represents the highest recall possible while keeping the number needed to screen to 2 (the capacity of the palliative team). The second graph shows of death cases (N=458 patients) who were detected and the earliest prior year admission that they are flagged. The final graph shows the difference in days that the model flags a patient for ACP versus when the clinical heuristic flags a patient in mutually detected cases (58.5 days, 95% CI 36.82 to 80.18; P<0.005; N=64).

In comparing detection of patients likely to benefit and the timing of detection, we see that the heuristic detects 71 (15.5%) patients who died while the model detects 392 (85.6%). The model co-detects 64 (90%) cases detected by the heuristic; thus, the heuristic only uniquely identifies 7 (1.8%) of all cases. On average the heuristic identifies patients 101 days (median = 44 days, SD = 110.5) before death, while the model flags patients 84 days (median = 40 days, SD = 95.8) before death. In cases that are co-detected by the model and heuristic, the model flags patients 58.5 days, approximately 2 months, before the heuristic.

**Discussion**

We compare a machine learning model for identifying patients likely to benefit from advanced care planning against an existing clinical heuristic for the same task. The model is significantly better at detecting patients in high-need for ACP, cuts the number of patients needed to screen by half and confers a lead-time advantage of almost 2 months over the heuristic. Yet, most patients are detected by the model less than three months before death. This is largely because only a quarter of the patients were admitted more than six months before death. Thus, patients may need to be identified outside of inpatient encounters in order to maximize lead time. We are currently working on models for this use case.

**References**