Measuring and Predicting Non-facetime Work to Inform Physician Compensation

**Background.** Primary care physicians (PCPs) have historically operated under a fee-for-service (FFS) system. In this system, PCP compensation is determined by the number of procedures or activities (e.g., office visits) conducted. However, FFS compensation has become financially unsustainable as healthcare costs continue to rise, so health systems are exploring population health approaches to compensation. Instead of receiving compensation for each activity conducted, PCPs would be compensated on the basis of managing a risk-adjusted patient panel.

The fundamental idea behind panel-based compensation is that PCPs managing panels requiring more work would receive higher compensation. The crux of this model is how the workloads of different panels are evaluated. Existing risk-adjustment methods make the critical assumption that a patient’s medical complexity is a direct indicator of the quantity of work induced on the patient’s PCP. With the advent of electronic health record (EHR) systems, however, we observe that the relationship between complexity and workload is more nuanced.

While PCP work until now has been measured using office visits, recent observational studies show that nearly half of PCP time is spent on EHR work (non-facetime work, or NFTW), which includes tasks such as post-visit documentation, staff messages, and patient communication. It is therefore crucial that risk-adjustment methods account for this NFTW.

The goals of this study are to 1) establish a methodology for measuring PCP work (facetime + NFTW), 2) investigate variability in the management of NFTW among PCPs in order to establish best practices, 3) explore the relationship between existing measures of patient complexity (risk scores) and workload, and 4) develop prediction models that use a patient’s characteristics to predict the PCP workload (facetime + NFTW) required to manage the patient.
Methods. Medical records of >10,000 patients from a primary care practice were analyzed including all electronic documentation generated about the patient. The developed analysis methods provide objective and replicable processes for measuring the quantity of workload (time) that any patient requires.

The total workload associated with individual patients is then broken down into work completed by different staff members (PCPs, nurses, medical assistants, secretaries, etc.). This enables us to evaluate how different PCPs manage the workload distribution amongst their assisting team members.

We then test the predictive power of current measures of patient complexity (risk scores) in terms of how well they predict future PCP workload. This is done by using the risk scores to predict PCP workload via machine learning models such as regularized linear regression (LASSO), classification and regression trees (CART), random forests, and neural networks. The performance of these models are evaluated using the metric of out-of-sample R-squared (OSR2).

Then, using EHR data, we generate ~600 characteristics that describe each patient according to their demographics, clinical history, and previous communication with the practice. These characteristics are used to train the same machine learning models to predict future PCP workload. We compare the performance of these models to the risk score models using OSR2.

Results. In line with the findings of small observational studies, our EHR analyses indicate that a tremendous amount of PCP time is spent on NFTW. The practice which manages ~10,000 patients generates ~740 messages and notes per workday, a quarter of which is done by PCPs.

When comparing NFTW between PCPs, we observe extreme workload differences. For example, two PCPs with similar patient panels (according to panel size and risk profile) generate vastly different quantities of NFTW – one PCP produces 2.5x more NFTW than the other PCP.
Furthermore, the distribution of work across staff levels also varies greatly between different PCP teams. Some PCPs manage 33% of their team’s NFTW while other PCPs manage 16% by allocating more NFTW to their team’s nurses.

In terms of the predictive modeling, current risk scores (e.g., publicly available scores developed by governmental agencies as well as proprietary scores developed by private companies) achieve OSR2 of ~0.07 when predicting PCP workload of patients. The models we developed using patients’ vast medical records achieve OSR2 of ~0.31 – providing more than a 4-fold increase in predictive power.

**Conclusions.** As PCP work increasingly involves usage of EHRs, it is crucial to quantify this workload and understand its drivers. While extensive efforts in medical education have been focused on establishing best practices for patient treatment during in-office visits, it is clear that the lack of best practices surrounding NFTW management represent significant opportunities for improvement. Furthermore, our findings challenge the assumption that a patient’s medical complexity is necessarily indicative of the work required to manage the patient. For example, a patient that suffers from chronic lower back pain, mood disorders, or fatigue would be considered by existing risk-scoring models as low-complexity even though our prediction models show that managing these conditions requires a substantial amount of work. This workload may be further amplified by patients’ personal issues (e.g., divorce or other stressors). Such life events are not captured by risk models although clinical intuition suggests they have significant impact on physician-patient interactions. Our prediction models leverage the data in EHRs to capture these effects as well.