Predicting Avoidable Health Care Utilization: Practical Considerations for Artificial Intelligence/Machine Learning Models in Population Health

Steven Lin, MD; Shreya Shah, MD; Amelia Sattler, MD; and Margaret Smith, MBA

In the United States, $1 in every $10 of total hospital expenditures is spent on potentially preventable conditions. Preventable hospitalizations and emergency department (ED) visits add up to $100 billion in costs to the US health care system each year, which is more money than the gross domestic product of 140 countries. Up to 13% of adult hospitalizations and 8% of pediatric hospitalizations are potentially preventable, and most of these preventable hospitalizations are due to poorly controlled chronic conditions.

One in 6 Medicare admissions is preventable. In 2017, Medicare admissions represented two-thirds of all potentially preventable hospitalizations and related health care costs. Preventable hospitalizations for diabetes among residents in the poorest communities are more than 400% higher than those in the wealthiest communities. These stark health disparities worsen as community-level income disparities widen.

Furthermore, there were more than 143 million ED visits in 2018, accounting for 12.5% of overall US health care costs. Approximately 1 in 5 adults use the ED at least once yearly, and the rate of potentially preventable ED visits has increased year over year for the past 2 decades. Studies have estimated rates of preventable ED visits to be as low as 3% and as high as 70%, a wide range given the lack of consensus on the definition of “preventable.” Regardless of the true figure, the problem is clear: the cost of managing primary care–treatable conditions is 10 to 12 times higher in the ED than in primary care, according to data from commercial payers. It takes an average of $2000 to manage a primary care–treatable condition in an ED compared with less than $200 in a primary care setting. Hospital facility fees increase the cost of an average ED visit by more than $1000, and laboratory and radiology services are 10 times more costly in the ED than in the office of a primary care physician (PCP).

The rise of quality-based reimbursement programs and alternative payment models incentivizes care delivery systems to improve population health and reduce health care costs. This transition from volume-based to value-based care includes a focus on ambulatory care–sensitive conditions, for which timely and effective primary care can lower the risk of an ED visit or hospitalization. Congestive heart failure is the costliest condition among preventable adult hospitalizations, followed by diabetes and chronic obstructive pulmonary disease. Asthma is the leading cause of preventable pediatric hospitalization. These and other chronic conditions, such as hypertension and depression, are the focus of many health systems and payers as they develop new methods to improve quality of care and reduce avoidable health care use. Studies of contemporary interventions to reduce ED visits and hospital readmissions have produced mixed evidence and revealed no silver bullets, reflecting both the complexity of the challenges as well as the need for new models and tools to address them.

Can modern artificial intelligence (AI) and machine learning (ML) predict...
avoidable health care utilization and reduce preventable hospitalizations and ED visits? The promise is there. The emergence of big data in health care and interoperability standards has catalyzed growth of predictive analytics, including recent collaborations between large technology companies and health systems.\textsuperscript{10,11} Although some studies have demonstrated AI/ML superiority over traditional non-AI/ML methods for predicting hospital and ED readmissions,\textsuperscript{12-16} few studies have looked at predicting all (including initial) hospitalizations and ED visits. A recent implementation of an AI/ML model across a network of 80 primary care centers outperformed previous approaches for identifying at-risk patients.\textsuperscript{17} More research is needed to evaluate real-world applications of these AI/ML models and their effects on preventable visits, quality of care, and total cost of care.

This is arguably one of the most exciting applications of AI/ML in medicine today. What are the practical considerations for translation and implementation that model developers might embrace in building AI/ML predictive models to improve population health?

**MODELS SHOULD NOT DEPEND SOLELY ON CLAIMS**

Existing tools are limited because they are largely retrospective and rely heavily on claims data. For example, a patient visits the ED or hospital; an insurance claim is processed, which takes up to 4-6 weeks; the health plan flags preventable visits and superusers (ie, patients with \geq 2 ED visits or 1 hospitalization in the past 6 months, as well as the 5% of patients who account for 50% of total spending); and then, finally, there may or may not be an intervention, with the health plan communicating with health systems and the PCP’s office, usually on a monthly to biannual basis (Figure). But this is simply too little, too late.
What we need are prospective, predictive models that do not depend on claims: in other words, models running on real-time electronic health record (EHR) data that can predict the risk of ED visits and hospitalizations for an entire patient population, delivering those insights to the PCP and primary care team, who can then execute interventions on a proactive and rolling, rather than on a retrospective and episodic, basis. Claims may provide additional data that augment the model, but the model itself should not depend solely on claims. Instead, the model should be powered by data elements illustrated in the Figure.

MODELS SHOULD PROVIDE PREDICTIONS ON EVERY PATIENT, NOT JUST THOSE WHO ENGAGE IN CARE
Most patients who visit the ED do not contact or see their PCP beforehand. Preventable ED visits are not a failure of PCPs recognizing that the patients in their office are deteriorating but rather are a failure of health systems to properly resource, structure, and equip primary care teams to identify patients at home who are sick and intervene before they end up in the ED or hospital. Therefore, models should provide situational awareness on every patient on a PCP’s panel all the time, not just for those who come into the clinic. Predictions should be delivered to primary care teams, not only PCPs, who are too busy seeing patients in the clinic and must rely on their teams to coordinate care for patients who require medical interventions. Ideally, predictive models should augment existing human-driven population health programs operating within value-based care initiatives.

MODELS SHOULD BE BUILT ON OUTPATIENT DATA
Most predictive models in development are based on inpatient, not outpatient, data. But inpatient care is just 4% of all care and does not reflect the type of care provided in outpatient settings. It seems obvious, but models for population health should be powered by outpatient data, of which primary care is the most robust and applicable, representing 52% of US physician office visits, more than all other specialties combined. Patients interact with PCPs often and in a variety of contexts (eg, acute, chronic, preventive care) that generate the broadest array of both clinical and biopsychosocial data. Unlike inpatient data, which are more dynamic and have shorter time horizons (measured in hours to days), data in outpatient settings, especially primary care, are more stable and span months to years, providing stronger predictions for events over longer time horizons.

The growing adoption of remote patient monitoring in outpatient settings will continue to shorten time horizons with continuous data from connected devices and wearables, enabling predictions for near-term outcomes. Outpatient data also include information about patterns of care delivery and systemic care gaps that affect hospitalization risk, such as fragmentation of care between primary and specialty care.

Although some AI/ML models are built to be implemented in outpatient care settings, we need more robust predictive models that integrate data from both outpatient and inpatient settings to guide interventions in primary care. Reducing preventable ED visits and hospitalizations requires structured, targeted outpatient interventions that occur consistently over time. Data used to create predictive models should align with the strategies and contexts in which these model outputs will be used to change outcomes.

MODELS SHOULD BE MINDFUL OF HEALTH EQUITY AND CONSIDER SOURCES OF DATA OUTSIDE OF EHRs
Models on EHR data alone may not be adequately representative or timely. We must remember that data in EHRs represent only those with access to care; but what about those who do not? In fact, uneven access to care among racial and ethnic minorities may lead models to underpredict their risks. Adding social determinants of health data, including neighborhood/environment, lifestyle behaviors/habits, language, transportation, income, financial strain, social
support, and education, to current prediction models can improve model accuracy for hospitalization, death, and costs of care. In addition, timely identification of health risks at the population level may be augmented by data in the public domain, such as Internet search data and social media data. For example, one of the most powerful predictors of outbreaks of infectious diseases is a cluster of individuals in a geographic region searching for similar symptoms online. Another example: One of the earliest predictors of depression and anxiety is how individuals behave on their social media. Therefore, model developers should include sources of data within and outside of EHRs while adhering to data privacy best practices. Current Health Insurance Portability and Accountability Act of 1996 laws are outdated and do not give adequate guidance for big data management in the era of AI/ML, highlighting a critical need for accountability from health systems and third parties with access to data. How can we incorporate EHR and public data in a privacy-protected way so that we can better care for the health of entire populations? That is a question that future research needs to address.

MODELS NEED TO BE CONNECTED TO AN EFFECTIVE PATIENT ENGAGEMENT PLATFORM

Even if the perfect model exists, connecting those predictions with clinical interventions requires an effective patient engagement platform. Patient portals tethered to EHRs seem to be the obvious choice, and their use has been associated with decreased ED visits and hospitalizations, increased quality of care for chronic conditions, and high patient satisfaction. However, portal drawbacks related to low patient adoption, health inequity, and provider burnout are unresolved. The individual delivering the patient outreach also matters. When a patient receives a communication from their health plan, the open rate of that message is only approximately 3%. That open rate increases dramatically if the communication is coming from someone the patient recognizes as part of their primary care team (30%-60%) and highest when it is from their PCP (60%-90%) (Stanford Medicine, unpublished data, 2021). Therefore, the best approach is to leverage the trusting relationship that patients already have with their PCP and primary care team, designing models with the end users in mind to transform model predictions into effective real-world interventions.

CONCLUSION

Preventable hospitalizations and ED visits are a serious source of human suffering and economic pain that AI/ML can address by predicting avoidable health care use. These models should not depend solely on claims data. They should provide situational awareness on every patient all the time, not just those who engage in care. They should be built on outpatient data. They should be mindful of health equity, consider sources of data outside of EHRs, and adhere to data privacy protections. Finally, they need to be connected to a robust patient engagement platform for care teams to operationalize their predictions into effective clinical interventions. Physicians and AI/ML model developers must work together to achieve the holy grail of modern population health: combining accurate predictions with effective interventions to engaged patients.

ACKNOWLEDGMENTS

We thank Ms Grace Hong for her research support for the Stanford Healthcare AI Applied Research Team.

Grant Support: This work was supported by the Division of Primary Care and Population Health, Department of Medicine, Stanford University School of Medicine.

Potential Competing Interests: The authors are principal investigators working with companies and nonprofit organizations through grants and sponsored research agreements administered by Stanford University. Current and previous collaborators include Amazon, American Academy of Family Physicians, American Board of Family Physicians, Center for Professionalism and Value in Health Care, Codex Health, DeepScribe, Eko Health, Google Health, Quadrant Technologies, Soap Health, Society of Teachers of Family Medicine, University of California San Francisco, and Verily.

Data Previously Presented: These data were presented at the 2021 Stanford Center for Artificial Intelligence in Medicine and Imaging Virtual Symposium.
Correspondence: Address to Steven Lin, MD, 211 Quarry Rd, Ste 405, MC 5985, Palo Alto, CA 94304 (stevenlin@stanford.edu).

REFERENCES


