MAYO CLINIC PROCEEDINGS

ORIGINAL ARTICLE

Predicting Primary Care Physician Burnout From Electronic Health Record Use Measures

Daniel Tawfik, MD, MS; Mohsen Bayati, PhD; Jessica Liu, PhD; Liem Nguyen; Amrita Sinha, MD; Thomas Kannampallil, PhD; Tait Shanafelt, MD; and Jochen Profit, MD, MPH

Abstract

Objective: To evaluate the ability of routinely collected electronic health record (EHR) use measures to predict clinical work units at increased risk of burnout and potentially most in need of targeted interventions.

Methods: In this observational study of primary care physicians, we compiled clinical workload and EHR efficiency measures, then linked these measures to 2 years of well-being surveys (using the Stanford Professional Fulfillment Index) conducted from April 1, 2019, through October 16, 2020. Physicians were grouped into training and confirmation data sets to develop predictive models for burnout. We used gradient boosting classifier and other prediction modeling algorithms to quantify the predictive performance by the area under the receiver operating characteristics curve (AUC).

Results: Of 278 invited physicians from across 60 clinics, 233 (84%) completed 396 surveys. Physicians were 67% women with a median age category of 45 to 49 years. Aggregate burnout score was in the high range (≥3.325/10) on 111 of 396 (28%) surveys. Gradient boosting classifier of EHR use measures to predict burnout achieved an AUC of 0.59 (95% CI, 0.48 to 0.77) and an area under the precision-recall curve of 0.29 (95% CI, 0.20 to 0.66). Other models' confirmation set AUCs ranged from 0.56 (random forest) to 0.66 (penalized linear regression followed by dichotomization). Among the most predictive features were physician age, team member contributions to notes, and orders placed with user-defined preferences. Clinic-level aggregate measures identified the top quartile of clinics with 56% sensitivity and 85% specificity.

Conclusion: In a sample of primary care physicians, routinely collected EHR use measures demonstrated limited ability to predict individual burnout and moderate ability to identify high-risk clinics.

 ■ Mayo Clin Proc. 2024;99(9):1411-1421

Electronic health record (EHR) adoption has facilitated improved quality monitoring, records archiving, and research capabilities in recent decades,[1-5] but with increased administrative and clerical burden among physicians. Outpatient physicians now spend up to twice as much time interacting with the EHR as they do with patients.[6-8] Concurrently, symptoms of burnout now affect more than 500,000 US physicians and cost the US health care system approximately $5.6 billion annually (in 2023 dollars).[9-11] Primary care physicians are consistently among the most affected[10-12]; almost 50% believe the amount of time spent on clerical tasks is unreasonable,[13] those with high EHR task load have higher likelihood of burnout,[14-16] and up to 75% believe the EHR contributes to their burnout.[17,18]

Burnout is currently exclusively assessed by surveys,[19] which create survey burden, carry the risk of response bias,[20] and may miss symptoms occurring between surveys. However, EHR metadata reflect the work environment and experiences of each

From Stanford University School of Medicine (D.T., J.L., T.S., J.P.), Stanford Graduate School of Business (M.B.), Stanford University School of Engineering (L.N.), and Stanford Medicine WellMD & WellPhD Center (T.S.), Stanford, CA; Harvard Medical School, Boston, MA (A.S.); and Washington University School of Medicine in St Louis, St Louis, MO (T.K.).

physician, ranging from highly granular audit logs of time-stamped user actions to monthly summaries of efficiency and proficiency with the EHR. These data may carry relevance for physician burnout arising from a complex landscape of predictors, as evidenced by observed associations between objective measures of In Basket, after hours work, and burnout.[21-23]

Predicting burnout through EHR use measures may mitigate the challenges of response bias by using data from the entire physician population of interest and facilitate ongoing assessment of risk without inducing survey fatigue. This study sought to generate a predictive model quantifying risk for primary care physician burnout based on routinely generated EHR use data.

METHODS

Study Design and Participants

This was a retrospective longitudinal cohort study of attending physicians in General Internal Medicine, Primary Care and Population Health, and General Pediatrics at a large academic center and its affiliated community primary care clinics. The study was approved by the Stanford institutional review board with waiver of informed consent.

Data Sources

The EHR use measures consisted of Epic EHR's Signal metrics—a summary of physician EHR activity—supplemented with clinical summary data extracted from the institution's clinical research database. Each health system conducted 2 routine well-being surveys of all physicians working more than 0.5 full-time equivalents in spring 2019 and fall 2020. These surveys contained the Stanford Professional Fulfillment Index (PFI), which includes as its burnout measures 4 items on emotional exhaustion and 6 items on interpersonal disengagement.[24,25]

Primary Outcome

The primary outcome was the presence of burnout symptoms, calculated from the average of all burnout measures from the PFI. The PFI employs a 5-item Likert scale of 0 to 4, which we multiplied by 2.5 to transform to a score of 0 to 10 for each individual. We considered burnout symptoms as a continuous variable along this scale as well as dichotomized, with a score of 3.325 or higher indicative of burnout symptoms, in line with the external validation results for the PFI.[26,27]

Feature Extraction and Engineering

We extracted initial candidate features from 412 Signal metrics, encompassing measures of EHR time—with subcategorization into In Basket, Orders, Notes & Letters, and Clinical Review—plus measures of workload and user-specific EHR customizations. We collated highly granular metrics to parent categories and de-duplicated redundant features, resulting in 228 Signal features. Included in these features are 2 vendor-calculated summary metrics: Provider Efficiency Score (a scaled score indicative of observed vs expected EHR time based on clinical workload) and Proficiency Score (a scaled summary score of the physician's use of available efficiency tools).

We then added approximations of the 7 core EHR use measures proposed by Sinsky and colleagues.[28] We also summarized the subjectivity and polarity of progress notes written by each physician during the study period; summed typographical errors per note (excluding medical terms, proper names, acronyms, and repeated occurrences of the same error within the same note); and identified notes with key terms suggestive of moral distress, interpersonal conflict, traumatic events, and end-of-life care.

We extracted these features for each of the 3 months leading up to each survey administration, plus 3-month averages and average change in feature value during the 3 months (normalized among providers in the same specialty and year). We additionally calculated averages for all other physicians in the same clinic (excluding the physician of interest) as team-based measures for the month leading up to the survey.

From the resultant pool of 1577 features, we favored a hybrid approach to feature selection incorporating both domain

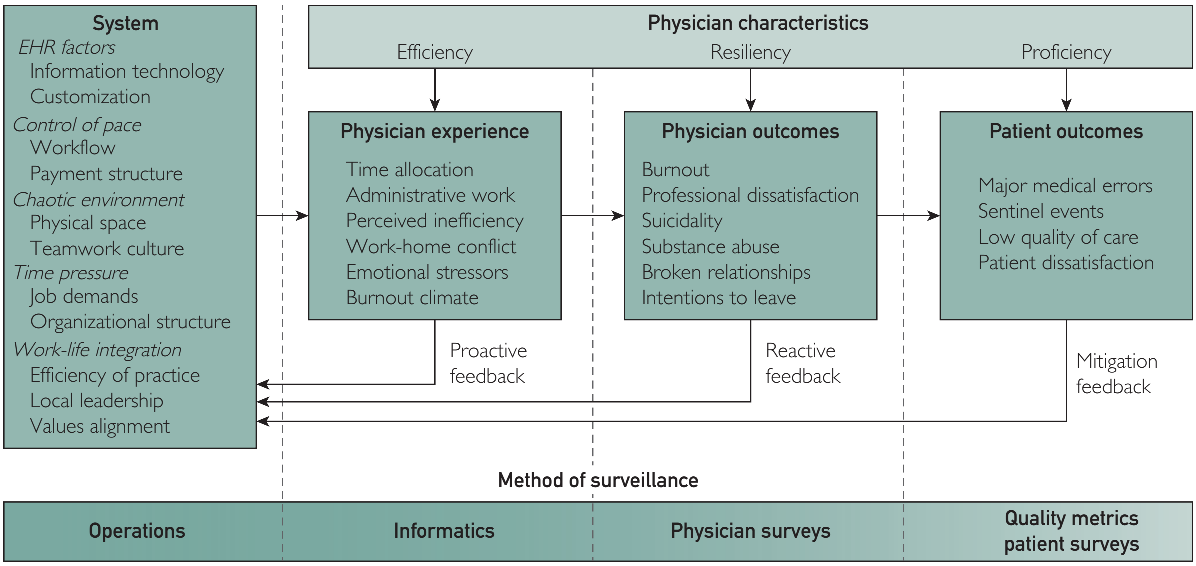

FIGURE 1. Conceptual framework outlining system factors, physician experience, and physician characteristics in relation to physician and patient outcomes along with traditional modes of feedback and opportunities for measurement. Nonexhaustive lists of examples are provided for system and physician factors as well as for outcomes. EHR, electronic health record.

knowledge and machine learning techniques. Guided by prior literature and expert opinion, we manually curated these characteristics to favor features with low collinearity, high cardinality, and known or potential associations with physician well-being from published literature[13-16,22,29-37] and our conceptual framework (Figure 1). This balanced approach, compared with relying solely on automated machine learning methods, helped refine our feature space in a context in which fully automated techniques might have been less effective. Prior literature finds that incorporating clinician expertise into feature generation can improve model performance.[38] This carefully curated process yielded a focused set of 191 candidate features, described in Supplemental Table 1 (available online at http://www.mayoclinicproceedings.org), that underwent further feature selection inherent to the approach applied for each machine learning method.

Statistical Approach

We randomly selected 20% of the physicians to hold aside as a confirmation sample. Among the remaining 80% training set, we divided physicians into 3 random folds, such that an individual physician could not appear on both sides of any split. Because nearly all candidate features were time varying, physicians with survey responses in both years were treated as unique observations within the same (training or confirmation) set.

We a priori selected gradient boosting classifier as our primary prediction algorithm of interest because of its capacity to capture nonlinear relationships and to incorporate variable selection from a large number of candidate variables. We used grid search to identify hyperparameters maximizing the area under the receiver operating characteristics curve (AUC) averaged among the 3 training folds. We then reported AUC and area under the precision-recall curve (AUCpr) for the confirmation set. Because of varying performance based on training-confirmation split, we repeated the entire process for 40 training-confirmation splits, reporting confirmation set performance for the best-performing model from training as well as the average of all 40 selected models. An AUC of 0.5 is comparable to the misclassification rate of a naive model; thus, we considered models for which these 40 iterations demonstrated a 95% CI above 0.5 to indicate at least marginally useful predictive power. Interpretation of AUCpr is similar, although the baseline threshold is dependent on the proportion of positive observations (0.28 in this population) rather than 0.5. Because relationships represented by gradient boosting models may be highly nonlinear, we also present results of the selected model using a diagram of SHapley Additive exPlanations (SHAP values).

We repeated the process for secondary algorithms: penalized logistic regression (more easily interpretable and able to achieve variable selection through penalization but limited in nonlinear relationships), distributed random forest (less prone to overfitting but often lower performance than gradient boosted models), k-nearest neighbors (nonparametric and intuitive, but performance may be affected by imbalanced data and outliers), and multilayer feedforward artificial neural network (potential to capture high-dimensionality relationships but limited interpretability and higher risk of overfitting). Grid search hyperparameters are outlined in Supplemental Table 2 (available online at http://www.mayoclinicproceedings.org).

In recognition that burnout may have climate-like effects in which individuals may influence team members[39] and with the goal of identifying work units at greatest need for intervention (as opposed to individuals), we evaluated the models' ability to predict clinic-level burnout. For clinic-level analyses, we excluded clinics with fewer than 3 survey respondents in a given year, to reduce small sample bias, and stratified by clinic to mitigate information leakage among physicians within the same clinic. We then calculated mean burnout score and mean predicted burnout score for each clinic. We ranked the clinics by mean burnout score, with a rank of 1 corresponding to the lowest mean of burnout scores and 36 corresponding to the highest mean of burnout scores. Clinics in the top 25% were considered to be clinics at high risk for burnout with the greatest need for intervention, and clinics in the lower 75% were considered standard risk. We then generated a confusion matrix and calculated sensitivity, specificity, and accuracy.

Sensitivity Analyses

In sensitivity analyses, we generated prediction models using the continuous burnout score, with performance measured directly by mean absolute error and indirectly by AUC (after conversion to probability of classification using a sigmoid function). We also generated prediction models using only respondents with burnout scores above or below 0.5 SD from the mean, effectively excluding respondents with intermediate burnout scores. We also generated prediction models separately for the emotional exhaustion and interpersonal disengagement scales of the PFI. For clinic-level analysis, we also evaluated clinic-level performance with models that were trained using clinic stratification to mitigate theoretical information leakage that may be due to burnout's climate-like effects.

We conducted analyses with Stata 17.0 software (StataCorp LLC) and Python 3.9.2 with TextBlob, sklearn, and h2oAutoML packages.

RESULTS

Summary and Demographic Characteristics

There were 396 complete surveys from 507 survey invitations (78% survey response rate), representing 233 of the 278 unique physicians (84% physician participation rate) from 60 primary care clinics in 2019 and 2020. Collectively, 163 of 278 eligible physicians (59%) completed 2 surveys, 70 of 278 (25%) completed 1 survey, and 45 of 278 (16%) completed 0 surveys. As shown in Table 1, most respondents were female (156/233 [67%]). The median age category was 45 to 49 years with 38 (16%) physicians; 106 (45%) physicians were younger than 45

TABLE 1. Participant and Clinic Characteristics

Physicians (N=233)	No. (%)
Female	156 (67)
Male	77 (33)
Age, years	
<35	41 (18)
35-39	31 (13)
40-44	34 (15)
45-49	38 (16)
50-54	24 (10)
55-59	27 (12)
60-64	24 (10)
≥65	14 (6)
Clinic type	
Academic	144 (52)
Community	134 (48)
Specialty	
General internal medicine	107 (46)
Family practice	91 (39)
General pediatrics	35 (15)
Clinics (N=60)	No. (%)
Clinic type	
Academic	42 (70)
Community	18 (30)
Clinic size	
<3 physicians	35 (58)
3-6 physicians	9 (15)
7-10 physicians	9 (15)
>10 physicians	7 (12)

years, and 89 (38%) physicians were 50 years or older.

Burnout scores ranged from 0 to 10 with a mean score of 2.5 (SD 1.8) and a median score of 2.5 (interquartile range, 1.0 to 3.5). With a burnout threshold of 3.325, 111 (28%) responses reported symptoms of burnout.

Performance of Gradient Boosting Classifier

The top-performing gradient boosting classifier model was selected on the basis of an average AUC of 0.68 across 3 folds of hyperparameter optimization. Chosen hyperparameters included a column sample rate of 0.8 (from candidate sample rates of 0.5 to 1), maximum depth of 10 (candidate depths 3 to 17), tree count of 23 (determined by early stopping up to a maximum of 100,000), and learning rate set at 0.1.

On the confirmation set, the classifier achieved an AUC of 0.59 (95% CI, 0.48 to 0.77) and an AUCpr of 0.29 (95% CI, 0.20 to 0.66), as shown in Table 2 and illustrated in the Supplemental Figure (available online at http://www.mayoclinicproceedings.org). Among the features most frequently included in decision trees, number of automated/administrative In Basket messages per day and average number of note characters written by other team members associated with higher burnout, whereas summary Provider Efficiency Score associated with lower burnout, with detailed directionality of relationships illustrated in Figure 2.

Performance of Other Algorithms

As shown in Table 2, the top-performing models from the secondary algorithms of interest achieved AUCs ranging from 0.56 (95% CI, 0.48 to 0.75) for random forest to 0.63 (95% CI, 0.46 to 0.77) for penalized logistic regression when tested on the confirmation set.

Clinic-Level Performance

Of the 60 primary care clinics, 36 (60%) had 3 or more respondents and were included in clinic-level analyses. Clinic-level predictions identified 5 of the 9 clinics with the top 25% of burnout scores (sensitivity, 56%) and 23 of the 27 standard-risk clinics (specificity, 85%), as illustrated in Figure 3. On classifying clinics into the top and bottom deciles, sensitivity was 50% (2 of 4 classified correctly) for both. The mean absolute error of clinic ranking was 5.5.

Sensitivity Analyses

When modeled as a regression task predicting linear burnout score, the best-performing models showed mean absolute errors ranging from 1.56 (95% CI, 1.38 to 1.73) for penalized linear regression to 1.71 (1.45-1.97) for k-nearest neighbors. The best-performing random forest model had an AUC of 0.66 for dichotomized predictions, outperforming the top AUC of 0.63 previously noted for classification models, which could suggest that the regression

TABLE 2. Model Performance for Binary Classification Models and Regression Models

Classification models	Model type	Confirmation set performance AUC (95% CI)	AUCpr (95% CI)	Log loss (95% CI)
	Gradient boosted classifier	0.59 (0.48-0.77)	0.29 (0.20-0.66)	0.63 (0.41-0.86)
	Distributed random forest	0.56 (0.48-0.75)	0.37 (0.22-0.62)	0.65 (0.42-0.82)
	Logistic regression	0.63 (0.46-0.77)	0.45 (0.19-0.63)	0.54 (0.49-0.70)
	K-nearest neighbors	0.58 (0.42-0.73)	0.34 (0.18-0.57)	0.60 (0.04-3.05)
	Neural network	0.58 (0.46-0.72)	0.37 (0.16-0.60)	1.73 (1.13-4.98)

		Confirmation set performance (regression)		Confirmation set performance (dichotomized)		
Regression models	Model type	MAE	RMSE	AUC	AUCpr	Log loss
	Gradient boosted regressor	1.62 (1.42-1.82)	1.94 (1.71-2.17)	0.54 (0.36-0.71)	0.31 (0.10-0.53)	0.63 (0.55-0.72)
	Distributed random forest	1.59 (1.40-1.78)	1.89 (1.66-2.12)	0.66 (0.51-0.81)	0.50 (0.29-0.70)	0.81 (0.72-0.91)
	Linear regression	1.56 (1.38-1.73)	1.85 (1.66-2.04)	0.56 (0.42-0.70)	0.48 (0.27-0.68)	0.68 (0.60-0.76)
	K-nearest neighbors	1.71 (1.45-1.97)	4.41 (3.23-5.59)	0.37 (0.21-0.53)	0.24 (0.02-0.47)	0.70 (0.63-0.78)
	Neural network	1.59 (1.30-1.89)	2.07 (1.72-2.42)	0.60 (0.42-0.78)	0.67 (0.47-0.87)	3.34 (3.21-3.47)

AUC, area under the receiver operating characteristics curve; AUCpr, area under the precision-recall curve; MAE, mean absolute error; RMSE, root mean squared error.

models profited from the richer information inherent in the numeric burnout score. Models predicting only the emotional exhaustion or interpersonal disengagement score showed AUCs of 0.57 (0.52 to 0.70) and 0.55 (0.51 to 0.60), respectively.

When responses within 0.5 SD of the mean were excluded, the best-performing model was a gradient boosting classifier model with an AUC of 0.61 on the confirmation set. When clinic-level group stratification was used, the clinic-level performance showed a sensitivity of 44% and specificity of 81% for identifying clinics in the top quartile of burnout scores.

DISCUSSION

In this longitudinal observational study of primary care physicians, we found that routinely collected EHR use measures provide limited ability to predict individual physicians at risk of experiencing burnout but had some ability to predict clinics with physicians at higher risk for burnout. Physician age, team-based note composition, order entry, and In Basket notifications were among the most predictive features. Although we identified several important predictors of burnout in this study, even the best-performing models demonstrated lackluster performance, highlighting the limitations of exclusively using EHR use measures as actionable predictors of burnout. Future work will be needed to determine whether additional features or more granular time-series analysis may improve performance or whether improved clinic-level aggregated predictions are accurate enough to be actionable.

Our findings build on similar findings of Lou et al[40] and Liu et al,[41] which reported limited predictive power to identify trainee physicians at risk of burnout by EHR use measures, in smaller samples of physicians with more frequent (monthly) burnout assessments. Although Lou et al used investigator-derived measures from raw event logs and Liu et al used a deep learning framework directly applied to raw event logs, the resultant model performances are comparable. These consistent findings lend support to the hypothesis that burnout is an individualized phenomenon affected by an array of work factors, personal factors, and the intersection of these domains, only a portion of which are captured through EHR use measures.

Physician age was among the most predictive features for burnout, with midcareer physicians at the highest risk for burnout,

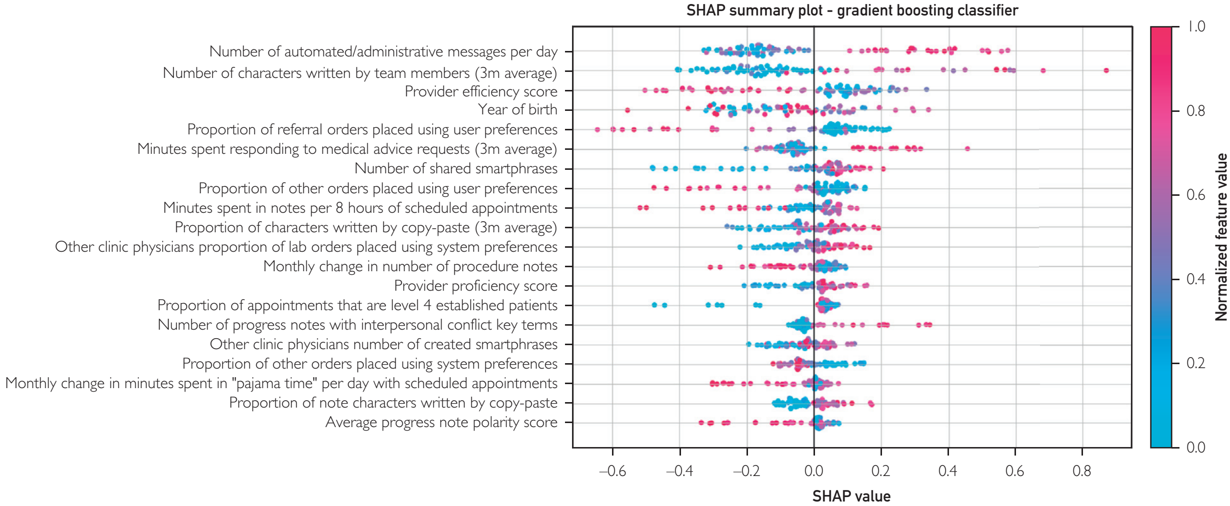

FIGURE 2. SHapley Additive exPlanations (SHAP) summary plot for selected gradient boosting classifier model, for top 20 features. Feature instances are color coded such that blue indicates low numeric values and red indicates high numeric values. Positive SHAP values (to the right) indicate higher predicted burnout score. Negative SHAP values (to the left) indicate lower predicted burnout score.

in line with prior findings.[42] Age may correlate with unmeasured contributors to burnout, including career stage, administrative or academic responsibilities, family and childcare needs, and financial burdens. Despite gender's role as a frequently considered predictor of work-life integration and burnout,[43-45] it showed very little importance among the full feature set, corroborating recent findings suggesting that there is little or nothing intrinsic about gender as a predictor of burnout that is not captured in the other included features, including EHR measures that have shown higher burden among women physicians.[43,45-48]

Team member contributions to progress notes also demonstrated importance for predicting burnout, with higher team contributions corresponding to higher risk of burnout. This finding merits further investigation but at a minimum suggests that caution is needed in transitioning toward a team-based care model to discern which shared tasks reduce work burden and which may exacerbate it.[49-53]

The EHR In Basket burden is an often-cited source of frustration for physicians,[13,22,36,54] a finding supported by our study as well. Remote, asynchronous communication has rapidly expanded in

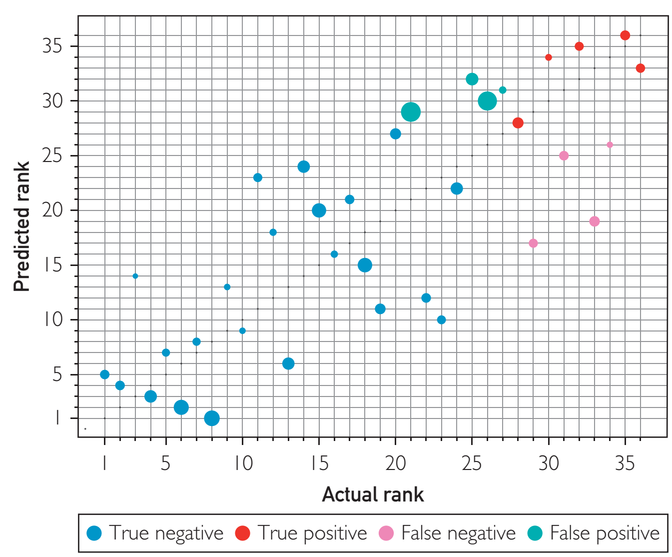

FIGURE 3. Predicted rank vs actual rank based on average burnout score for 36 clinics. Larger rank indicates higher average burnout score. The top 25% of clinics with highest burnout scores are shown in red and pink; the predicted top 25% are shown in red and teal. Marker sizes correspond to number of physicians employed in each clinic.

recent years, catalyzed by the COVID-19 pandemic,[55,56] but does not yet have consistent established infrastructure to efficiently route messages, to protect time spent on In Basket activities, and to allow compensation for professional services rendered on the platform. The resultant increased effort and decreased reward may contribute to burnout. In addition, the concomitant reduction in face-to-face contact with team members and patients may reduce professional fulfillment.[57,58]

Overall, the limited ability of EHR measures to predict physician burnout lies in contrast to the strong subjective relationships between EHR use and burnout previously identified.[14,17,30,59-63] It is possible that even if these objective features accurately capture the work done in the EHR, they may not directly correspond to the subjective experience of the work. Different individuals may interact with the EHR in such a way that generates similar measures but with different levels of frustration, which is strongly correlated with burnout.[29,34,64,65] Factors outside of the EHR are likely to be major contributors to burnout, with studies finding a minority of physician burnout variability to associate with EHR use or perceived EHR usability.[14,29,66]

Clinic-level predictions showed improved performance over individual predictions, lending further support to the potential of using burnout prediction tools as a trigger for system intervention at the work unit level rather than an individual clinician level. This work unit level approach avoids the potential ethical issues related to privacy if organizations use EHR metrics in an attempt to diagnose burnout in individuals—particularly considering that this misguided approach may increase the temptation to treat burnout as an individual failure rather than as a system failure. Treating burnout as an individual failure within an organization may promote unwarranted stigma, limit accuracy and completeness of survey responses if there are fears of repercussions, or promote “gaming” of EHR metrics perceived to promote a diagnosis of burnout.

This study has several limitations. Although we included a diverse sample of academic and community physicians and achieved a favorable survey response rate, this population may not be representative of all primary care physicians and cannot be extrapolated to other specialties or provider types as the workflows and experiences of physicians may vary widely. Although we included a large number of candidate measures, many are vendor derived and may not be transferable to other EHR vendors. There are many unmeasured EHR factors not captured in this feature set, and EHR use measures do not capture other work environment characteristics that may drive burnout, including team dynamics, communication outside of the EHR, leader behavior, values alignment, self-valuation, work-life integration, and mistreatment experiences.[27,44,67-71] Data collection for this study was completed in the early phase of the COVID-19 pandemic response, and further evaluation will be needed to identify novel features relevant to workflow changes implemented after this study.

CONCLUSION

Electronic health record measures appear to have limited discriminatory ability for identifying individual physicians at risk for burnout. In contrast, aggregate measures of these EHR measures at the work unit level show promise for identifying clinics that place individuals at high risk. Future work may improve predictions to provide actionable results. If the predictive value of these measures can be improved with inclusion of additional characteristics, such unit-level measures may help organizations identify clinics to prioritize for system-level interventions.

POTENTIAL COMPETING INTERESTS

Dr Kannampallil reports grants from NIH/NIA, grants from NIH/NLM, grants from AHRQ, grants from NIH/NCATS, personal fees from Elsevier, personal fees from Pfizer, and personal fees from Springer, outside the submitted work. Dr Shanafelt is co-inventor of the Well-being Index instruments

(Physician Well-being Index, Nurse Well-being Index, Medical Student Well-being Index, the Well-being Index) and the Mayo Leadership Index. Mayo Clinic holds the copyright for these instruments and has licensed them for use outside of Mayo Clinic. Dr Shanafelt receives a portion of any royalties. As an expert on the well-being of health care professionals, Dr Shanafelt frequently gives grand rounds/keynote lecture presentations and provides advising for health care organizations. He receives honoraria for some of these activities. Given his role as Section Editor, Dr Shanafelt had no involvement in the peer review of this article and has no access to information regarding its peer review.

ACKNOWLEDGMENTS

This research used data provided by STARR, "STAnford medicine Research data Repository," a clinical data warehouse containing live Epic data from Stanford Health Care, the Stanford Children's Hospital, the University Healthcare Alliance, and Packard Children's Health Alliance clinics and other auxiliary data from hospital applications, such as radiology PACS. STARR platform is developed and operated by Stanford Medicine Research Technology team and is made possible by Stanford School of Medicine Research Office.

SUPPLEMENTAL ONLINE MATERIAL

Supplemental material can be found online at http://www.mayoclinicproceedings.org. Supplemental material attached to journal articles has not been edited, and the authors take responsibility for the accuracy of all data.

Abbreviations and Acronyms: **AUC,** area under the receiver operating characteristics curve; **AUCpr,** area under the precision-recall curve; **EHR,** electronic health record; **PFI,** Professional Fulfillment Index

Grant Support: This research was supported by research grants from the Agency for Healthcare Research and Quality (K08 HS027837, PI: Tawfik), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD084679, PI: Profit), and the American Medical Association's Practice Transformation Initiative (PI: Shanafelt).

Data Previously Presented: Preliminary findings of this study were presented at the International Conference on Physician Health, October 13-15, 2022, Orlando, Florida.

Correspondence: Address to Daniel Tawfik, MD, MS, 770 Welch Rd, Ste 435, Palo Alto, CA 94304 (dtawfik@stanford.edu).

ORCID

Daniel Tawfik: https://orcid.org/0000-0003-2901-2737

REFERENCES

1. Washington V, DeSalvo K, Mostashari F, Blumenthal D. The HITECH era and the path forward. *N Engl J Med*. 2017; 377(10):904-906.
2. Martin SA, Sinsky CA. The map is not the territory: medical records and 21st century practice. *Lancet*. 2016;388:2053-2056.
3. Berenson RA, Basch P, Sussex A. Revisiting E&M visit guidelines—a missing piece of payment reform. *N Engl J Med*. 2011;364(20):1892-1895.
4. Halamka JD, Tripathi M. The HITECH era in retrospect. *N Engl J Med*. 2017;377(10):907-909.
5. Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: are we ignoring the real cause? *Ann Intern Med*. 2018;169(1):50-51.
6. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med*. 2016;165(11):753-760.
7. Tai-Seale M, Olson CW, Li J, et al. Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. *Health Aff (Millwood)*. 2017; 36(4):655-662.
8. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med*. 2017;15(5):419-426.
9. Han S, Shanafelt TD, Sinsky CA, et al. Estimating the attributable cost of physician burnout in the United States. *Ann Intern Med*. 2019;170(11):784-790.
10. Shanafelt TD, West CP, Sinsky C, et al. Changes in burnout and satisfaction with work-life integration in physicians and the general US working population between 2011 and 2020. *Mayo Clin Proc*. 2022;97(3):491-506.
11. Shanafelt TD, West CP, Dyrbye LN, et al. Changes in burnout and satisfaction with work-life integration in physicians during the first 2 years of the COVID-19 pandemic. *Mayo Clin Proc*. 2022;97(12):2248-2258.
12. Shanafelt TD, West CP, Sinsky C, et al. Changes in burnout and satisfaction with work-life integration in physicians and the general US working population between 2011 and 2017. *Mayo Clin Proc*. 2019;94(9):1681-1694.
13. Tran B, Lenhart A, Ross R, Dorr DA. Burnout and EHR use among academic primary care physicians with varied clinical workloads. *AMIA Jt Summits Transl Sci Proc*. 2019;2019:136-144.
14. Shanafelt TD, Dyrbye LN, Sinsky C, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. *Mayo Clin Proc*. 2016;91(7):836-848.
15. Babbott S, Manwell LB, Brown R, et al. Electronic medical records and physician stress in primary care: results from the MEMO Study. *J Am Med Inform Assoc*. 2014;21(e1):e100-e116.
16. Tawfik DS, Phibbs CS, Sexton JB, et al. Factors associated with provider burnout in the NICU. *Pediatrics*. 2017;139(5): e20164134.

17. Robertson SL, Robinson MD, Reid A. Electronic health record effects on work-life balance and burnout within the I^3 population collaborative. *J Grad Med Educ*. 2017;9(4):479-484.
18. Poll TH. How doctors feel about electronic health records. Accessed September 11, 2018. http://med.stanford.edu/content/dam/sm/ehr/documents/EHR-Poll-Presentation.pdf
19. Maslach C, Jackson SE. *Maslach Burnout Inventory*. Consulting Psychologists Press; 1981.
20. Rotenstein LS, Torre M, Ramos MA, et al. Prevalence of burnout among physicians: a systematic review. *JAMA*. 2018;320(11):1131-1150.
21. Kurbatov V, Shaughnessy M, Baratta V, et al. Application of advanced bioinformatics to understand and predict burnout among surgical trainees. *J Surg Educ*. 2020;77(3):499-507.
22. Adler-Milstein J, Zhao W, Willard-Grace R, Knox M, Grumbach K. Electronic health records and burnout: time spent on the electronic health record after hours and message volume associated with exhaustion but not with cynicism among primary care clinicians. *J Am Med Inform Assoc*. 2020;27(4):531-538.
23. Dyrbye LN, Gordon J, O'Horo J, et al. Relationships between EHR-based audit log data and physician burnout and clinical practice process measures. *Mayo Clin Proc*. 2023;98(3):398-409.
24. Brady KJ, Ni P, Carlasare L, et al. Establishing crosswalks between common measures of burnout in US physicians. *J Gen Intern Med*. 2022;37(4):777-784.
25. Trockel M, Bohman B, Lesure E, et al. A brief instrument to assess both burnout and professional fulfillment in physicians: reliability and validity, including correlation with self-reported medical errors, in a sample of resident and practicing physicians. *Acad Psychiatry*. 2018;42(1):11-24.
26. Shanafelt TD, Makowski MS, Wang H, et al. Association of burnout, professional fulfillment, and self-care practices of physician leaders with their independently rated leadership effectiveness. *JAMA Netw Open*. 2020;3(6):e207961. Published correction appears in *JAMA Netw Open*. 2020;3(9):e2021812.
27. Shanafelt TD, Wang H, Leonard M, et al. Assessment of the association of leadership behaviors of supervising physicians with personal-organizational values alignment among staff physicians. *JAMA Netw Open*. 2021;4(2):e2035622.
28. Sinsky CA, Rule A, Cohen G, et al. Metrics for assessing physician activity using electronic health record log data. *J Am Med Inform Assoc*. 2020;27(4):639-643.
29. Melnick ER, Dyrbye LN, Sinsky CA, et al. The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clin Proc*. 2020;95(3):476-487.
30. Eschenroeder HC, Manzione LC, Adler-Milstein J, et al. Associations of physician burnout with organizational electronic health record support and after-hours charting. *J Am Med Inform Assoc*. 2021;28(5):960-966.
31. Perez HR, Beyrouty M, Bennett K, et al. Chaos in the clinic: characteristics and consequences of practices perceived as chaotic. *J Healthc Qual*. 2017;39(1):43-53.
32. Linzer M, Poplau S, Grossman E, et al. A cluster randomized trial of interventions to improve work conditions and clinician burnout in primary care: results from the Healthy Work Place (HWP) study. *J Gen Intern Med*. 2015;30(8):1105-1111.
33. Linzer M, Poplau S, Babbott S, et al. Worklife and wellness in academic general internal medicine: results from a national survey. Sep. *J Gen Intern Med*. 2016;31(9):1004-1010
34. Gardner RL, Cooper E, Haskell J, et al. Physician stress and burnout: the impact of health information technology. *J Am Med Inform Assoc*. 2019;26(2):106-114.
35. Kroth PJ, Morioka-Douglas N, Veres S, et al. Association of electronic health record design and use factors with clinician stress and burnout. *JAMA Netw Open*. 2019;2(8):e199609.
36. Hilliard RW, Haskell J, Gardner RL. Are specific elements of electronic health record use associated with clinician burnout more than others? *J Am Med Inform Assoc*. 2020;27(9):1401-1410.
37. Marckini DN, Samuel BP, Parker JL, Cook SC. Electronic health record associated stress: a survey study of adult congenital heart disease specialists. *Congenit Heart Dis*. 2019;14(3):356-361.
38. Vallon JJ, Panjwani N, Ling X, et al. Patient-level clinical expertise enhances prostate cancer recurrence predictions with machine learning. Preprint. Posted online March 23, 2022. *medRxiv*. 22272635. https://doi.org/10.1101/2022.03.22.22272635.
39. Profit J, Sharek PJ, Amspoker AB, et al. Burnout in the NICU setting and its relation to safety culture. *BMJ Qual Saf*. 2014;23(10):806-813.
40. Lou SS, Liu H, Warner BC, Harford D, Lu C, Kannampallil T. Predicting physician burnout using clinical activity logs: model performance and lessons learned. *J Biomed Inform*. 2022;127:104015.
41. Liu H, Lou SS, Warner BC, Harford DR, Kannampallil T, Lu C. HiPAL: a deep framework for physician burnout prediction using activity logs in electronic health records. Presented at: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining; August 14-18, 2022; Washington, DC. Accessed August 28, 2023. https://doi.org/10.1145/3534678.3539056.
42. Dyrbye LN, Varkey P, Boone SL, Satele DV, Sloan JA, Shanafelt TD. Physician satisfaction and burnout at different career stages. *Mayo Clin Proc*. 2013;88(12):1358-1367.
43. Marshall AL, Dyrbye LN, Shanafelt TD, et al. Disparities in burnout and satisfaction with work-life integration in U.S. physicians by gender and practice setting. *Acad Med*. 2020;95(9):1435-1443.
44. Tawfik DS, Shanafelt TD, Dyrbye LN, et al. Personal and professional factors associated with work-life integration among US physicians. *JAMA Netw Open*. 2021;4(5):e2111575.
45. Rotenstein L, Harry E, Wickner P, et al. Contributors to gender differences in burnout and professional fulfillment: a survey of physician faculty. *Jt Comm J Qual Patient Saf*. 2021;47(11):723-730.
46. Rotenstein LS, Apathy NC, Bates DW, Landon B. Variation in support for documentation among primary care physicians by gender. Published online September 12. *J Am Board Fam Med*. 2022. https://doi.org/10.3122/jabfm.2022.AP.220071
47. Rotenstein L, Holmgren AJ. COVID exacerbated the gender disparity in physician electronic health record inbox burden. *J Am Med Inform Assoc*. 2023;30(10):1720-1724.
48. Rittenberg E, Liebman JB, Rexrode KM. Primary care physician gender and electronic health record workload. *J Gen Intern Med*. 2022;37(13):3295-3301.
49. Anderson P, Halley MD. A new approach to making your doctor-nurse team more productive. *Fam Pract Manag*. 2008;15(7):35-40.
50. Shaw JG, Winget M, Brown-Johnson C, et al. Primary Care 2.0: a prospective evaluation of a novel model of advanced team care with expanded medical assistant support. *Ann Fam Med*. 2021;19(5):411-418.
51. Sinsky CA, Bodenheimer T. Powering-up primary care teams: advanced team care with in-room support. *Ann Fam Med*. 2019;17(4):367-371.
52. Sinsky CA, Sinsky TA, Althaus D, Tranel J, Thiltgen M. 'Core teams': nurse-physician partnerships provide patient-centered care at an Iowa practice. *Health Aff (Millwood)*. 2010;29(5):966-968.
53. Brown-Johnson CG, Chan GK, Winget M, et al. Primary Care 2.0: design of a transformational team-based practice model to meet the quadruple aim. *Am J Med Qual*. 2018;34(4):339-347.
54. Tai-Seale M, Dillon EC, Yang Y, et al. Physicians' well-being linked to in-basket messages generated by algorithms in electronic health records. *Health Aff (Millwood)*. 2019;38(7):1073-1078.
55. Holmgren AJ, Downing NL, Tang M, Sharp C, Longhurst C, Huckman RS. Assessing the impact of the COVID-19 pandemic on clinician ambulatory electronic health record

use. *J Am Med Inform Assoc*. 2022;29(3):453-460. Published correction appears in J Am Med Inform Assoc. 2022;29(4):749.
56. Hansen MA, Chen R, Hirth J, Langabeer J, Zoorob R. Impact of COVID-19 lockdown on patient-provider electronic communications. Published online January 19. *J Telemed Telecare.* 2023. https://doi.org/10.1177/1357633x221146810
57. DePuccio MJ, Sullivan EE, Breton M, McKinstry D, Gaughan AA, McAlearney AS. The impact of COVID-19 on primary care teamwork: a qualitative study in two states. *J Gen Intern Med*. 2022;37(8):2003-2008.
58. Aung E, Pasanen L, LeGautier R, McLachlan SA, Collins A, Philip J. The role of telehealth in oncology care: a qualitative exploration of patient and clinician perspectives. *Eur J Cancer Care (Engl)*. 2022;31(2):e13563.
59. Yan Q, Jiang Z, Harbin Z, Tolbert PH, Davies MG. Exploring the relationship between electronic health records and provider burnout: a systematic review. *J Am Med Inform Assoc*. 2021; 28(5):1009-1021.
60. Melnick ER, West CP, Nath B, et al. The association between perceived electronic health record usability and professional burnout among US nurses. *J Am Med Inform Assoc*. 2021; 28(8):1632-1641.
61. Nguyen OT, Jenkins NJ, Khanna N, et al. A systematic review of contributing factors of and solutions to electronic health record—related impacts on physician well-being. *J Am Med Inform Assoc*. 2021;28(5):974-984.
62. Olson K, Sinsky C, Rinne ST, et al. Cross-sectional survey of workplace stressors associated with physician burnout measured by the Mini-Z and the Maslach Burnout Inventory. *Stress Health*. 2019;35(2):157-175.
63. Koranne R, Williams ES, Poplau S, et al. Reducing burnout and enhancing work engagement among clinicians: the Minnesota experience. *Health Care Manage Rev*. 2022; 47(1):49-57.
64. Tawfik DS, Sinha A, Bayati M, et al. Frustration with technology and its relation to emotional exhaustion among health care workers: cross-sectional observational study. *J Med Internet Res*. 2021;23(7):e26817.
65. Peccoralo LA, Kaplan CA, Pietrzak RH, Charney DS, Ripp JA. The impact of time spent on the electronic health record after work and of clerical work on burnout among clinical faculty. *J Am Med Inform Assoc*. 2021;28(5):938-947.
66. McPeek-Hinz E, Boazak M, Sexton JB, et al. Clinician burnout associated with sex, clinician type, work culture, and use of electronic health records. *JAMA Netw Open*. 2021;4(4): e215686.
67. Shanafelt TD, Gorringe G, Menaker R, et al. Impact of organizational leadership on physician burnout and satisfaction. *Mayo Clin Proc*. 2015;90(4):432-440.
68. Dyrbye LN, Major-Elechi B, Hays JT, Fraser CH, Buskirk SJ, West CP. Physicians' ratings of their supervisor's leadership behaviors and their subsequent burnout and satisfaction: a longitudinal study. *Mayo Clin Proc*. 2021;96(10):2598-2605.
69. Trockel M, Sinsky C, West CP, et al. Self-valuation challenges in the culture and practice of medicine and physician well-being. *Mayo Clin Proc*. 2021;96(8):2123-2132.
70. Dyrbye LN, West CP, Sinsky CA, et al. Physicians' experiences with mistreatment and discrimination by patients, families, and visitors and association with burnout. *JAMA Netw Open*. 2022; 5(5):e2213080.
71. Rowe SG, Stewart MT, Van Horne S, et al. Mistreatment experiences, protective workplace systems, and occupational distress in physicians. *JAMA Netw Open*. 2022;5(5): e2210768.