

Original Paper

# Cluster Analysis of Primary Care Physician Phenotypes for Electronic Health Record Use: Retrospective Cohort Study

Allan Fong<sup>1</sup>, MS; Mark Iscoe<sup>2</sup>, MD; Christine A Sinsky<sup>3</sup>, MD; Adrian D Haimovich<sup>2</sup>, MD, PhD; Brian Williams<sup>4</sup>, MD; Ryan T O'Connell<sup>4</sup>, MD; Richard Goldstein<sup>4</sup>, MD, PhD; Edward Melnick<sup>2</sup>, MHS, MD

<sup>1</sup>National Center for Human Factors in Healthcare, MedStar Health, Washington, DC, United States

<sup>2</sup>Department of Emergency Medicine, Yale School of Medicine, New Haven, CT, United States

<sup>3</sup>American Medical Association, Chicago, IL, United States

<sup>4</sup>Northeast Medical Group, Yale New Haven Health, Stratford, CT, United States

**Corresponding Author:**

Allan Fong, MS

National Center for Human Factors in Healthcare

MedStar Health

3007 Tilden St NW

Washington, DC, 20008

United States

Phone: 1 2022449807

Email: [allan.fong@medstar.net](mailto:allan.fong@medstar.net)

## Abstract

**Background:** Electronic health records (EHRs) have become ubiquitous in US office-based physician practices. However, the different ways in which users engage with EHRs remain poorly characterized.

**Objective:** The aim of this study is to explore EHR use phenotypes among ambulatory care physicians.

**Methods:** In this retrospective cohort analysis, we applied affinity propagation, an unsupervised clustering machine learning technique, to identify EHR user types among primary care physicians.

**Results:** We identified 4 distinct phenotype clusters generalized across internal medicine, family medicine, and pediatrics specialties. Total EHR use varied for physicians in 2 clusters with above-average ratios of work outside of scheduled hours. This finding suggested that one cluster of physicians may have worked outside of scheduled hours out of necessity, whereas the other preferred ad hoc work hours. The two remaining clusters represented physicians with below-average EHR time and physicians who spend the largest proportion of their EHR time on documentation.

**Conclusions:** These findings demonstrate the utility of cluster analysis for exploring EHR use phenotypes and may offer opportunities for interventions to improve interface design to better support users' needs.

(*JMIR Med Inform* 2022;10(4):e34954) doi: [10.2196/34954](https://doi.org/10.2196/34954)

**KEYWORDS**

electronic health record; phenotypes; cluster analysis; unsupervised machine learning; machine learning; EHR; primary care

## Introduction

As of 2021, the vast majority of US office-based physicians used an electronic health record (EHR) [1]. The transition from paper to electronic records has many potential benefits but has also introduced new burdens. Furthermore, EHR use dominates clinical time [2] and is associated with burnout [3-5]. Despite the ubiquity of EHRs, patterns of clinician use are poorly characterized.

A 2019 survey study of clinicians reported widely divergent, subjective experiences with their EHR use and found that individual user differences accounted for over half of the variation in EHR use [6]. User-level variation can be due to disparities in proficiency that could potentially be remedied with appropriate training [7-10]. Emerging evidence suggests there are elements aside from proficiency that differentiate EHR users. For example, recent cross-sectional analyses of ambulatory care physicians' EHR use have found significant differences in time spent on EHRs based on gender [11,12], specialty [12,13], and country [14].

Audit logs offer a wealth of information derived from granular observations of users' EHR actions [15,16]. For example, research using log data has demonstrated associations between physicians' EHR activities and vendor-defined metrics of efficiency [17] and that efficiency varied based on physicians' years of experience and shift type [18]. In this study, we propose to use audit log data for the de novo identification of EHR user types (ie, EHR use phenotypes). Phenotype was first introduced by Richesson et al [19] as a biological concept to describe a set of observable biological traits. In the context of EHR use measures, phenotype will be used to describe observable use patterns across gender and specialty differences as defined by an unsupervised clustering approach called affinity propagation. First, 5 EHR use measures will be standardized using z-scores, which will then be used to calculate the similarities between physicians. A grid search and algorithm constraints will then be used to identify optimal clusters across a cohort of ambulatory care physicians.

## Methods

### Study Setting and Data Sources

This study retrospectively examined EHR log data of nontrainee, primary care physicians employed by a large ambulatory practice network (Northeast Medical Group) in northeastern United States (Connecticut, New York, and Rhode Island) between March 2018 and February 2020. Physicians were included if they specialized in general internal medicine, family medicine, or general pediatrics.

### Ethics Approval

All data were anonymized, with the investigators blinded to the participants' identities. The study protocol was approved by Northeast Medical Group's Institutional Review Board (IRB number 2000026556).

### EHR Use Measures

We retrieved data from the Epic Signal platform (Epic Systems) stratified by month and derived 5 proposed, time-based core EHR use measures normalized to 8 hours of scheduled patient time (Table 1) [20]. The first measure is EHR-Time<sub>8</sub>, defined as the time a physician spends on EHRs (both during and outside of scheduled patient hours) [20]. The second measure is work outside of work (WOW<sub>8</sub>), not to be confused with WOW carts (ie, workstations on wheels, a common industry term). WOW<sub>8</sub> is defined as the time a physician works on EHRs outside of scheduled patient hours [20]. The third measure is Note-Time<sub>8</sub>, defined as the time a physician spends on documentation [20]. The fourth and fifth measures are IB-Time<sub>8</sub> and Order-Time<sub>8</sub>, defined as the times a physician spends on inbox activities and on orders, respectively [20]. To account for relationships between EHR-Time<sub>8</sub> and its composite measures, we reported the ratios of WOW<sub>8</sub>, Note-Time<sub>8</sub>, IB-Time<sub>8</sub>, and Order-Time<sub>8</sub> to EHR-Time<sub>8</sub>, denoted as WOW-EHR, Note-EHR, IB-EHR, and Order-EHR, respectively. These measures (Table 1) were calculated and extracted from the Epic Signal platform, which have been validated and used in previous studies [20,21]. Each physician's EHR use measures were averaged across study months to account for variation in metric calculations introduced by changes in measure definitions over time due to the vendor's continuous quality improvement processes. For this analysis, we only considered physicians with valid metric months. Months with fewer than 30 clinical hours scheduled and less than 1 hour of EHR use were excluded from the analysis as invalid metric months. These thresholds were determined based on previous manual chart review validation and analysis of EHR vendor data [13].

**Table 1.** Electronic health record (EHR) use measures and definitions.

Measure	Definition
EHR-Time <sub>8</sub>	Time a physician spends on EHRs (both during and outside of scheduled patient hours) normalized to 8 hours of scheduled patient time
WOW-EHR	Ratio of EHR time that occurs during work outside of work (WOW <sub>8</sub> <sup>a</sup> ) hours: WOW <sub>8</sub> /EHR-Time <sub>8</sub>
Note-EHR	Ratio of EHR time a physician spends on documentation: Note-Time <sub>8</sub> <sup>b</sup> /EHR-Time <sub>8</sub>
IB-EHR	Ratio of EHR time a physician spends on inbox (IB) activities: IB-Time <sub>8</sub> <sup>c</sup> /EHR-Time <sub>8</sub>
Order-EHR	Ratio of EHR time a physician spends on orders: Order-Time <sub>8</sub> <sup>d</sup> /EHR-Time <sub>8</sub>

<sup>a</sup>WOW<sub>8</sub>: work outside of work hours normalized to 8 hours of scheduled patient time.

<sup>b</sup>Note-Time<sub>8</sub>: note time hours normalized to 8 hours of scheduled patient time.

<sup>c</sup>IB-Time<sub>8</sub>: inbox time hours normalized to 8 hours of scheduled patient time.

<sup>d</sup>Order-Time<sub>8</sub>: order time hours normalized to 8 hours of scheduled patient time.

### Cluster Analysis

Clusters were required to include individuals from at least two primary care specialties. Moreover, we did not require that all individuals be assigned to a phenotype cluster while also seeking

to minimize the total number of phenotypes. Affinity propagation, an algorithm that takes a set of pairwise similarities between data points and finds clusters on the basis of maximizing the total similarity between data points in a cluster, was used for phenotype discovery [22]. Affinity propagation

has advantages over other clustering algorithms, such as not predefining a number of clusters. A major disadvantage of affinity propagation is its high computational cost and resource requirement; however, this approach was deemed feasible given this study's sample size [22]. First, a standard z-score for each measure was calculated in order to center and scale the data. Similarities between data points were then calculated using Euclidean distance, which is defined for two 2D points as the length of the line formed by the two points. A grid search was then performed by varying the damping factor and preference from 0.5 to 1 and from 2 to 4, respectively, to identify the optimal clustering given the initial cluster conditions. Physicians in clusters that did not have representation from at least two specialties were excluded. Finally, physician gender and specialty distributions were described between clusters. All analyses were performed using Python software (version 3.7;

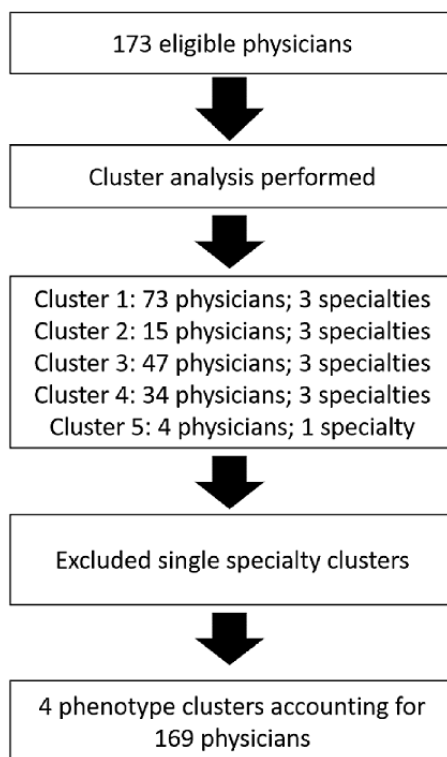
Python Software Foundation) and scikit-learn (version 0.24; scikit-learn developers) [23].

## Results

### Identifying Clusters

Of 332 ambulatory, nontrainee physicians, 290 (87.3%) have valid month metrics. Of those, a further 173 (52.1%) eligible physicians were of the specialties of interest: 117 (67.6%) in internal medicine, 36 (20.8%) in family medicine, and 20 (11.6%) in pediatrics. Gender distribution of the eligible physicians was 47.4% (82/173) female and 52.6% (91/173) male. We identified 4 clusters that met our a priori defined clustering conditions, accounting for 97.7% (169/173) of eligible physicians (Figure 1).

**Figure 1.** Summary of workflow and exclusion criteria.



### EHR Use Measures and Phenotypes Clusters

The phenotype clusters are “Lower EHR time,” “Higher note time,” “Work outside of work,” and “Notes outside of work.” The EHR use measures across clusters are summarized in Table 2. There was a significant association between phenotype

clusters and each EHR use measure: EHR-Time<sub>8</sub> (Kruskal-Wallis  $H=72.7$ ,  $P<.001$ ), WOW-EHR ( $H=84.3$ ,  $P<.001$ ), Note-EHR ( $H=89.0$ ,  $P<.001$ ), IB-EHR ( $H=45.8$ ,  $P<.001$ ), and Order-EHR ( $H=46.8$ ,  $P<.001$ ). The z-scores for the measures are displayed in Figure 2 to illustrate the relative differences between clusters.

**Table 2.** Electronic health record (EHR) use measures by phenotype cluster.

Measure	Phenotype clusters, median (IQR)				
	Lower EHR time	Higher note time	Work outside of work	Notes outside of work	All
EHR-Time <sub>g</sub> <sup>a</sup>	4.62 (4.20-5.43)	5.81 (4.41-6.22)	6.83 (5.95-8.36)	5.90 (5.37-6.36)	5.62 (4.57-6.40)
WOW-EHR <sup>b</sup>	0.07 (0.04-0.12)	0.05 (0.03-0.07)	0.21 (0.17-0.26)	0.13 (0.10-0.19)	0.11 (0.06-0.19)
Note-EHR <sup>c</sup>	0.24 (0.20-0.28)	0.46 (0.43-0.49)	0.31 (0.27-0.36)	0.37 (0.33-0.40)	0.29 (0.24-0.38)
IB-EHR <sup>d</sup>	0.14 (0.12-0.18)	0.06 (0.05-0.08)	0.15 (0.11-0.17)	0.10 (0.08-0.12)	0.13 (0.09-0.16)
Order-EHR <sup>e</sup>	0.19 (0.17-0.24)	0.14 (0.12-0.17)	0.16 (0.14-0.18)	0.14 (0.12-0.17)	0.17 (0.14-0.20)

<sup>a</sup>EHR-Time<sub>g</sub>: time a physician spends on EHRs normalized to 8 hours of scheduled patient time.

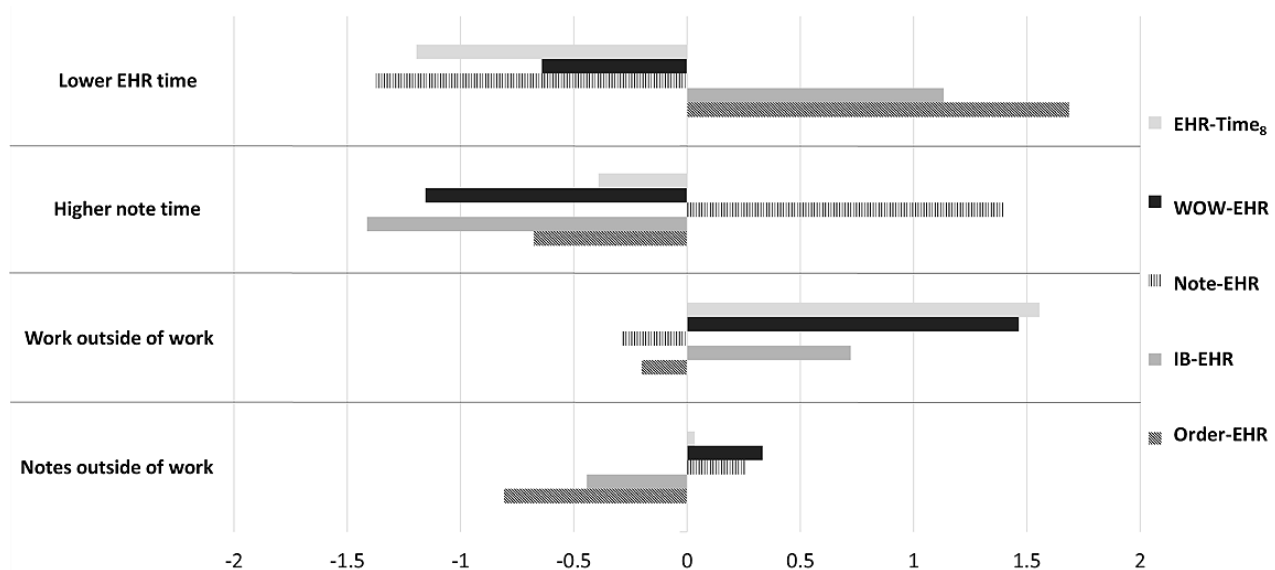
<sup>b</sup>WOW-EHR: ratio of EHR time that occurs during work outside of scheduled hours.

<sup>c</sup>Note-EHR: ratio of EHR time that a physician spends on documentation.

<sup>d</sup>IB-EHR: ratio of EHR time that a physician spends on inbox activities.

<sup>e</sup>Order-EHR: ratio of EHR time that a physician spends on orders.

**Figure 2.** Z-scores for electronic health record (EHR) use measure across clusters. EHR-Time<sub>g</sub>: time a physician spends on EHRs normalized to 8 hours of scheduled patient time; IB-EHR: ratio of EHR time that a physician spends on inbox activities; Note-EHR: ratio of EHR time that a physician spends on documentation; Order-EHR: ratio of EHR time that a physician spends on orders; WOW-EHR: ratio of EHR time that occurs during work outside of scheduled hours.



### “Lower EHR Time” Cluster

The “Lower EHR time” cluster was the largest cluster, constituting 42.2% (73/173) of eligible physicians. Physicians in this cluster spent the least amount of time on EHRs (EHR-Time<sub>g</sub>: median 4.62, IQR 4.20-5.43). “Lower EHR time” cluster physicians had the lowest median Note-EHR ratio of 0.24 (IQR 0.20-0.28) and the second lowest median WOW-EHR ratio of 0.07 (IQR 0.04-0.12). They also had the highest median IB-EHR and Order-EHR ratios of 0.14 (IQR 0.12-0.18) and 0.19 (IQR 0.17-0.24), respectively.

### “Higher Note Time” Cluster

“Higher note time” cluster physicians, constituting only 8.7% (15/173) of the total, had near-average normalized EHR time (EHR-Time<sub>g</sub>: median 5.81, IQR 4.41-6.22). Physicians in this cluster spent the largest proportion of their EHR time

documenting notes (Note-Time: median 0.46, IQR 0.43-0.49) compared to physicians in other clusters. They also spent the lowest proportions of that time on EHRs outside of scheduled hours and on inbox activities, with median WOW-EHR and IB-EHR ratios of 0.05 (IQR 0.03-0.07) and 0.06 (IQR 0.05-0.08), respectively.

### “Work Outside of Work” Cluster

“Work outside of work” cluster physicians, constituting 27.2% (47/173) of the total, spent the most time on EHRs (EHR-Time<sub>g</sub>: median 6.83, IQR 5.95-8.36) and the largest proportion of that time outside of work hours (WOW-EHR: median 0.21, IQR 0.17-0.26). This cluster of physicians had average median Note-EHR and Order-EHR ratios of 0.31 (IQR 0.27-0.36) and 0.16 (IQR 0.14-0.18), respectively, and an above-average median IB-EHR ratio of 0.15 (IQR 0.11-0.17).

### “Notes Outside of Work” Cluster

“Notes outside of work” cluster physicians, constituting 19.7% (34/173) of the total, had the second-highest median WOW-EHR ratio of 0.13 (IQR 0.10-0.19) but had near-average total normalized EHR time (EHR-Time<sub>8</sub>: median 5.90, IQR 5.37-6.36). This cluster of physicians had an above-average median Note-EHR ratio of 0.37 (IQR 0.33-0.40) and below-average median IB-EHR and Order-EHR ratios of 0.10 (IQR 0.08-0.12) and 0.14 (IQR 0.12-0.17), respectively.

### Phenotype Clusters by Specialty and Gender

Physician distribution across phenotype clusters by specialty and gender are reported in Table 3. There was a significant association between the clusters and specialty ( $X^2_6=26.67$ ,

$P<.001$ ). Pediatricians primarily fell into the “Higher note time” and “Notes outside of work” clusters (16/20, 80%) and accounted for 47% (7/15) of the total physicians in the “Higher note time” cluster. Family and internal medicine physicians were primarily distributed across the “Lower EHR time” and “Work outside of work” clusters (family medicine: 29/36, 81%; internal medicine: 87/113, 77%). In addition, there was a significant association between gender and clusters ( $X^2_3=18.28$ ,  $P<.001$ ). Female physicians were more prominent in the “Work outside of work” and “Notes outside of work” clusters, accounting for 64% (30/47) and 62% (21/34) of the clusters, respectively. Male physicians accounted for 71% (52/73) of the “Lower EHR time” cluster.

**Table 3.** Physician specialty and gender distribution by phenotype cluster.

Distribution	Number of physicians (N=173), n (%)	Phenotype clusters				P value
		Lower EHR <sup>a</sup> time (n=73), n (%)	Higher note time (n=15), n (%)	Work outside of work (n=47), n (%)	Notes outside of work (n=34), n (%)	
<b>Specialty</b>						<.001
Family medicine	36 (21)	19 (26)	2 (13)	10 (21)	5 (15)	
Internal medicine	113 (65)	52 (71)	6 (40)	35 (74)	20 (59)	
Pediatrics	20 (12)	2 (3)	7 (47)	2 (4)	9 (26)	
<b>Gender</b>						<.001
Female	80 (46)	21 (29)	8 (53)	30 (64)	21 (62)	
Male	89 (51)	52 (71)	7 (47)	17 (36)	13 (38)	
Total	169 (98)	73 (42)	15 (9)	47 (27)	34 (20)	

<sup>a</sup>EHR: electronic health record.

## Discussion

### Principal Findings

In this unsupervised clustering machine learning analysis of a cohort of primary care physicians, we identified 4 distinct EHR use phenotypes characterized by the total time spent on EHR activities and the ratios of those times in comparison to one another. These phenotypes were differentiated and described by patterns of use consistent with overall efficiency, higher documentation time, and working outside of work hours; each of these patterns of use were generally associated with the “Lower EHR time,” “Higher note time,” and “Work/Notes outside of work” clusters, respectively. While exploratory, these results provide insights into EHR use phenotypes across gender and specialties that can complement and provide additional context for current EHR use research.

### Work Outside of Scheduled Hours

We identified 2 phenotype clusters that had above-average ratios for work outside of scheduled hours. Although “Work outside of work” and “Notes outside of work” clusters both had high WOW-EHR ratios, only the “Work outside of work” cluster

had significantly higher than average EHR-Time<sub>8</sub>. A possible explanation for this is that physicians in the “Work outside of work” cluster work from home partly out of necessity because they require more time on EHRs, whereas physicians in the “Notes outside of work” cluster may elect to finish work at home, suggesting a preference for ad hoc work hours.

### Note Time

Time spent on clinical documentation accounted for the largest proportion of total EHR time in each cluster. There was, however, considerable variation in the ratio of note time to EHR time across clusters: from 0.24 of EHR time in the “Lower EHR time” cluster to 0.46 in the “Higher note time” cluster despite similar total EHR time in both clusters. Potential explanations for this variation include differences in clinic- or physician-specific workflows (eg, scribe support or team-based documentation; differences in depth and complexity of encounters and expectations for documentation; and use of form, copied, or auto-populated notes) and differences in documentation style, particularly among the “Higher note time” cluster that may include physicians who deliberately spend more time on documentation.



## Limitations

This exploratory work only used time-based metrics and did not account for patient acuity or complexity. Although the data were gathered over a 2-year period, systemic differences in patient volume and care could have affected the results. In addition, this work was limited to a single ambulatory practice network in one region of the United States and was limited to primary care physicians. Some types of EHR activities (eg, chart review) were not included in the metrics, and it is possible that other activities or practice domains could also affect clustering. Furthermore, it should be noted that this study only identified EHR use phenotypes and did not explore reasons behind differences in EHR use or assign value to the phenotypes.

## Conclusions

Our findings may highlight opportunities for interventions to improve EHR design and use to better support EHR users'

needs. Potential differences in users' needs were identified for each phenotype cluster. The "Higher note time" and "Notes outside of work" clusters might benefit from scribe support more than the other two clusters. The "Work outside of work" cluster might benefit from inbox support and restructuring their practice for a more team-based approach. Physicians in the "Lower EHR time" cluster could be consulted as local champions to help their peers improve their EHR efficiency. By identifying and classifying individual EHR use and user needs, we can better understand and target interventions at the individual or department level. Future work should validate these phenotypes in larger cohorts and in diverse settings, explore differences in physicians' training and demographics across phenotypes, and investigate the relationships among EHR use phenotypes, patient outcomes, and clinician satisfaction and burnout.

## Acknowledgments

We thank the physicians of the Northeast Medical Group. We also thank the following individuals for their contribution and support: Prem Thomas, MD, Center for Medical Informatics, Yale School of Medicine (data analysis, not compensated); and Becky Tylutki, Team Coordinator, Joint Data Analytics Team, Yale New Haven Health System (data collection and honest broker, not compensated). Financial support for this study was provided in part by two American Medical Association Practice Transformation Initiatives (contract numbers 36648 and 36650). The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report.

## Conflicts of Interest

CAS is employed by the American Medical Association. All other authors declare no other conflicts of interest. The opinions expressed in this article are those of the authors and should not be interpreted as American Medical Association policy.

## References

1. Office-based physician health IT adoption. The Office of the National Coordinator for Health Information Technology. 2021. URL: <https://www.healthit.gov/data/apps/office-based-physician-health-it-adoption> [accessed 2021-09-24]
2. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med* 2016 Dec 06;165(11):753-760. [doi: [10.7326/M16-0961](https://doi.org/10.7326/M16-0961)] [Medline: [27595430](https://pubmed.ncbi.nlm.nih.gov/27595430/)]
3. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. *Mayo Clin Proc* 2016 Jul;91(7):836-848. [doi: [10.1016/j.mayocp.2016.05.007](https://doi.org/10.1016/j.mayocp.2016.05.007)] [Medline: [27313121](https://pubmed.ncbi.nlm.nih.gov/27313121/)]
4. Gardner RL, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, et al. Physician stress and burnout: the impact of health information technology. *J Am Med Inform Assoc* 2019 Feb 01;26(2):106-114 [FREE Full text] [doi: [10.1093/jamia/ocy145](https://doi.org/10.1093/jamia/ocy145)] [Medline: [30517663](https://pubmed.ncbi.nlm.nih.gov/30517663/)]
5. Olson K, Sinsky C, Rinne ST, Long T, Vender R, Mukherjee S, 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 Apr;35(2):157-175. [doi: [10.1002/smi.2849](https://doi.org/10.1002/smi.2849)] [Medline: [30467949](https://pubmed.ncbi.nlm.nih.gov/30467949/)]
6. Longhurst CA, Davis T, Maneker A, Eschenroeder HC, Dunscombe R, Reynolds G, Arch Collaborative. Local investment in training drives electronic health record user satisfaction. *Appl Clin Inform* 2019 Mar;10(2):331-335 [FREE Full text] [doi: [10.1055/s-0039-1688753](https://doi.org/10.1055/s-0039-1688753)] [Medline: [31091545](https://pubmed.ncbi.nlm.nih.gov/31091545/)]
7. Robinson KE, Kersey JA. Novel electronic health record (EHR) education intervention in large healthcare organization improves quality, efficiency, time, and impact on burnout. *Medicine (Baltimore)* 2018 Sep;97(38):e12319 [FREE Full text] [doi: [10.1097/MD.00000000000012319](https://doi.org/10.1097/MD.00000000000012319)] [Medline: [30235684](https://pubmed.ncbi.nlm.nih.gov/30235684/)]
8. Sieja A, Markley K, Pell J, Gonzalez C, Redig B, Kneeland P, et al. Optimization sprints: improving clinician satisfaction and teamwork by rapidly reducing electronic health record burden. *Mayo Clin Proc* 2019 May;94(5):793-802 [FREE Full text] [doi: [10.1016/j.mayocp.2018.08.036](https://doi.org/10.1016/j.mayocp.2018.08.036)] [Medline: [30824281](https://pubmed.ncbi.nlm.nih.gov/30824281/)]
9. DiAngi YT, Stevens LA, Halpern-Felsher B, Pageler NM, Lee TC. Electronic health record (EHR) training program identifies a new tool to quantify the EHR time burden and improves providers' perceived control over their workload in the EHR. *JAMIA Open* 2019 Jul;2(2):222-230 [FREE Full text] [doi: [10.1093/jamiaopen/ooz003](https://doi.org/10.1093/jamiaopen/ooz003)] [Medline: [31984357](https://pubmed.ncbi.nlm.nih.gov/31984357/)]

10. Hollister-Meadows L, Richesson RL, De Gagne J, Rawlins N. Association between evidence-based training and clinician proficiency in electronic health record use. *J Am Med Inform Assoc* 2021 Mar 18;28(4):824-831 [[FREE Full text](#)] [doi: [10.1093/jamia/ocaa333](https://doi.org/10.1093/jamia/ocaa333)] [Medline: [33575787](https://pubmed.ncbi.nlm.nih.gov/33575787/)]
11. Tait SD, Oshima SM, Ren Y, Fenn AE, Boazak M, Hinz EM, et al. Electronic health record use by sex among physicians in an academic health care system. *JAMA Intern Med* 2021 Feb 01;181(2):288-290 [[FREE Full text](#)] [doi: [10.1001/jamainternmed.2020.5036](https://doi.org/10.1001/jamainternmed.2020.5036)] [Medline: [33284311](https://pubmed.ncbi.nlm.nih.gov/33284311/)]
12. Rotenstein LS, Holmgren AJ, Downing NL, Bates DW. Differences in total and after-hours electronic health record time across ambulatory specialties. *JAMA Intern Med* 2021 Jun 01;181(6):863-865 [[FREE Full text](#)] [doi: [10.1001/jamainternmed.2021.0256](https://doi.org/10.1001/jamainternmed.2021.0256)] [Medline: [33749732](https://pubmed.ncbi.nlm.nih.gov/33749732/)]
13. Melnick ER, Ong SY, Fong AL, Socrates V, Ratwani RM, Nath B, et al. Characterizing physician EHR use with vendor derived data: a feasibility study and cross-sectional analysis. *J Am Med Inform Assoc* 2021 Jul 14;28(7):1383-1392 [[FREE Full text](#)] [doi: [10.1093/jamia/ocab011](https://doi.org/10.1093/jamia/ocab011)] [Medline: [33822970](https://pubmed.ncbi.nlm.nih.gov/33822970/)]
14. Holmgren AJ, Downing NL, Bates DW, Shanafelt TD, Milstein A, Sharp CD, et al. Assessment of electronic health record use between US and non-US health systems. *JAMA Intern Med* 2021 Feb 01;181(2):251-259 [[FREE Full text](#)] [doi: [10.1001/jamainternmed.2020.7071](https://doi.org/10.1001/jamainternmed.2020.7071)] [Medline: [33315048](https://pubmed.ncbi.nlm.nih.gov/33315048/)]
15. Adler-Milstein J, Adelman JS, Tai-Seale M, Patel VL, Dymek C. EHR audit logs: A new goldmine for health services research? *J Biomed Inform* 2020 Jan;101:103343 [[FREE Full text](#)] [doi: [10.1016/j.jbi.2019.103343](https://doi.org/10.1016/j.jbi.2019.103343)] [Medline: [31821887](https://pubmed.ncbi.nlm.nih.gov/31821887/)]
16. Amroze A, Field TS, Fouayzi H, Sundaresan D, Burns L, Garber L, et al. Use of electronic health record access and audit logs to identify physician actions following noninterruptive alert opening: descriptive study. *JMIR Med Inform* 2019 Feb 07;7(1):e12650 [[FREE Full text](#)] [doi: [10.2196/12650](https://doi.org/10.2196/12650)] [Medline: [30730293](https://pubmed.ncbi.nlm.nih.gov/30730293/)]
17. Kannampallil TG, Denton CA, Shapiro JS, Patel VL. Efficiency of emergency physicians: insights from an observational study using EHR log files. *Appl Clin Inform* 2018 Jan;9(1):99-104 [[FREE Full text](#)] [doi: [10.1055/s-0037-1621705](https://doi.org/10.1055/s-0037-1621705)] [Medline: [30184241](https://pubmed.ncbi.nlm.nih.gov/30184241/)]
18. Wang JK, Ouyang D, Hom J, Chi J, Chen JH. Characterizing electronic health record usage patterns of inpatient medicine residents using event log data. *PLoS One* 2019 Feb 06;14(2):e0205379 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0205379](https://doi.org/10.1371/journal.pone.0205379)] [Medline: [30726208](https://pubmed.ncbi.nlm.nih.gov/30726208/)]
19. Richesson RL, Hammond WE, Nahm M, Wixted D, Simon GE, Robinson JG, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *J Am Med Inform Assoc* 2013 Dec 01;20(e2):e226-e231 [[FREE Full text](#)] [doi: [10.1136/amiajnl-2013-001926](https://doi.org/10.1136/amiajnl-2013-001926)] [Medline: [23956018](https://pubmed.ncbi.nlm.nih.gov/23956018/)]
20. Sinsky CA, Rule A, Cohen G, Arndt BG, Shanafelt TD, Sharp CD, et al. Metrics for assessing physician activity using electronic health record log data. *J Am Med Inform Assoc* 2020 Apr 01;27(4):639-643 [[FREE Full text](#)] [doi: [10.1093/jamia/ocz223](https://doi.org/10.1093/jamia/ocz223)] [Medline: [32027360](https://pubmed.ncbi.nlm.nih.gov/32027360/)]
21. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan WJ, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med* 2017 Sep;15(5):419-426 [[FREE Full text](#)] [doi: [10.1370/afm.2121](https://doi.org/10.1370/afm.2121)] [Medline: [28893811](https://pubmed.ncbi.nlm.nih.gov/28893811/)]
22. Frey BJ, Dueck D. Clustering by passing messages between data points. *Science* 2007 Feb 16;315(5814):972-976. [doi: [10.1126/science.1136800](https://doi.org/10.1126/science.1136800)] [Medline: [17218491](https://pubmed.ncbi.nlm.nih.gov/17218491/)]
23. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825-2830. [doi: [10.48550/arXiv.1201.0490](https://doi.org/10.48550/arXiv.1201.0490)]

## Abbreviations

**EHR:** electronic health record

**EHR-Time<sub>g</sub>:** time a physician spends on EHRs normalized to 8 hours of scheduled patient time

**IB-EHR:** percent of EHR time that a physician spends on inbox activities

**IB-Time<sub>g</sub>:** inbox time hours normalized to 8 hours of scheduled patient time

**Note-EHR:** percent of EHR time that a physician spends on documentation

**Note-Time<sub>g</sub>:** note time hours normalized to 8 hours of scheduled patient time

**Order-EHR:** percent of EHR time that a physician spends on orders

**Order-Time<sub>g</sub>:** order time hours normalized to 8 hours of scheduled patient time

**WOW<sub>g</sub>:** work outside of work hours normalized to 8 hours of scheduled patient time

**WOW-EHR:** percent of EHR time that occurs during work outside of scheduled hours

*Edited by C Lovis; submitted 14.11.21; peer-reviewed by R Verheij, A Joseph, M Pradhan; comments to author 08.01.22; revised version received 03.02.22; accepted 11.03.22; published 15.04.22*

*Please cite as:*

*Fong A, Iscoe M, Sinsky CA, Haimovich AD, Williams B, O'Connell RT, Goldstein R, Melnick E*

*Cluster Analysis of Primary Care Physician Phenotypes for Electronic Health Record Use: Retrospective Cohort Study*

*JMIR Med Inform 2022;10(4):e34954*

URL: <https://medinform.jmir.org/2022/4/e34954>

doi: [10.2196/34954](https://doi.org/10.2196/34954)

PMID: [35275070](https://pubmed.ncbi.nlm.nih.gov/35275070/)

©Allan Fong, Mark Iscoe, Christine A Sinsky, Adrian D Haimovich, Brian Williams, Ryan T O'Connell, Richard Goldstein, Edward Melnick. Originally published in JMIR Medical Informatics (<https://medinform.jmir.org>), 15.04.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication on <https://medinform.jmir.org/>, as well as this copyright and license information must be included.