

Automating Measurement of Trainee Work Hours

Hossein Soleimani, PhD¹, Julia Adler-Milstein, PhD^{2,3}, Russell J Cucina, MD, MS^{1,3}, Sara G Murray, MD, MAS^{1,3*}

¹Health Informatics, University of California, San Francisco, San Francisco, California; ²Center for Clinical Informatics and Improvement Research, University of California, San Francisco, San Francisco, California; ³Department of Medicine, University of California, San Francisco, San Francisco, California.

BACKGROUND: Medical training programs across the country are bound to a set of work hour regulations, generally monitored via self-report.

OBJECTIVE: We developed a computational method to automate measurement of intern and resident work hours, which we validated against self-report.

DESIGN, SETTING, AND PARTICIPANTS: We included all electronic health record (EHR) access log data between July 1, 2018, and June 30, 2019, for trainees enrolled in the internal medicine training program. We inferred the duration of continuous in-hospital work hours by linking EHR sessions that occurred within 5 hours as “on-campus” work and further accounted for “out-of-hospital” work which might be taking place at home.

MAIN OUTCOMES AND MEASURES: We compared daily work hours estimated through the computational method with self-report and calculated the mean absolute error between the two groups. We used the

computational method to estimate average weekly work hours across the rotation and the percentage of rotations where average work hours exceeded the 80-hour workweek.

RESULTS: The mean absolute error between self-reported and EHR-derived daily work hours for first- (PGY-1), second- (PGY-2), and third- (PGY-3) year trainees were 1.27, 1.51, and 1.51 hours, respectively. Using this computational method, we estimated average (SD) weekly work hours of 57.0 (21.7), 69.9 (12.2), and 64.1 (16.3) for PGY-1, PGY-2, and PGY-3 residents.

CONCLUSION: EHR log data can be used to accurately approximate self-report of work hours, accounting for both in-hospital and out-of-hospital work. Automation will reduce trainees’ clerical work, improve consistency and comparability of data, and provide more complete and timely data that training programs need. *Journal of Hospital Medicine* 2021;16:404-408. © 2021 Society of Hospital Medicine

Across the country, residents are bound to a set of rules from the Accreditation Council for Graduate Medical Education (ACGME) designed to minimize fatigue, maintain quality of life, and reduce fatigue-related patient safety events. Adherence to work hours regulations is required to maintain accreditation. Among other guidelines, residents are required to work fewer than 80 hours per week on average over 4 consecutive weeks.¹ When work hour violations occur, programs risk citation, penalties, and harm to the program’s reputation.

Residents self-report their adherence to program regulations in an annual survey conducted by the ACGME.² To collect more frequent data, most training programs monitor resident work hours through self-report on an electronic tracking platform.³ These data generally are used internally to identify problems and opportunities for improvement. However, self-report approaches are subject to imperfect recall and incomplete reporting, and require time and effort to complete.⁴

The widespread adoption of electronic health records

(EHRs) brings new opportunity to measure and promote adherence to work hours. EHR log data capture when users log in and out of the system, along with their location and specific actions. These data offer a compelling alternative to self-report because they are already being collected and can be analyzed almost immediately. Recent studies using EHR log data to approximate resident work hours in a pediatric hospital successfully approximated scheduled hours, but the approach was customized to their hospital’s workflows and might not generalize to other settings.⁵ Furthermore, earlier studies have not captured evening out-of-hospital work, which contributes to total work hours and is associated with physician burnout.⁶

We developed a computational method that sought to accurately capture work hours, including out-of-hospital work, which could be used as a screening tool to identify at-risk residents and rotations in near real-time. We estimated work hours, including EHR and non-EHR work, from these EHR data and compared these daily estimations to self-report. We then used a heuristic to estimate the frequency of exceeding the 80-hour workweek in a large internal medicine residency program.

METHODS

The population included 82 internal medicine interns (PGY-1) and 121 residents (PGY-2 = 60, PGY-3 = 61) who rotated through University of California, San Francisco Medical Center (UCSFMC) between July 1, 2018, and June 30, 2019, on inpatient rotations.

*Corresponding Author: Sara G Murray, MD, MAS; Email: Sara.murray@ucsf.edu.

Published online first April 16, 2021.

Find additional supporting information in the online version of this article.

Received: September 25, 2020; Revised: January 4, 2021; Accepted: February 3, 2021

© 2021 Society of Hospital Medicine DOI 10.12788/jhm.3607

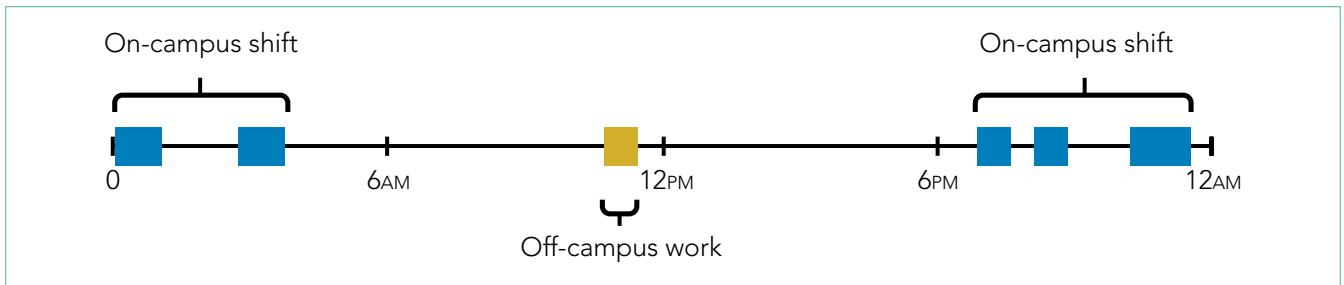


FIG 1. Approach to Linking EHR Sessions to Measure the Total Workday. EHR sessions within 5 hours were linked, forming an entire on-campus workday. Out-of-hospital logins were added as individual entities.

In the UCSF internal medicine residency program, interns spend an average of 5 months per year and residents spend an average of 2 months per year on inpatient rotations at UCSFMC. Scheduled inpatient rotations generally are in 1-month blocks and include general medical wards, cardiology, liver transplant, night-float, and a procedures and jeopardy rotation where interns perform procedures at UCSFMC and serve as backup for their colleagues across sites. Although expected shift duration differs by rotation, types of shifts include regular length days, call days that are not overnight (but expected duration of work is into the late evening), 28-hour overnight call (PGY-2 and PGY-3), and night-float.

Data Source

This computational method was developed at UCSFMC. This study was approved by the University of California, San Francisco institutional review board. Using the UCSF Epic Clarity database, EHR access log data were obtained, including all Epic logins/logoffs, times, and access devices. Access devices identified included medical center computers, personal computers, and mobile devices.

Trainees self-report their work hours in MedHub, a widely used electronic tracking platform for self-report of resident work hours.⁷ Data were extracted from this database for interns and residents who matched the criteria above. The self-report data were considered the gold standard for comparison, because it is the best available despite its known limitations.

We used data collected from UCSF's physician scheduling platform, AMiON, to identify interns and residents assigned to rotations at UCSF hospitals.⁸ AMiON also was used to capture half-days of off-site scheduled clinics and teaching, which count toward the workday but would not be associated with on-campus logins.

Developing a Computational Method to Measure Work Hours

We developed a heuristic to accomplish two goals: (1) infer the duration of continuous in-hospital work hours while providing clinical care and (2) measure "out-of-hospital" work. Logins from medical center computers were considered to be "on-campus" work. Logins from personal computers were considered to be "out-of-hospital." "Out-of-hospital" login sessions were further subdivided into "out-of-hospital work" and "out-of-hospital

study" based on activity during the session; if any work activities listed in Appendix Table 1 were performed, the session was attributed to work. If only chart review was performed, the session was attributed to study and did not count towards total hours worked. Logins from mobile devices also did not count towards total hours worked.

We inferred continuous in-hospital work by linking on-campus EHR sessions from the first on-campus login until the last on-campus logoff (Figure 1). Based on our knowledge of workflows, residents generally print their patient lists when they arrive at the hospital and use the EHR to update hand-off information before they leave. To computationally infer a continuous workday, we determined the maximum amount of time between an on-campus logoff and a subsequent on-campus login that could be inferred as continuous work in the hospital. We calculated the probability that an individual would log in on-campus again at any given number of hours after they last logged out (Appendix Figure 1). We found that for any given on-campus logoff, there was a 93% chance an individual will log in again from on-campus within the next 5 hours, indicating continuous on-campus work. However, after more than 5 hours have elapsed, there is a 90% chance that at least 10 hours will elapse before the next on-campus login, indicating the break between on-campus workdays. We therefore used 5 hours as the maximum interval between on-campus EHR sessions that would be linked together to classify on-campus EHR sessions as a single workday. This window accounts for resident work in direct patient care, rounds, and other activities that do not involve the EHR.

If there was overlapping time measurement between on-campus work and personal computer logins (for example, a resident was inferred to be doing on-campus work based on frequent medical center computer logins but there were also logins from personal computers), we inferred this to indicate that a personal device had been brought on-campus and the time was only attributed to on-campus work and was not double counted as out-of-hospital work. Out-of-hospital work that did not overlap with inferred on-campus work time contributed to the total hours worked in a week, consistent with ACGME guidelines.

Our internal medicine residents work at three hospitals: UCSFMC and two affiliated teaching hospitals. Although this study measured work hours while the residents were on an inpatient rotation at UCSFMC, trainees also might have occasional half-day clinics or teaching activities at other sites not captured

TABLE 1. Total Days Worked at UCSFMC, Number of Rotations Worked at UCSFMC, Total Days With Self-Reported Hours, and Proportion of Days for Which There Was Self-Reporting

PGY	No. of days worked at UCSFMC		No. of rotations worked at UCSFMC		No. of days worked at UCSFMC with self-reported hours		Percentage of days worked at UCSFMC with self-reported hours	
	Total	Mean (SD) per trainee	Total	Mean (SD) per trainee	Total	Mean (SD) per trainee	Overall	Mean (SD) per trainee
1	8445	103.0 (19.4)	406	5.0 (1.5)	5401	87.1 (29.1)	64.0	60.2 (40.7)
2	3488	58.1 (19.2)	140	2.4 (0.8)	2688	47.2 (21.9)	77.1	75.1 (30.4)
3	2677	44.6 (20.5)	95	1.7 (0.8)	1672	30.4 (20.2)	62.5	61.9 (36.7)

Abbreviations: PGY, postgraduate year; UCSFMC, University of California, San Francisco Medical Center.

by these EHR log data. The allocated time for that scheduled activity (extracted from AMiON) was counted as work hours. If the trainee was assigned to a morning half-day of off-site work (eg, didactics), this was counted the same as an 8 AM to noon on-campus EHR session. If a trainee was assigned an afternoon half-day of off-site work (eg, a non-UCSF clinic), this was counted the same as a 1 PM to 5 PM on-campus EHR session. Counting this scheduled time as an on-campus EHR session allowed half-days of off-site work to be linked with inferred in-hospital work.

Comparison of EHR-Derived Work Hours Heuristic to Self-Report

Because resident adherence with daily self-report is imperfect, we compared EHR-derived work to self-report on days when both were available. We generated scatter plots of EHR-derived work hours compared with self-report and calculated the mean absolute error of estimation. We fit a linear mixed-effect model for each PGY, modeling self-reported hours as a linear function of estimated hours (fixed effect) with a random intercept (random effect) for each trainee to account for variations among individuals. StatsModels, version 0.11.1, was used for statistical analyses.⁹

We reviewed detailed data from outlier clusters to understand situations where the heuristic might not perform optimally. To assess whether EHR-derived work hours reasonably overlapped with expected shifts, 20 8-day blocks from separate interns and residents were randomly selected for qualitative detail review in comparison with AMiON schedule data.

Estimating Hours Worked and Work Hours Violations

After validating against self-report on a daily basis, we used our heuristic to infer the average rate at which the 80-hour workweek was exceeded across all inpatient rotations at UCSFMC. This was determined both including "out-of-hospital" work as derived from logins on personal computers and excluding it. Using the estimated daily hours worked, we built a near real-time dashboard to assist program leadership with identifying at-risk trainees and trends across the program.

RESULTS

Data from 82 interns (PGY-1) and 121 internal medicine residents (PGY-2 and PGY-3) who rotated at UCSFMC between July 1, 2018, and June 30, 2019, were included in the study. Table 1 shows the number of days and rotations worked at UCSFMC

as well as the frequency of self-report of work hours according to program year. Figure 2 shows scatter plots for self-report of work hours compared with work hours estimated from our computational method. The mean absolute error in estimation of self-report with the heuristic is 1.38 hours. Explanations for outlier groups also are described in Figure 2. Appendix Figure 2 shows the distribution of the differences between estimated and self-reported daily work hours.

Qualitative review of EHR-derived data compared with schedule data showed that, although residents often reported homogenous daily work hours, EHR-derived work hours often varied as expected on a day-to-day basis according to the schedule (Appendix Table 2).

Because out-of-hospital EHR use does not count as work if done for educational purposes, we evaluated the proportion of out-of-hospital EHR use that is considered work and found that 67% of PGY-1, 50% of PGY-2, and 53% of PGY-3 out-of-hospital sessions included at least one work activity, as denoted in Appendix Table 1. Out-of-hospital work therefore represented 85% of PGY-1, 66% of PGY-2, and 73% of PGY-3 time spent in the EHR out-of-hospital. These sessions were counted towards work hours in accordance with ACGME rules and included 29% of PGY-1 workdays and 21% of PGY-2 and PGY-3 workdays. This amounted to a median of 1.0 hours per day (95% CI, 0.1-4.6 hours) of out-of-hospital work for PGY-1, 0.9 hours per day (95% CI, 0.1-4.1 hours) for PGY-2, and 0.8 hours per day (95% CI, 0.1-4.7 hours) for PGY-3 residents. Out-of-hospital logins that did not include work activities, as denoted in Appendix Table 1, were labeled out-of-hospital study and did not count towards work hours; this amounted to a median of 0.3 hours per day (95% CI, 0.02-1.6 hours) for PGY-1, 0.5 hours per day (95% CI, 0.04-0.25 hours) for PGY-2, and 0.3 hours per day (95% CI, 0.03-1.7 hours) for PGY-3. Mobile device logins also were not counted towards total work hours, with a median of 3 minutes per day for PGY-1, 6 minutes per day for PGY-2, and 5 minutes per day for PGY-3.

The percentage of rotation months where average hours worked exceeded 80 hours weekly is shown in Table 2. Inclusion of out-of-hospital work hours substantially increased the frequency at which the 80-hour workweek was exceeded. The frequency of individual residents working more than 80 hours weekly on average is shown in Appendix Figure 3. A narrow majority of PGY-1 and PGY-2 trainees and a larger majority of PGY-3 trainees never worked in excess of 80 hours per week when averaged over the

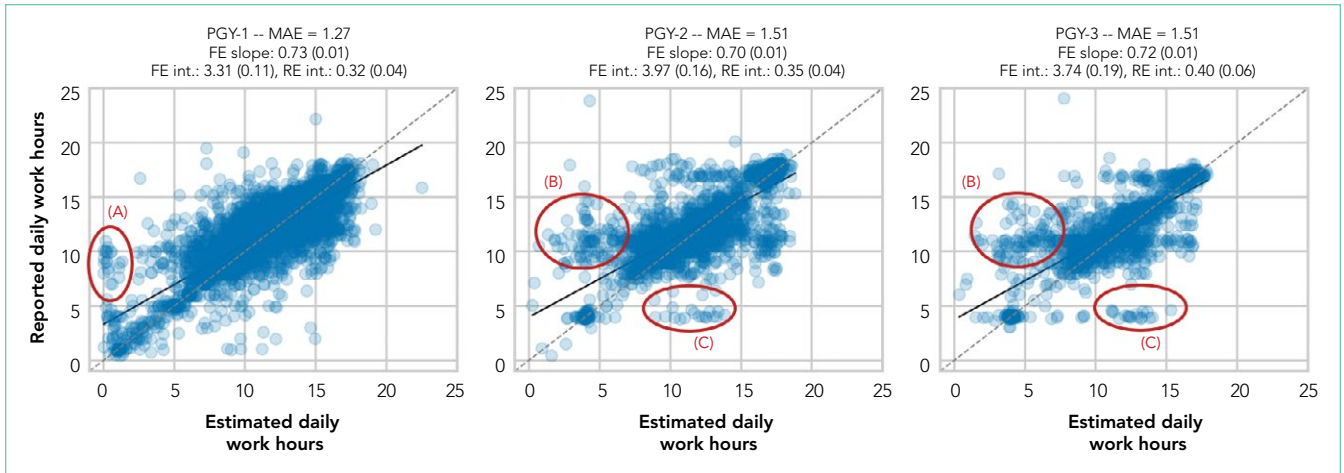


FIG 2. Daily Work Hours Estimated With the Computational Heuristic in Comparison to Self-Report. Mean absolute error (MAE) per year is shown in the graph. The linear regression line indicates the computational heuristic overestimates daily work hours when self-reported hours are large and underestimates when small. We report the fixed effect (FE) slope and intercept coefficients and the random effect intercept (RE) along with their corresponding standard error in the graph. Outlier groups: (A) The cohort of PGY-1 interns where the heuristic appears to underestimate self-report is largely made up of a backup rotation where residents may cover for sick colleagues at other hospitals. (B) The cohort of PGY-2 and PGY-3 residents where the heuristic appears to underestimate self-report is largely comprised of a UCSF-based night-float rotation where they might work at other hospitals, and, in rare cases, prolonged overnight electronic health record silence while residents were sleeping for long periods in the hospital while on-call. (C) The cohort of PGY-2 and PGY-3 in which the heuristic appears to overestimate work hours are largely circumstances where residents are working nightshifts but attribute all work hours to 1 day rather than splitting time before and after midnight.

course of a rotation, but several trainees did on several occasions.

Estimations from the computational method were built into a dashboard for use as screening tool by residency program directors (Appendix Figure 4).

DISCUSSION

EHR log data can be used to automate measurement of trainee work hours, providing timely data to program directors for identifying residents at risk of exceeding work hours limits. We demonstrated this by developing a data-driven approach to link on-campus logins that can be replicated in other training programs. We further demonstrated that out-of-hospital work substantially contributed to resident work hours and the frequency with which they exceed the 80-hour workweek, making it a critical component of any work hour estimation approach. Inclusive of out-of-hospital work, our computational method found that residents exceeded the 80-hour workweek 10% to 21% of the time, depending on their year in residency, with a small majority of residents never exceeding the 80-hour workweek.

Historically, most ACGME residency programs have relied on resident self-report to determine work hours.³ The validity of this method has been extensively studied and results remain mixed; in some surveys, residents admit to underreporting their hours while other validation studies, including the use of clock-in and clock-out or time-stamped parking data, align with self-report relatively well.¹⁰⁻¹² Regardless of the reliability of self-report, it is a cumbersome task that residents have difficulty adhering to, as shown in our study, where only slightly more than one-half of the days worked had associated self-report. By relying on resident self-report, we are adding to the burden of clerical work, which is associated with physician burnout.¹³ Furthermore, because self-report typically does not happen in real-time, it limits

a program's ability to intervene on recent or impending work-hour violations. Our computational method enabled us to build a dashboard that is updated daily and provides critical insight into resident work hours at any time, without waiting for retrospective self-report.

Our study builds on previous work by Dziorny et al using EHR log data to algorithmically measure in-hospital work.⁵ In their study, the authors isolated shifts with a login gap of 4 hours and then combined shifts according to a set of heuristics. However, their logic integrated an extensive workflow analysis of trainee shifts, which might limit generalizability.⁵ Our approach computationally derives the temporal threshold for linking EHR sessions, which in our data was 5 hours but might differ at other sites. Automated derivation of this threshold will support generalizability to other programs and sites, although programs will still need to manually account for off-site work such as didactics. In a subsequent study evaluating the 80-hour workweek, Dziorny et al evaluated shift duration and appropriate time-off between shifts and found systematic underreporting of work.¹⁴ In our study, we prioritized evaluation of the 80-hour workweek and found general alignment between self-report and EHR-derived work-hour estimates, with a tendency to underestimate at lower reported work hours and overestimate at higher reported work hours (potentially because of underreporting as illustrated by Dziorny et al). We included the important out-of-hospital logins as discrete work events because out-of-hospital work contributes to the total hours worked and to the number of workweeks that exceed the 80-hour workweek, and might contribute to burnout.¹⁵ The incidence of exceeding the 80-hour workweek increased by 7% to 8% across all residents when out-of-hospital work was included, demonstrating that tools such as ResQ (ResQ Medical) that rely pri-

TABLE 2. Impact of Out-Of-Hospital Work on the Percentage of Rotation Months That Exceed the 80-Hour Workweek

PGY	Without out-of-hospital work		Including out-of-hospital work	
	Mean (SD) weekly work hours	Rate (95% CI) of exceeding 80 hours /week, %	Mean (SD) weekly work hours	Rate (95% CI) of exceeding 80 hours per week, %
1	54.8 (20.8)	7.0 (5.0-10.1)	57.0 (21.7)	15.0 (11.9-18.3)
2	68.3 (11.9)	13.0 (8.3-19.4)	69.9 (12.2)	21.4 (14.4-28.9)
3	62.3 (15.9)	3.0 (1.1-8.9)	64.1 (16.3)	10.5 (5.8-18.3)

marily on geolocation data might not sufficiently capture the ways in which residents spend their time working.¹⁶

Our approach has limitations. We determined on-campus vs out-of-hospital locations based on whether the login device belonged to the medical center or was a personal computer. Consequently, if trainees exclusively used a personal computer while on-campus and never used a medical center computer, we would have captured this work done while logged into the EHR but would not have inferred on-campus work. Although nearly all trainees in our organization use medical center computers throughout the day, this might impact generalizability for programs where trainees use personal computers exclusively in the hospital. Our approach also assumes trainees will use the EHR at the beginning and end of their workdays, which could lead to underestimation of work hours in trainees who do not employ this practice. With regards to work done on personal computers, our heuristic required that at least one work activity (as denoted in Appendix Table 1) be included in the session in order for it to count as work. Although this approach allows us to exclude sessions where trainees might be reviewing charts exclusively for educational purposes, it is difficult to infer the true intent of chart review.

There might be periods of time where residents are doing in-hospital work but more than 5 hours elapsed between EHR user sessions. As we have started adapting this computational method for other residency programs, we have added logic that allows for long periods of time in the operating room to be considered part of a continuous workday. There also are limitations to assigning blocks of time to off-site clinics; clinics that are associated with after-hours work but use a different EHR would not be captured in total out-of-hospital work.

Although correlation with self-report was good, we identified clusters of inaccuracy. This likely resulted from our residency program covering three medical centers, two of which were not included in the data set. For example, if a resident had an off-site clinic that was not accounted for in AMiON, EHR-derived work hours might have been underestimated relative to self-report. Operationally leveraging an automated system for measuring work hours in the form of dashboards and other tools could provide the impetus to ensure accurate documentation of schedule anomalies.

CONCLUSION

Implementation of our EHR-derived work-hour model will allow ACGME residency programs to understand and act upon

trainee work-hour violations closer to real time, as the data extraction is daily and automated. Automation will save busy residents a cumbersome task, provide more complete data than self-report, and empower residency programs to intervene quickly to support overworked trainees.

Acknowledgments

The authors thank Drs Bradley Monash, Larissa Thomas, and Rebecca Berman for providing residency program input.

Disclosures: The authors have nothing to disclose.

References

1. Accreditation Council for Graduate Medical Education. Common program requirements. Accessed August 12, 2020. <https://www.acgme.org/What-We-Do/Accreditation/Common-Program-Requirements>
2. Accreditation Council for Graduate Medical Education. Resident/fellow and faculty surveys. Accessed August 12, 2020. <https://www.acgme.org/Data-Collection-Systems/Resident-Fellow-and-Faculty-Surveys>
3. Petre M, Geana R, Cipparrone N, et al. Comparing electronic and manual tracking systems for monitoring resident duty hours. *Ochsner J*. 2016;16(1):16-21.
4. Gonzalo JD, Yang JJ, Ngo L, Clark A, Reynolds EE, Herzog SJ. Accuracy of residents' retrospective perceptions of 16-hour call admitting shift compliance and characteristics. *Grad Med Educ*. 2013;5(4):630-633. <https://doi.org/10.4300/jgme-d-12-00311.1>
5. Dziorny AC, Orenstein EW, Lindell RB, Hames NA, Washington N, Desai B. Automatic detection of front-line clinician hospital shifts: a novel use of electronic health record timestamp data. *Appl Clin Inform*. 2019;10(1):28-37. <https://doi.org/10.1055/s-0038-1676819>
6. 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. <https://doi.org/10.1093/jamia/ocy145>
7. MedHub. Accessed April 7, 2021. <https://www.medhub.com>
8. AMiON. Accessed April 7, 2021. <https://www.amion.com>
9. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. Proceedings of the 9th Python in Science Conference. <https://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf>
10. Todd SR, Fahy BN, Paukert JL, Mersinger D, Johnson ML, Bass BL. How accurate are self-reported resident duty hours? *J Surg Educ*. 2010;67(2):103-107. <https://doi.org/10.1016/j.jsurg.2009.08.004>
11. Chadaga SR, Keniston A, Casey D, Albert RK. Correlation between self-reported resident duty hours and time-stamped parking data. *J Grad Med Educ*. 2012;4(2):254-256. <https://doi.org/10.4300/JGME-D-11-00142.1>
12. Drolet BC, Schwede M, Bishop KD, Fischer SA. Compliance and falsification of duty hours: reports from residents and program directors. *J Grad Med Educ*. 2013;5(3):368-373. <https://doi.org/10.4300/JGME-D-12-00375.1>
13. Shanafelt TD, Dyrbye LN, West CP. Addressing physician burnout: the way forward. *JAMA*. 2017;317(9):901. <https://doi.org/10.1001/jama.2017.0076>
14. Dziorny AC, Orenstein EW, Lindell RB, Hames NA, Washington N, Desai B. Pediatric trainees systematically under-report duty hour violations compared to electronic health record defined shifts. *PLOS ONE*. 2019;14(12):e0226493. <https://doi.org/10.1371/journal.pone.0226493>
15. Saag HS, Shah K, Jones SA, Testa PA, Horwitz LI. Pajama time: working after work in the electronic health record. *J Gen Intern Med*. 2019;34(9):1695-1696. <https://doi.org/10.1007/s11606-019-05055-x>
16. ResQ Medical. Accessed April 7, 2021. <https://resqmedical.com>