

BRIEF RESEARCH REPORT

The Practice of Emergency Medicine

Accuracy of physician self-estimation of time spent during patient care in the emergency department

Martin A. Reznik MD, MBA^{1,2} | Virginia Mangolds PhD, FNP-C¹ |
Kevin A. Kotkowski MD, MBA^{1,2} | Kian D. Samadian MD³ | James Joseph MD⁴ |
Celine Larkin PhD¹

¹Department of Emergency Medicine, UMass Chan Medical School, Worcester, Massachusetts, USA

²Department of Emergency Medicine, UMass Memorial Health, Worcester, Massachusetts, USA

³Department of Emergency Medicine, Massachusetts General Hospital & Brigham and Women's Hospital, Boston, Massachusetts, USA

⁴Department of Emergency Medicine, Alpert Medical School of Brown University, Providence, Rhode Island, USA

Correspondence

Martin A. Reznik, MD, MBA, Department of Emergency Medicine, University of Massachusetts Chan Medical School, UMass Memorial Health, Worcester, MA, USA.
Email: mreznek@hotmail.com

Funding and support: By JACEP Open policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist.

Abstract

Objective: Accurate measurement of physicians' time spent during patient care stands to inform emergency department (ED) improvement efforts. Direct observation is time consuming and cost prohibitive, so we sought to determine if physician self-estimation of time spent during patient care was accurate.

Methods: We performed a prospective, convenience-sample study in which research assistants measured time spent by ED physicians in patient care. At the conclusion of each observed encounter, physicians estimated their time spent. Using Mann-Whitney *U* tests and Spearman's rho, we compared physician estimates to actual time spent and assessed for associations of encounter characteristics and physician estimation.

Results: Among 214 encounters across 10 physicians, we observed a medium-sized correlation between actual and estimated time (Spearman's rho = 0.63, $p < 0.001$), and in aggregate, physicians underestimated time spent by a median of 0.1 min. An equal number of encounters were overestimated and underestimated. Underestimated encounters were underestimated by a median of 5.1 min (interquartile range [IQR] 2.5–9.8) and overestimated encounters were overestimated by a median of 4.3 min (IQR 2.5–11.6)—26.3% and 27.9% discrepancy, respectively. In terms of actual time spent, underestimated encounters (median 19.3 min, IQR 13.5–28.3) were significantly longer than overestimated encounters (median 15.3 min, IQR 11.3–20.5) ($p < 0.001$).

Conclusions: Physician self-estimation of time spent was accurate in aggregate, providing evidence that it is a valid surrogate marker for larger-scale process improvement and research activities, but likely not at the encounter level. Investigations exploring mechanisms to augment physician self-estimation, including modeling and technological support, may yield pathways to make self-estimation valid also at the encounter level.

Supervising Editor: Christian Tomaszewski, MD, MBA

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. JACEP Open published by Wiley Periodicals LLC on behalf of American College of Emergency Physicians.

1 | INTRODUCTION

1.1 | Background

Direct observation remains the gold standard for physician as well as nonphysician healthcare practitioner time and effort assessments, and reliable surrogate markers have yet to be identified.¹⁻³ Direct observation, however, is time consuming and cost prohibitive at scale,²⁻⁵ heightening the urgency for an alternative. We hypothesized that emergency physician self-reporting represented a potentially feasible and low-cost alternative to direct observation, but emergency physician accuracy in such reporting was unknown.

1.2 | Importance

Accurate measurement of time spent by physicians during patient care activities stands to inform emergency department (ED) efforts to improve quality, patient experience, provider wellness, and financial outcomes. In addition, improved understanding of physician resource needs in ED settings has potential to guide policy and advocacy efforts.

1.3 | Goals of this investigation

Our goal was to determine if emergency physicians were able to estimate accurately their time spent during patient care.

2 | METHODS

2.1 | Study design and setting

We performed a prospective, convenience-sample study in which trained research assistants (RAs) observed attending physicians during clinical shifts in 2 EDs: an urban, academic ED with approximately 65,000 annual adult patient encounters and an urban, general ED with approximately 43,000 annual patient encounters. The investigation was approved by the University of Massachusetts Chan Medical School Institutional Review Board.

2.2 | Selection of participants

Before the study period, we sent email communication describing the investigation to attending physicians ($n = 65$) in a single academic department of emergency medicine. The email described the study design as well as the institutional review board approval. Within the email, the physicians were assured that their participation would not interfere with their employment. Decisions to participate or opt-out would be kept confidential, known only to a study coordinator who had no role in physician employment status. Potential subjects also were instructed that in agreeing to participate in the study there would be

The Bottom Line

Accurate estimates, usually done by direct observation, of physician engagement in patient care activities may help improve emergency department efficiencies. In a prospective study of 214 patient encounters across 10 physicians, self-estimates by physicians were very accurate, underestimating median times by only 0.1 minutes, and thereby providing a potential surrogate measure.

no benefit to them such as additional scribe support or any other clinical or administrative assistance. A power calculation, assuming at least a correlation of 0.2 with 80% power at an alpha level of 0.05, called for 194 patient encounter observations, so a convenience sample of 10 physicians was planned with a goal of 20 patient encounters per physician. From the physicians not having opted out ($n = 63$), 10 subjects were randomly selected (by choosing every other physician on an alphabetic-order list of those who did not opt out), with the exception that, if a selected physician's schedule was anticipated not to align with RA schedules (eg, night-shift-only physicians, physicians working a low number of shifts at the study site EDs) or if a selected physician worked very few shifts without significant trainee presence, they were excluded ($n = 4$) and another was selected using the same methodology. The study coordinator oriented the 10 physician subjects to the investigation including instructions to provide their typical patient care and not to consciously track any time they spent in patient care activities. The physician subjects also were reminded that the RAs were to observe only and not perform any clinical care-related activities.

2.3 | Measurements

During a 12-month period, 4 trained RAs: 2 senior medical students and 2 experienced ED scribes (acting only as an RA with no scribing duties during observations), observed physician subjects during patient care in the ED. A single RA observed a single attending physician during a clinical shift. This was repeated over multiple shifts until at least 20 patient encounters were observed for each of the 10 physicians. Only complete patient encounters (patients not signed out to or from another ED provider) were included, and we excluded encounters co-managed with trainees (students and residents). RAs timed and recorded patient care-related activities in 15 sec increments and classified activities into the following categories: patient/family/surrogate interactions, ED staff interactions, consultant interactions, computer interface/documentation activities, procedure-related activities including consent and preparation, and other (including medical decision making—subjects were instructed to inform RAs if they were thinking about patient care outside of one of the other time engagement categories). If no activity was observed for 1 or more care activity categories during an observed encounter,

RAs recorded the time as 0 for those activities at the completion of the encounter. At the end of each complete patient encounter, the RA prompted the physician subject to estimate the total time in minutes they spent in care-related activities for that encounter.

For each encounter, the RA calculated the total time spent for each activity category as well as the total time spent overall. These data and the physician's self-reported total time estimate were recorded on a standardized reporting tool. The tool included patient identifiers to allow for subsequent patient-specific data abstraction. The RAs and/or study coordinator abstracted patient specific data from the electronic health record (EHR) (Epic, Epic Systems Corporation, Verona, WI) including date of service, age, gender, primary language, triage score from the Emergency Severity Index (ESI), disposition, and final diagnosis. For physician subjects, we recorded gender and years in practice as an attending physician (first year in practice = 1 year). The study coordinator assigned a unique identification number to each physician observed, so their identities were known only to RAs through direct observation and the study coordinator. The study coordinator also assigned unique identification numbers to the patients. Deidentified data were shared with investigators for analysis.

We performed an interrater reliability (IRR) assessment midway through the study. All 4 RAs simultaneously observed the primary investigator, an attending emergency physician, during a clinical shift. The RAs independently recorded time intervals following the methodology for the primary investigation.

2.4 | Outcomes

Our primary outcome was emergency physician accuracy in estimating their time spent on activities related to patients' clinical care. Secondary outcomes included encounter/patient characteristics associated with discrepancies between actual and estimated time spent on patient care.

2.5 | Analysis

For statistical analysis, we used SPSS version 28 (IBM, Armonk, NY). We calculated descriptive statistics and tested data for normality. Data for time taken and estimated time were positively skewed, so median/interquartile ranges and non-parametric tests were employed. For subject estimation accuracy assessment, we examined correlations between actual time and estimated time (both continuous variables). We also calculated a new variable to measure discrepancy, which was the estimated time minus the actual time. We used the discrepancy variable to create a categorical variable for under versus over-estimation: if the discrepancy value was negative, the estimate was categorized as an underestimate and if it was positive, it was categorized as an overestimate. We employed Mann-Whitney *U* tests to examine differences in median time between dichotomous groups and Spearman's rho to examine correlation between non-normally distributed continuous variables. We performed interrater compari-

son using 2-way mixed intraclass correlation (ICC) analysis. The RAs reported observations in 15-sec intervals, and we considered a "match" when RAs reported identical observations.

3 | RESULTS

3.1 | Characteristics of patient encounters and study subjects

From November 16, 2020 to November 23, 2021, 214 patient encounters were observed across 10 physicians (1–4 shifts per subject to reach goal of 20 encounters per subject). Characteristics of patients and physicians are in Table 1. Chief complaints among the encounters varied, but among the most common were chest pain, cough, and cellulitis.

3.2 | Main results

Physicians spent a median of 16.5 min (interquartile range [IQR] 12.1–23.9) total time per patient encounter. Across the patient care activity categories, physician subjects spent the most time interacting with patients/family/surrogates (median 6.5 min, IQR 4.3–10.0) and on computer-related activities (median 6.5 min, IQR 4.3–9.9). Physicians spent less time in communicating with ED staff (median 0.3 min, IQR 0.0–1.0), procedures/consent/gowning (median 0.0 min, IQR 0–1.5) and consults (median 0.0 min, IQR 0–1.5). Physicians estimated their total time spent per patient encounter at a median of 17.0 min (IQR 11.0–25.0).

We observed a medium-sized correlation between actual and estimated time (Spearman's rho = 0.63, $p < 0.001$) (Figure 1). An equal number of encounters were overestimated as underestimated, and across all encounters, physicians underestimated time spent by a median of 0.1 min. In the subset of underestimated encounters, subjects underestimated by a median of 5.1 min (IQR 2.5–9.8), and in the overestimated encounters, subjects overestimated by a median of 4.3 min (IQR 2.5–11.5). Underestimated encounters (median 19.3 min, IQR 13.5–28.3) were significantly longer than overestimated encounters (median 15.3 min, IQR 11.3–20.5; $p < 0.001$) in terms of actual time spent. Table 2 shows associations between encounter/patient characteristics and discrepancies between actual and estimated time.

3.3 | Interrater reliability

Within the IRR assessment, the 4 RAs observed 10 encounters. For total time, ICCs for absolute agreement were in the "excellent" range (1.000, 95% confidence interval [CI] 0.999–1.000). Among the activity subcategories, patient/family/surrogate interactions and computer interface/documentation, ICC for absolute agreement was "excellent" (1.000, 95% CI 0.999–1.000 and 0.999, 95% CI 0.998–1.000, respectively). The remaining activity subcategories were observed only in 3

TABLE 1 Patient encounter and physician characteristics.

Patients		n	Proportion/distribution
Gender	Female	112	52.3%
	Male	102	47.7%
Median age		40 years	IQR 28–56 years
Primary language	English	157	73.4%
	Spanish	32	15.0%
	Portuguese	10	4.7%
	Other	15	7.0%
ESI level	2	33	15.4%
	3	111	51.9%
	4	62	29.0%
	5	8	3.7%
Disposition	Admitted	33	15.4%
	Not admitted	181	84.6%
Physicians			
Gender	Female	3	30%
	Male	7	70%
Median years in practice		4.5 years	IQR 2.5–23 years

Abbreviations: ESI, Emergency Severity Index; IQR, interquartile range.

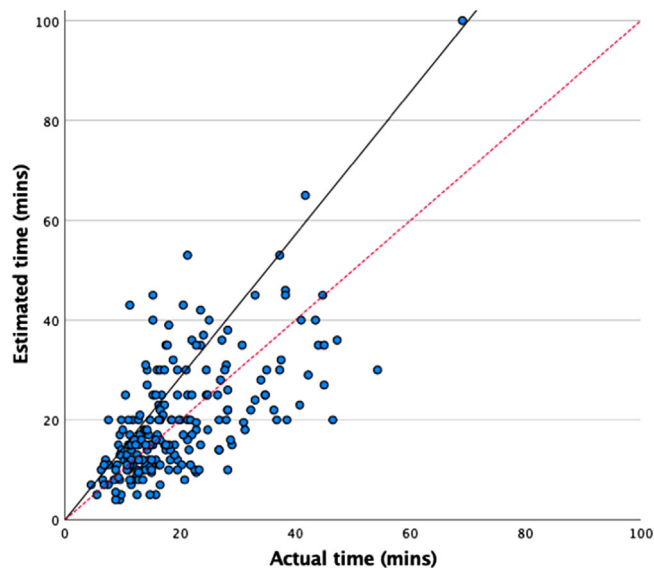


FIGURE 1 Scatterplot of the correlation between actual time taken and estimated time taken. *solid line shows actual correlation; dotted line shows reference line for perfect correlation.

or fewer of the encounters, so IRR could not be analyzed for those subcategories.

4 | LIMITATIONS

The physician subjects worked in a single academic department so it is possible that unmeasured local culture/practices may have affected

the results. Physician subjects were selected after an opt-out period and an RA alignment/feasibility assessment, but we believe it unlikely that these biased the results. The number of physicians observed limited our ability to assess physician-specific characteristics' associations with estimation discrepancy. However, physician subjects were chosen essentially at random, so it remains unclear if unmeasured physician-specific characteristics may have affected the results. Notably, there were fewer female than male physicians in our investigation. The proportion of women was lower than that of our department overall (~40%), but it aligned with reported proportions for board certified emergency physicians nationally.^{6,7} There were no ESI 1 encounters observed in our study. Simulation-based studies have shown that clinicians tend to underestimate time spent in resuscitation scenarios.^{8,9} We are not aware of any real-life studies in this regard, and in our study, there did not appear to be a trend related to ESI 2 through 5. It remains unclear if lack of ESI 1 encounters may have affected the results. Nonetheless, the distribution of ESI levels observed in our investigation was reflective generally of EDs nationally.¹⁰ As only complete patient encounters were included, we could not assess providers' ability to estimate time spent during sign-out activities or time spent caring for patients received in sign-out. Although we measured physician time per patient, it is important to note that our study was not designed to investigate time spent by physicians in ED care but rather to determine a physician's ability to accurately estimate the time spent. Although sufficiently powered to answer our study question, the number of patient encounters limited generalized conclusions related to actual time spent by emergency physicians. Physician-related activities that may have occurred after an encounter, including documentation, were not measured in our investigation. Physicians had scribes dur-

TABLE 2 Patient encounter characteristics association with discrepancies between physician estimated time and actual time.

		Median discrepancy in minutes (estimated minus actual)	IQR	Test statistic	p
Gender	Female	-0.3	-5.5 to +4.3	U = 5704	0.99
	Male	+0.3	-5.0 to +4.6		
Primary language	Non-English	-3.0	-9.1 to +3.8	U = 3423	0.009
	English	+0.3	-3.5 to +5.7		
Disposition	Admitted	-1.3	-6.0 to +8.3	U = 2999	0.97
	Discharged	+0.3	-5.0 to +4.3		
ESI level	2	-1.0	-5.6 to +3.9	H = 1.05	0.79
	3	+0.3	-5.5 to +7.8		
	4	+0.3	-4.8 to +3.5		
	5	-0.6	-6.0 to +2.3		
Age		Continuous	Variable	rho = 0.12	0.07

Abbreviations: ESI, Emergency Severity Index; IQR, interquartile range.

ing the shifts observed (unless a late notice call-out occurred of which we were not aware), so post-shift documentation was likely minimal. Using the primary investigator for the IRR analysis was done to facilitate alignment of RA scheduling; it was possible that the RAs had heightened vigilance during the IRR observations.

5 | DISCUSSION

Our investigation revealed mixed results regarding emergency physicians' accuracy in self-estimation of time spent during patient care. There was a moderate correlation of estimated and actual time spent, and remarkably, the overall median subject estimate was inaccurate by only 0.1 min—less than 1% of the median actual time spent. The latter finding however appeared to result from subjects over- and underestimating with similar frequency and similar magnitude: for underestimated encounters, subjects underestimated by 26.3% and for overestimated encounters, subjects overestimated by 27.9%. In aggregate, our observations implied that physician self-estimation of time spent during patient care may represent a viable surrogate for direct observation; however, there are important limitations. Although self-estimation was relatively accurate overall, on an individual encounter basis, estimates were fairly inaccurate, implying that self-estimation may be a valid surrogate for larger-scale process improvement and research activities that require accuracy only in aggregate, but it likely is not a valid surrogate at the individual encounter level.

With regard to the apparent inaccuracy at the encounter level, it is prudent to note that within our study population, we observed (1) underestimation occurring when actual time spent was longer and overestimation when actual time spent was shorter and (2) underestimation for encounters with patients for whom English was not their primary language. These observations open the possibility that

observed inaccuracy at the encounter level may have some element of underlying predictability. In other words, there may be underlying consistency in reporting inaccuracy that, if identified and characterized, may ultimately enable the use of physician self-reporting for smaller scale investigations or even at the encounter level. A study methodology similar to ours but multicenter, on a larger scale and including more patient encounter-related characteristics and physician-related characteristics may allow for the development of modelling that would enable physician self-estimation to offer value even at the encounter level.

Another possible avenue to improve validity of physician self-estimation at the encounter level is to train physicians to be more accurate in their self-assessments; however, we are not aware of investigations related to such efforts. One could envision physician self-reporting being used in combination with other modalities to generate better estimates. Radio frequency identification (RFID) has been effective in some time/motion studies and may be able to capture physician time at the patient bedside; however, some barriers to its use in health care have been cited.^{11,12} Although RFID may be able to accurately reflect a physician's time at a location, it would not provide sufficient information, in isolation, for activities such as work at the computer or conversations because it would not discriminate to which patient encounter to allocate the measured time. EHR systems could be audited to determine relevant information such as when a patient chart was opened by a physician. However, these data may lose accuracy if a chart is left open outside of active use. Combining RFID data to identify when a physician is at the computer and trackable EHR data could assist. However, neither of these electronic solutions are likely to enable the capture of live, interpersonal interactions. Strategic prompts might improve physician estimation accuracy to some degree, but prompting is likely to be intrusive and erode physician efficiency and wellness. Further investigation is warranted

to determine if a multi-modal approach might prove feasible and accurate.

In summary, physician self-estimation of time spent in patient care was accurate in aggregate, providing evidence that it is a valid surrogate marker at least for larger-scale process improvement and research activities. Investigations exploring mechanisms to augment physician self-estimation, including technological support and mathematical modeling, may yield pathways to make self-estimation valid at the encounter level.

AUTHOR CONTRIBUTIONS

Martin A Reznek conceived the study; Martin A Reznek and Virginia Mangolds designed the study; Virginia Mangolds, Kian D Samadian, and James Joseph supervised the conduct of the study and data collection; Virginia Mangolds managed the data, including quality control; Celine Larkin provided statistical advice on study design and analyzed the data; Martin A Reznek, Kevin A Kotkowski, and Celine Larkin drafted the manuscript, and all authors contributed substantially to its revision. Martin A Reznek takes responsibility for the paper as a whole.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

1. Kilpatrick K. Development and validation of a time and motion tool to measure cardiology acute care nurse practitioner activities. *Can J Cardiovasc Nurs*. 2011;21(4):18-26.
2. Bratt JH, Foreit J, Chen PL, West C, Janowitz B, de Vargas T. A comparison of four approaches for measuring clinician time use. *Health Policy Plan*. 1999;14(4):374-381. doi:10.1093/heapol/14.4.374
3. Zuckerman S, Merrell K, Berenson R, Mitchell S, Upadhyay D, Lewis R. *Research Report—Collecting empirical physician time data: piloting an approach for validating work relative value units*. Urban Institute; 2016. Available at: Accessed: December 30, 2022 https://www.urban.org/sites/default/files/publication/87771/2001123-collecting-empirical-physician-time-data-piloting-approach-for-validating-work-relative-value-units_0.pdf
4. Tipping MD, Forth VE, Magill DB, Englert K, Williams MV. Systematic review of time studies evaluating physicians in the hospital setting. *J Hosp Med*. 2010;5(6):353-359. doi:10.1002/jhm.647
5. Abdulwahid MA, Booth A, Turner J, Mason SM. Understanding better how emergency doctors work. Analysis of distribution of time and activities of emergency doctors: a systematic review and critical

appraisal of time and motion studies. *Emerg Med J*. 2018;35(11):692-700. Epub 2018 Sep 5. PMID: 30185505. doi:10.1136/emmermed-2017-207107

6. Agrawal P, Madsen TE, Lall M, Zeidan A. Gender Disparities in academic emergency medicine: strategies for the recruitment, retention, and promotion of women. *AEM Educ Train*. 2020;4(suppl 1):S67-S74. doi:10.1002/aet2.10414
7. ABEM News. Baren JM, Our workforce and our board must reflect our patients. November 6, 2019. Accessed September 9, 2022 <https://www.abem.org/public/news-events/news/2019/11/06/our-workforce-and-our-board-must-reflect-our-patients>
8. Brabrand M, Folkestad L, Hosbond S. Perception of time by professional health care workers during simulated cardiac arrest. *Am J Emerg Med*. 2011;29(1):124-126. doi:10.1016/j.ajem.2010.08.013
9. Trevisanuto D, De Bernardo G, Res G, et al. Time perception during neonatal resuscitation. *J Pediatr*. 2016;177:103-107. doi:10.1016/j.jpeds.2016.07.003
10. Chmielewski N, Moretz J. ESI triage distribution in U.S. emergency departments. *Adv Emerg Nurs J*. 2022;44(1):46-53. doi:10.1097/TME.0000000000000390
11. Wamba SA, Anand A, Carter L. A literature review of RFID-enabled healthcare applications and issues. *Int J Inf Manage*. 2013;33(5):875-891.
12. Haddara M, Staaby A. RFID applications and adoptions in healthcare: a review on patient safety. *Procedia Comput Sci*. 2018;138:80-88.

How to cite this article: Reznek MA, Mangolds V, Kotkowski KA, Samadian KD, Joseph J, Larkin L. Accuracy of physician self-estimation of time spent during patient care in the emergency department. *JACEP Open*. 2023;4:e12923. <https://doi.org/10.1002/emp2.12923>

AUTHOR BIOGRAPHY



Martin Reznek, MD, MBA, is an attending physician and Professor and Executive Vice Chair for Clinical Operations and Education in the Department of Emergency Medicine at University of Massachusetts Chan Medical School in Worcester, Massachusetts.