

## Opinion Paper

Carl T. Berdahl\* and David L. Schriger

# Study design and ethical considerations related to using direct observation to evaluate physician behavior: reflections after a recent study

<https://doi.org/10.1515/dx-2020-0029>

Received February 28, 2020; accepted April 19, 2020;

published online June 26, 2020

**Abstract:** In a recent study using direct observation of physicians, we demonstrated that physician-generated clinical documentation is vulnerable to error. In fact, we found that physicians consistently overrepresented their actions in certain areas of the medical record, such as the physical examination. Because of our experiences carrying out this study, we believe that certain investigations, particularly those evaluating physician behavior, should not rely on documentation alone. Investigators seeking to evaluate physician behavior should instead consider using observation to obtain objective information about occurrences in the patient-physician encounter. In this article, we describe our experiences using observation, and we offer investigators our perspectives related to study design and ethical questions to consider when performing similar work.

**Keywords:** documentation; emergency medicine; ethics; study design; observation.

## Introduction

In a recent study using direct observation of physicians in the emergency department, we demonstrated that physician-generated clinical documentation is vulnerable to error. In fact, we found that some sections of the medical record, such as the physical examination, consistently overrepresented physician actions—and the accuracy of specific data elements was only 53% [1].

---

\*Corresponding author: Carl T. Berdahl, MD, MS, Cedars-Sinai Medical Center, 8687 Melrose Ave G-562, West Hollywood, CA, 90069, USA, E-mail: [carl.berdahl@csmc.edu](mailto:carl.berdahl@csmc.edu). <https://orcid.org/0000-0002-4374-3280>

David L. Schriger: Department of Emergency Medicine, UCLA David Geffen School of Medicine, Los Angeles, CA, USA

Several factors are likely to contribute to inaccuracy in clinical documentation. For example, physicians may cut-and-paste outdated information [2]; use shortcuts that generate large blocks of text [3] that they neglect to customize; or add misleading information that facilitates reimbursement or defensive medicine [4, 5]. Unfortunately, even clinical documentation composed by professional scribes has error rates similar to that of physician-authored text [6]. Because of our experience quantifying documentation accuracy and because of our reviews of pertinent literature [7, 8], we believe that certain investigations, particularly those evaluating physician behavior, cannot rely on documentation alone.

Investigators seeking to evaluate physician behavior should consider using observation to obtain objective information describing occurrences during the patient-physician encounter. The main advantage of observation is that study personnel can be trained to identify specific events of interest as they occur in real time. In this article, we describe our experiences with using observation in our study, and we offer investigators our perspectives related to study design and ethical questions to consider when performing similar work.

## Study design

Our study objective was to quantify the accuracy of physician documentation in the emergency department setting. We believed that certain sections of the medical record, such as the physical examination, were particularly vulnerable to error because of anecdotal experience suggesting that physicians used shortcuts to generate large blocks of generic text. To determine the accuracy of documentation, we required two data sources: (1) physicians' documentation and (2) more objective records of physicians' actions in the room with patients. Obtaining documentation would be straightforward, but we faced a number of decisions as we strategized to obtain information about physicians' actions. In the sections that follow,

we present several study design decisions we made, our rationalizations for each decision, and relevant advantages and disadvantages.

## Simulation vs. observation of real-time patient care activities

Over the last several decades, researchers have used both simulation and observation to audit physician behavior [8–10]. Simulation offers the advantage of a controlled environment with time and space to reflect on results immediately after a patient-physician encounter. However, while some experts in simulation have explored ways to impose distractions and time pressures on research subjects [11, 12], it may not be possible to replicate the complex conditions of a busy emergency department. An alternative to simulation is to observe patient care activities as they unfold, which is a technique that experts in human factors advocate [8, 13, 14]. A third option is to use unannounced standardized patients, which may be an appropriate middle ground for clinical settings where patients have pre-arranged appointments [15, 16]. While there is precedent for using unannounced standardized patients in the emergency department [17], the American College of Emergency Physicians has a policy statement that opposes the use of “fictitious patients” because of unintended negative effects such as delays in care for patients with critical conditions [18]. To determine the best course of action, our research team held discussions with ED administrators. For this particular study, we collectively decided to proceed with direct observation rather than simulation or unannounced standardized patients after considering the aforementioned factors.

## Mode of observation

To obtain data describing physician actions, we considered the use of audio-recordings, video-recordings, real-time observation, or a combination of those modalities to generate data to describe physician actions during encounters with patients. Audio alone would be inadequate for our purposes, since we wanted to obtain data describing physicians’ observable actions related to the physical examination. Video might be sufficient, but we worried that patients would be less likely to participate, especially if they were seeking evaluation of conditions they considered embarrassing. Additionally, comprehensive evaluation with video would require two or more angles, which would be costly to set up and potentially

burdensome to troubleshoot in the event it malfunctioned. Finally, real-time observation would be labor-intensive, but it also had the advantage that it would operate within the norms and expectations of a teaching facility, since it was typical for medical students or other observers to shadow physicians in our emergency departments. Given these factors and our specific clinical environment, we settled on the use of real-time observation to obtain data describing the physical examination.

## Identity of observers

After deciding to utilize real-time observation to obtain physical examination data, we contemplated using several different types of observers. The most resource-intensive option would be to use physicians. Besides being costly, this option might induce the physicians under observation to alter their behavior [19]. Medical students were a second option, and using them seemed attractive because they would already have some clinical experience, including introductory training about how to perform physical examinations. However, we knew that we would require several dedicated individuals with flexible schedules to complete our goal of observing 240 encounters at two different institutions. For this reason, we decided to recruit undergraduate students to shoulder most of the burden.

## Recruitment and training of observers

Undergraduate students, particularly those planning to pursue careers related to health, have a reputation for being highly dedicated, competitive individuals [20]. Thus, we designed a multi-stage recruitment and training process leveraging competition, which ultimately led to the recruitment of a final cohort of observers who had demonstrated outstanding performance at every stage. The first stage had already been completed for us. A self-selected undergraduate student organization, known as the Emergency Medicine Research Associates Program, had already recruited a group of about 100 students to participate in various emergency medicine research projects (Inclusion in the group required submission of transcripts, letters of recommendation, and two rounds of interviews) [21]. For that group, we held an educational session where we described and demonstrated physical examination maneuvers. We then presented our study idea and asked for volunteers to attend future training sessions. Thereafter, we held two more 2-h sessions in which 25 students practiced with a physical examination checklist,

participated in a live examination, and completed a final, video-based examination. All participating observers scored higher than 97% on the video-based examination. Ultimately, we selected the ten undergraduate observers with the best scores to be members of our observation team.

When we commenced data collection for the study itself, both physician authors participated for the first 20 cases and compared our results with those of the trained undergraduate students. The initial interrater reliability among the group was promising (>97%), and we discussed disagreements immediately following the encounters. This led us to make some specific decisions about standardization of data collection process (e. g., Does an extremity examination require exposure of the entire extremity?), and we presented these recommendations to the team of observers to enhance the reliability of the entire team's work. Throughout the study, we continued to use multiple observers for each patient-physician encounter when it was feasible so that we could generate robust interrater reliability scores. Overall, two observers were present for 53 of the encounters. We audited their agreement scores periodically throughout the study period, and this led us to believe that our observational data were reliable. Details describing inter-observer agreement are available in the original research article and its accompanying supplement [1].

## **To deceive or not to deceive**

The final and perhaps most significant design consideration we faced was whether to deceive the study participants about the research goal. We feared that, if we described our true intent, they would alter their behavior and invalidate our results. Thus, we opted to deceive them: We told recruited physician study participants that we would be performing a time-and-motion study to understand how they performed histories and physical examinations. (In other words, we intentionally neglected to mention that we would be reviewing documentation.) In accordance with federal rules and regulations about protecting research subjects, we obtained specific institutional review board approval for the act of deception [22, 23]. Consequently, we were required to report the true intent of the study to participants after data were collected and give them a chance to withdraw participation.

Practically speaking, deceiving the physician participants meant that we had to maintain secrecy from everyone at the institution except core members of the research team. This proved challenging, because data collection occurred over the course of nearly 2 years.

During this period, several participating physicians asked what we were “really studying”, so it became clear that some were suspicious of our intent.

## **Protections for research subjects**

During the study's initial conceptualization, we had engaged experts in study design to ask how to best protect the participating physicians. The protections we chose included typical research subject protections, such as keeping the identities of participants anonymous and destroying data, as well as more atypical ones, such as keeping the name of the institutions unpublished and masking the exact interval of data collection. Furthermore, we obtained a Certificate of Confidentiality from the federal government to prevent our research team from being required to disclose our results in the event of any legal action. (Note that the National Institutes of Health automatically awards Certificates of Confidentiality for any work that they fund. However, our study was funded internally, so we had to complete a brief application to receive the certificate.) [24]

## **Ethics and consequences of deception**

### **Guidance regarding the acceptability of deception**

The ethics of deception in research have been debated for decades, and some stakeholders believe deception is never ethical [25]. However, most published guidance seems to suggest that deception is allowable under certain conditions. For example, the Belmont Report recommended that “research involving incomplete disclosure” should have “an adequate plan for debriefing subjects, when appropriate” [26]. Currently, deception in research is discouraged by societies representing groups of professionals including psychologists and sociologists. However, these groups still allow for deception in low-risk settings as long as deception is disclosed to participants and participants can withdraw participation after the disclosure [27, 28]. Importantly, ethicists maintain that disclosure in a debriefing does not eliminate the “wrong” of deception. They encourage researchers to explore other options besides deception, and they recommend that researchers electing to deceive research subjects need to justify this decision and explain why alternatives are inferior [22, 29–32].

## Our experience with disclosure of deception

We first disclosed the study's true intent and our results to physician research subjects during a conference call. At first, the research subjects were quiet, perhaps because they were considering what the implications of the results were (i. e., Am I a bad doctor? Are we all bad doctors? Is the system the problem?). A few spoke up and expressed concern that the results would harm the residency program's reputation. We validated their concerns and then offered our opinion that the results could influence national policy toward adopting new rules and regulations that did not incentivize expanding documentation. A few participants seemed to like that idea. We closed the call by offering the participants further opportunities to discuss the results on a one-to-one basis and explicitly presented the option that individuals could withdraw their data if they wished.

A few weeks after the phone call, we presented our results in departmental grand rounds, which stimulated both support and opposition from participants and non-participants. Over the ensuing several weeks, we spoke with participants in one-on-one settings. Several individuals expressed dismay that the research team betrayed them. Others expressed fear that publication of the results would lead to legal consequences for them or the institutions where the work was performed. Ultimately, three out of our 12 physician participants decided to have their data withdrawn prior to publication because of such fears. While we were disappointed with this result, we respected (and still respect) the decisions made by those individuals. After all, the initial consent process was indeed misleading, and they ultimately perceived that risk to their professional reputations would exist for several years after publication of the study.

## Reflections on downstream implications of deceiving research subjects

One author (CTB) is relatively junior within his organization. As such, he received comments from a few mentors during the study planning phase that discouraged him from completing the work because the ethics were controversial. Two such comments were: "This study is career suicide for you, since the results will be unpopular for your institution." and "How will you ever recruit physicians to participate in future work?" Our research group thought carefully about whether the work was important to advancing the practice of medicine and whether the junior

author's career would realistically experience adverse effects. In the end, we decided that our motivation to present this work was strong because of its potential impact, and we believed that negative career consequences were unlikely to be major—though it is indeed possible that recruitment of physicians for future work may indeed prove challenging because we deceived our physician colleagues on this occasion.

## Conclusions

Using real-time observation to evaluate physician behavior proved to be a challenging but worthwhile endeavor. We believe that our results should influence others to consider observation as a data source over physician documentation because documentation may not accurately represent physician behavior. We hope that describing our experience will enable other investigators to choose study design options that are best suited for their research goals and practice environments. Finally, we hope sharing our story will lead investigators to be thoughtful in how they protect their research subjects, particularly if they are considering deceiving them.

**Research funding:** This study was funded through contributions by Cedars-Sinai Medical Center, the UCLA National Clinician Scholars Program, and the Korein Foundation.

**Author contributions:** All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

**Competing interests:** Authors state no conflict of interest.

## References

1. Berdahl CT, Moran GJ, McBride O, Santini AM, Verzhbinsky IA, Schrager DL. Concordance between electronic clinical documentation and physicians' observed behavior. *JAMA Netw Open* 2019;2:e1911390.
2. Tsou AY, Lehmann CU, Michel J, Solomon R, Possanza L, Gandhi T. Safe practices for copy and paste in the EHR. Systematic review, recommendations, and novel model for health IT collaboration. *Appl Clin Inform* 2017;8:12–34.
3. Roman-Belmonte JM, De la Corte-Rodríguez H, Rodríguez-Merchan EC. Comparative analysis of two methods of data entry into electronic medical records: a randomized clinical trial (research letter). *J Eval Clin Pract* 2017;23:1478–81.
4. Coding Trends of Medicare Evaluation and Management Services. Office of the inspector general. Baltimore, MD: US Department of Health and Human Services; 2012. Contract No.: OEI-04-10-00180.

5. Studdert DM, Mello MM, Sage WM, DesRoches CM, Peugh J, Zapert K, et al. Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *J Am Med Assoc* 2005;293:2609–17.
6. Yan C, Rose S, Rothberg MB, Mercer MB, Goodman K, Misra-Hebert AD. Physician, scribe, and patient perspectives on clinical scribes in primary care. *J Gen Intern Med* 2016;31:990–5.
7. Tzeng HM. Using multiple data sources to answer patient safety-related research questions in hospital inpatient settings: a discursive paper using inpatient falls as an example. *J Clin Nurs* 2011;20:3276–84.
8. Catchpole K, Neyens DM, Abernathy J, Allison D, Joseph A, Reeves ST. Framework for direct observation of performance and safety in healthcare. *BMJ Qual Saf* 2017;26:1015–21.
9. Nestel D, Scerbo MW, Kardong-Edgren SE. A contemporary history of healthcare simulation research. *Healthcare simulation research*. Springer; 2019. pp. 9–14. <https://www.springer.com/gp/book/9783030268367>.
10. Trowbridge RL, Reilly JB, Clauser JC, Durning SJ. Using computerized virtual cases to explore diagnostic error in practicing physicians. *Diagnosis (Berl)* 2018;5:229–33.
11. Langenfeld J. In situ simulation. In: Carstens PK, Paulman P, Paulman A, Stanton MJ, Monaghan BM, Dekker D, editors. *Comprehensive healthcare simulation: mobile medical simulation*. Cham: Springer International Publishing; 2020. pp. 283–99.
12. Davison M, Kinnear FB, Fulbrook P. Evaluation of a multiple-encounter in situ simulation for orientation of staff to a new paediatric emergency service: a single-group pretest/post-test study. *BMJ Simul Technol Enhanc Learn* 2017;3:149–53.
13. Morgan L, Robertson E, Hadi M, Catchpole K, Pickering S, New S, et al. Capturing intraoperative process deviations using a direct observational approach: the glitch method. *BMJ Open* 2013;3:e003519.
14. Dixon-Woods M, Bosk C. Learning through observation: the role of ethnography in improving critical care. *Curr Opin Crit Care* 2010;16:639–42.
15. Zabar S, Hanley K, Stevens D, Murphy J, Burgess A, Kalet A, et al. Unannounced standardized patients: a promising method of assessing patient-centered care in your health care system. *BMC Health Serv Res* 2014;14:157.
16. Rethans JJ, Gorter S, Bokken L, Morrison L. Unannounced standardised patients in real practice: a systematic literature review. *Med Educ* 2007;41:537–49.
17. Zabar S, Ark T, Gillespie C, Hsieh A, Kalet A, Kachur E, et al. Can unannounced standardized patients assess professionalism and communication skills in the emergency department? *Acad Emerg Med* 2009;16:915–8.
18. Fictitious patients. American college of emergency physicians policy compendium. American College of Emergency Physicians, Irving, TX; 2020.
19. Goodwin MA, Stange KC, Zyzanski SJ, Crabtree BF, Borawski EA, Flocke SA. The Hawthorne effect in direct observation research with physicians and patients. *J Eval Clin Pract* 2017;23:1322–8.
20. Lin KY, Parnami S, Fuhrel-Forbis A, Anspach RR, Crawford B, De Vries RG. The undergraduate premedical experience in the United States: a critical review. *Int J Med Educ* 2013;4:26–37.
21. Emergency Medicine Research Associates. UCLA David Geffen School of Medicine; 2020. Available from: <https://organizations.dgsom.ucla.edu/emra/pages/>.
22. Kimmel AJ. Ethical issues in behavioral research: basic and applied perspectives. 2nd ed. Malden, MA: Blackwell Pub.; 2007. Vol. xxii, 405 p.
23. Basic HHS Policy for Protection of Human Research Subjects; 2018, p. 46. 45 C.F.R.
24. What is a certificate of confidentiality? National Institutes of Health, US Department of Health and Human Services; 2019. Available from: <https://grants.nih.gov/policy/humansubjects/coc/what-is.htm>.
25. Baumrind D. Research using intentional deception. Ethical issues revisited. *Am Psychol* 1985;40:165–74.
26. United States. National commission for the protection of human subjects of biomedical and behavioral research. The Belmont report: ethical principles and guidelines for the protection of human subjects of research. The Commission, Bethesda, MD; 1978. 20 p. <https://www.worldcat.org/title/belmont-report-ethical-principles-and-guidelines-for-the-protection-of-human-subjects-of-research/oclc/5403553?page=citation>.
27. Code of ethics. American Sociological Association, Washington, DC; 2018.
28. Ethical principles of psychologists and code of conduct. American Psychological Association, Washington, DC; 2017.
29. Tai MC-T. Deception and informed consent in social, behavioral, and educational research (SBER). *Tzu Chi Med J* 2012;24:218–22.
30. Wilson AT. Counterfactual consent and the use of deception in research. *Bioethics* 2015;29:470–7.
31. Miketta S, Friesen M. Debriefed but still troubled? About the (in) effectiveness of postexperimental debriefings after ego threat. *J Pers Soc Psychol* 2019;117:282–309.
32. Kimmel AJ, Smith NC, Klein JG. Ethical decision making and research deception in the behavioral sciences: an application of social contract theory. *Ethics Behav* 2011;21:222–51.