Simulation, Selection, and Mechanical Turk: Can Cases Presented Online Help Us Learn About Shared Decisionmaking and Medical Malpractice?

Gregory W. Ramsey, PhD, MSIA*

*Corresponding Author. E-mail: gregory.ramsey@morgan.edu.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2019.03.003

A **podcast** for this article is available at www.annemergmed.com.

SEE RELATED ARTICLE, P. 126.

[Ann Emerg Med. 2019;74:137-139.]

Schoenfeld et al[1] have demonstrated support for the importance of shared decisionmaking in settings in which patients can participate in the decisionmaking process. In summary, they performed a simulation study to examine how 3 levels of shared decisionmaking—none, brief, and thorough—affect the likelihood of patients' seeking some type of legal action for an adverse outcome resulting from treatment in an emergency department. This unique study used crowd-sourced simulation with nonphysician respondents as the tool for evaluating the hypothesis that increased shared decisionmaking, as opposed to no shared decisionmaking, results in a lower likelihood of complaints or litigation in the event of an adverse outcome. The authors point out that approximately 75% of emergency physicians will be named in a malpractice claim at some time in their career. The study is unique for 3 reasons: nonphysician respondents were used to evaluate actions of physicians, respondents completed surveys according to their assessments of clinical scenarios, and crowd sourcing was the mechanism used for recruiting respondents and conducting the survey.

This editorial is structured in the following manner: the positioning of this study relative to current simulation studies in medicine, an explanation of crowd sourcing and its use in medical simulations, and a discussion of some benefits, limitations, and concerns associated with crowd-sourced simulations within the field of medicine.

Schoenfeld et al present a set of vignettes, which are short, focused narratives. Properly designed vignettes draw participants into the portrayed scenarios and stimulate realistic decisions and responses. Previous studies have shown that simulations based on vignettes yield results nearly equivalent to those performed with standardized patients.[2,3] Simulations using standardized patients are limited by the number of patients who can be recruited; however, a vignette, once developed, has no marginal cost and thus few constraints on its use. Internet-enabled vignettes can be widely disseminated in written or multimedia format.

Most simulation studies focus on the physician as the decisionmaker. Recently, with increased emphasis on patient-centered care, simulation studies have evaluated patient behavior in the context of shared decisionmaking, including examination of patients' desires in the contexts of major surgeries and illnesses requiring inpatient care.[4,5] For instance, McKinstry[5] found that patients' preferences for shared decisionmaking was a function of the presenting problem, age, smoking status, and social class. Schoenfeld et al add to this stream of research by using crowd sourcing through an online labor market for performing a shared decisionmaking vignette–based simulation.

Crowd sourcing is the process of outsourcing a task, typically using Internet-enabled technology, to a distributed group of independent workers (the crowd) for a financial reward.[6] It is an efficient means for accomplishing a human intelligence task (ie, a task that computers are incapable of doing). Companies such as Amazon, Clickworker, and Toluna provide platforms for crowd-sourcing tasks. These platforms manage the recruitment of workers, collection of completed work, and payment for work. In the current study, the authors chose Amazon Mechanical Turk (MTurk; https://www.mturk.com/) as their crowd-sourcing platform for administering the survey used in the study. Researchers perform a power calculation to determine sample size and so establish the number of respondents who will be accepted for crowd-sourced work. Crowd-sourcing platforms allow researchers to specify the amount respondents will be paid, screen candidates for suitable characteristics for inclusion in a study, and review responses before accepting them. The fact that responses can be reviewed suggests that researcher-developed procedural controls need to be put in place to prevent biasing of results by arbitrary rejection of surveys (https://blog.mturk.com/tutorial-reconciling-worker-responses-280c96f1a696). MTurk provides an effective way to deliver vignettes in a randomized manner.

Medical researchers are taking advantage of easy access to large pools of people who complete surveys in exchange for payment.[7,8] I shall examine these opportunities with respect to MTurk. MTurk provides access to a participant pool of greater than 100,000 respondents.[9] The MTurk participant pool is more diverse than the typical undergraduate population[10] frequently used as study participants and more diverse than samples used in Internet-based, laboratory, or field experiments.[11] Indeed, for a study looking to survey individuals who may not be medically oriented, the online approach provides researchers access to a population who might otherwise be challenging and expensive to reach. Additionally, the level of attentiveness and multi-item scale reliabilities of study respondents are higher for MTurk compared with that for student samples[12]; recruiting study participants is faster than using direct recruitment methods,[13] and although compensation rates affect willingness to participate, they do not affect data quality. Horton et al[14] have shown that behavioral experiments conducted with online labor markets have internal and external validity comparable to that obtained with laboratory and field experiments. Data obtained with MTurk have been shown to be as reliable as data captured through direct population sampling through the Internet.[15] Finally, MTurk facilitates recruitment, payment, and data collection, allowing researchers to complete studies faster but with less access to respondents to verify information.[8]

So how is crowd sourcing relevant to emergency medicine? As Schoenfeld et al have demonstrated, it provides a potent means for conducting patient-centered research on topics of concern to the practice of emergency medicine. Their approach can be extended by making the vignettes richer, with animation, recorded actor portrayals, or adaptive delivery of material based on participants' responses during the completion of surveys. By shortening the research time through crowd sourcing, this approach may speed translation of behavioral research to practice.

Simulations like the one performed in this study also help formulate a theory of shared decisionmaking. Such a theory, cast as cognitive computational models, can be used to predict actions taken by decision agents, including physicians and patients.[16,17] By explicitly defining the cognitive processes associated with shared decisionmaking, the effects of changing inputs, such as the amount and type of information shared or the degree of equipping patients to make decisions, can be used to predict how outputs of the model will change. Ultimately, these computational models can be executed as computer programs and experiments can be conducted on a broad range of model inputs. Such insights into cognitive processes can potentially influence behavior, possibly through directed training or policies.

A hypothetical example of a shared decisionmaking computational model is one that receives inputs of patient condition, key words spoken by the patient and physician, outcome of treatment, and number of related treatment encounters. An algorithm derived from a combination of theory and data from previous patient-physician interactions then generates the expected probability that the patient will seek legal action. Possible outgrowths of this type of model include more accurate risk management planning and mitigation of risky interactions through staff training.

The work by Schoenfeld et al suggests that in the conditions covered by the vignettes, shared decisionmaking may lead to fewer complaints and lawsuits. This conclusion may break down on account of personal dynamics and communication. Actual narratives involve emotions that interfere with providers and patients rationally engaging in shared decisionmaking.[18-20] In bedside discussion, trust, rapport, and interactive dynamics play significant roles.[21-24] With fixed vignettes, the use of more words to explain options to a patient reflects increasing levels of shared decisionmaking. This simplification essential to the research may constrain the human dynamics found in choice of words, timing of statements, and tonality, all of which likely color the effectiveness of shared decisionmaking.[20]

Possible limitations to extending the conclusions include the inability to generalize the findings to additional scenarios in which shared decisionmaking may seem appropriate and biases among simulation participants such that smaller subpopulations are neither represented nor captured. This latter limitation is worth discussing further to highlight potential dangers in not accounting for subpopulations. The applicability or most appropriate approach to shared decisionmaking may be related to socioeconomics, culture, education, or race. If groups are not suitably represented as study participants, the ability to generalize to these groups may be impaired. Aside from bias in interpretation, a potentially more pernicious issue relates to algorithms that are developed as part of translating research findings to the clinical realm. The introduction of algorithmic bias, based on unrecognized initial sampling bias, may ultimately disenfranchise subgroups and minority populations. Identifying potential sources of bias during study design can prevent its occurrence.[25-28] If a known demographic profile needs to be achieved, it is possible on a crowd-sourcing platform to structure recruitment to meet the desired profile.[7]

In practical terms, the simulation study performed by Schoenfeld et al further supports the importance and potential protections and benefits of shared decisionmaking. From this study, it appears that the risks associated with shared decisionmaking are low and the potential benefits great. So, too, with appropriate attention to design, recruitment, and bias, may be the potential for crowd-sourced behavioral research in medicine.

Supervising editor: Stephen Schenkel, MD, MPP. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From Department of Information Science & Systems, E. Graves School of Business and Management, Morgan State University, Baltimore, MD.

Authorship: All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding and support: By *Annals* 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 author has stated that no such relationships exist.

REFERENCES

1. Schoenfeld EM, Mader S, Houghton C, et al. The effect of shared decisionmaking on patients' likelihood of filing a complaint or lawsuit: a simulation study. *Ann Emerg Med*. 2019;74:126-136.
2. Peabody JW, Luck J, Glassman P, et al. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA*. 2000;283:1715-1722.
3. Peabody JW, Luck J, Glassman P, et al. Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. *Ann Intern Med*. 2004;141:771-780.
4. Mansell D, Poses RM, Kazis L, et al. Clinical factors that influence patients' desire for participation in decisions about illness. *Arch Intern Med*. 2000;160:2991-2996.
5. McKinstry BJB. Do patients wish to be involved in decision making in the consultation? a cross sectional survey with video vignettes. *BMJ*. 2000;321:867-871.
6. Estellés-Arolas E, González-Ladrón-De-Guevara F. Towards an integrated crowdsourcing definition. *J Inf Sci*. 2012;38:189-200.
7. Cunningham JA, Godinho A, Kushnir V. Using Mechanical Turk to recruit participants for internet intervention research: experience from recruitment for four trials targeting hazardous alcohol consumption. *BMC Med Res Methodol*. 2017;17:156.
8. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? *Perspect Psychol Sci*. 2011;6:3-5.
9. Difallah D, Filatova E, Ipeirotis P. Demographics and dynamics of Mechanical Turk workers. In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM; February 5-9, 2018:135-143.
10. Buhrmester MD, Talaifar S, Gosling SD. An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspect Psychol Sci*. 2018;13:149-154.
11. Paolacci G, Chandler J, Ipeirotis PG. Running experiments on Amazon Mechanical Turk. *Judgment & Decision Making*. 2010;5:411-419.
12. Kees J, Berry C, Burton S, et al. An analysis of data quality: professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*. 2017;46:141-155.
13. Bartneck C, Duenser A, Moltchanova E, et al. Comparing the similarity of responses received from studies in Amazon's Mechanical Turk to studies conducted online and with direct recruitment. *PLoS One*. 2015;10:e0121595.
14. Horton JJ, Rand DG, Zeckhauser RJ. The online laboratory: conducting experiments in a real labor market. *Exp Econ*. 2011;14:399-425.
15. Levay KE, Freese J, Druckman JN. The demographic and political composition of Mechanical Turk samples. *SAGE Open*. 2016;6:1-17.
16. Carley KM; Computational approaches to sociological theorizing. *Handbook of Sociological Theory*. New York, NY: Springer; 2001:69-83.
17. Sun RJ. Introduction to computational cognitive modeling. In: *Cambridge Handbook of Computational Psychology*. Cambridge, UK: Cambridge University Press; 2018:3-19.
18. Wofford MM, Wofford JL, Bothra J, et al. Patient complaints about physician behaviors: a qualitative study. *Acad Med*. 2004;79:134-138.
19. Lerner JS, Li Y, Valdesolo P, et al. *Emotion and decision making*. 2015;66.
20. Epstein RM, Street RL Jr. Shared mind: communication, decision making, and autonomy in serious illness. *Ann Fam Med*. 2011;9:454-461.
21. Moore PJ, Adler NE, Robertson PA. Medical malpractice: the effect of doctor-patient relations on medical patient perceptions and malpractice intentions. *West J Med*. 2000;173:244-250.
22. Beckman HB, Markakis KM, Suchman AL, et al. The doctor-patient relationship and malpractice: lessons from plaintiff depositions. *Arch Intern Med*. 1994;154:1365-1370.
23. Thom DH, Campbell B. Patient-physician trust: an exploratory study. *J Fam Pract*. 1997;44:169-177.
24. Roter D. The patient-physician relationship and its implications for malpractice litigation. *J. Health Care L. & Pol'y*. 2006;9:304-314.
25. Kirkpatrick K. Battling algorithmic bias: how do we ensure algorithms treat us fairly? Communications of the ACM. 2016;59:16-17.
26. Lepri B, Oliver N, Letouzé E, et al. Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*. 2018;31:611-627.
27. O'Neil C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Broadway Books; 2017.
28. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med*. 2018;15:e1002689.