



Serious games may improve physician heuristics in trauma triage

Deepika Mohan^{a,b}, Baruch Fischhoff^{c,d,1}, Derek C. Angus^a, Matthew R. Rosengart^b, David J. Wallace^{a,e}, Donald M. Yealy^e, Coreen Farris^f, Chung-Chou H. Chang^a, Samantha Kerti^a, and Amber E. Barnato^g

^aDepartment of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA 15261; ^bDepartment of Surgery, University of Pittsburgh, Pittsburgh, PA 15213; ^cDepartment of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213; ^dInstitute for Politics and Strategy, Carnegie Mellon University, Pittsburgh, PA 15213; ^eDepartment of Emergency Medicine, University of Pittsburgh, Pittsburgh, PA 15213; ^fRAND Corporation, Pittsburgh, PA 15213; and ^gThe Dartmouth Institute, Geisel School of Medicine at Dartmouth, Lebanon, NH 03766

Contributed by Baruch Fischhoff, July 11, 2018 (sent for review March 29, 2018; reviewed by Arthur S. Elstein and Avery Nathens)

Trauma triage depends on fallible human judgment. We created two “serious” video game training interventions to improve that judgment. The interventions’ central theoretical construct was the representativeness heuristic, which, in trauma triage, would mean judging the severity of an injury by how well it captures (or “represents”) the key features of archetypes of cases requiring transfer to a trauma center. Drawing on clinical experience, medical records, and an expert panel, we identified features characteristic of representative and nonrepresentative cases. The two interventions instantiated both kinds of cases. One was an adventure game, seeking narrative engagement; the second was a puzzle-based game, emphasizing analogical reasoning. Both incorporated feedback on diagnostic errors, explaining their sources and consequences. In a four-arm study, they were compared with an intervention using traditional text-based continuing medical education materials (active control) and a no-intervention (passive control) condition. A sample of 320 physicians working at nontrauma centers in the United States was recruited and randomized to a study arm. The primary outcome was performance on a validated virtual simulation, measured as the proportion of undertriaged patients, defined as ones who had severe injuries (according to American College of Surgeons guidelines) but were not transferred. Compared with the control group, physicians exposed to either game undertriaged fewer such patients [difference = -18%, 95% CI: -30 to -6%, $P = 0.002$ (adventure game); -17%, 95% CI: -28 to -6%, $P = 0.003$ (puzzle game)]; those exposed to the text-based education undertriaged similar proportions (difference = +8%, 95% CI: -3 to +19%, $P = 0.15$).

heuristics | triage | serious games | decision making | judgment

Americans can expect to experience at least one meaningful medical diagnostic error in their lifetime (1). Often, those errors occur for reasons outside individual physicians’ control, including poorly designed healthcare systems, imperfect health information technology, and the complexity of medicine (2). However, physicians’ heuristic judgment is a source as well (3). Heuristics are mental shortcuts that often produce valid judgments but can lead to errors in atypical or rare events. Because they reflect natural processes, heuristics are not easily, or even productively, replaced. However, they might be improved if people could be trained to use them more effectively. We report results of a randomized controlled trial of two interventions designed to do just that, using simulated versions of the diagnostic challenges that can face emergency department (ED) physicians in clinical practice.

Severely injured patients benefit from treatment at trauma centers (4–6). Getting patients to the most appropriate hospital requires accurate assessment of their injuries. That triage can occur either in the field or after physicians have evaluated patients at a nontrauma center, where half of all severely injured patients are initially taken. We use “undertriage” to describe severely injured patients who are not appropriately transferred

from a nontrauma center to a trauma center (more specifically, a level I/II center accredited by the American College of Surgeons to provide definitive care for severe injuries).

Transfer decisions occur under stressful, time-pressured conditions that invite, and perhaps require, heuristic thinking. Physicians must draw on their clinical training and experience when interpreting a patient’s presenting condition, in light of whatever is known about that individual’s general health state and personal circumstances. Although trauma triage is often routine, some cases have great, but not always apparent, uncertainty. Moreover, feedback loops are often weak for emergency medicine physicians, who do not routinely get follow-up information on patient outcomes, particularly for those transferred to another facility for definitive management or aftercare (7). As evidence of these challenges, diagnostic errors persist despite decades of continuing medical education efforts (5–11).

Given the nature of trauma triage, the most relevant heuristic is representativeness, which entails judging the subjective probability of an event by how well it represents the archetype of a situation (12). In the case of trauma triage, that would mean asking how much a patient’s condition “looks like” one severe enough to require transfer to a trauma center. Judgment by representativeness could lead physicians astray if a patient’s presentation did not reflect a severe injury archetype. That could

Significance

Americans can expect to experience at least one meaningful diagnostic medical error in their lifetime. One plausible source of those errors is physicians’ reliance on heuristics that are generally useful but can fail in diagnostically challenging situations. Based on previous research and clinical experience, we identified heuristics that might cause diagnostic errors in trauma triage. We sought to improve physicians’ heuristic judgment by providing simulated experience with two “serious” video games. In a randomized controlled trial, both games had positive effects, whereas equivalent exposure to traditional medical education had none. By complementing physicians’ natural ways of thinking, such simulated experiences might transfer to actual triage and other high-pressure decisions.

Author contributions: D.M., B.F., D.C.A., M.R.R., D.M.Y., C.F., and A.E.B. designed research; D.M., D.C.A., M.R.R., D.J.W., D.M.Y., C.F., and A.E.B. performed research; C.-C.H.C. and S.K. contributed new reagents/analytic tools; D.M., B.F., D.C.A., M.R.R., D.J.W., D.M.Y., C.F., C.-C.H.C., S.K., and A.E.B. analyzed data; and D.M., B.F., D.C.A., M.R.R., D.J.W., D.M.Y., C.F., C.-C.H.C., S.K., and A.E.B. wrote the paper.

Reviewers: A.S.E., University of Illinois, Chicago; and A.N., University of Toronto.

The authors declare no conflict of interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence should be addressed. Email: baruch@cmu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1805450115/-DCSupplemental.

Published online August 27, 2018.

happen if the injury itself was not obvious (e.g., multiple rib fractures) or the patient was vulnerable to future complications (e.g., frail).

We created two interventions intended to help physicians recalibrate their heuristics in ways that would carry over to clinical practice. The design of those interventions faced two challenges. One was creating sufficiently immersive conditions for participants to absorb the lessons. We addressed that challenge by drawing on two theories of immersive learning. One was the theory of narrative engagement, according to which stories help people achieve active mastery of decision principles that can be recalled when related situations arise (13). We applied it here with an adventure video game in which players solved a mystery (involving a missing grandfather), while role-playing a physician faced with trauma triage decisions. An initial version of this game provided preliminary evidence of its efficacy (14). The second theory was that of analogical reasoning, according to which structured case comparisons can be an effective way to train people to master and apply decision principles (15). We applied it here with a newly developed puzzle game, in which players triaged sets of trauma patients, created to illustrate diagnostic principles. We chose this strategy in response to comments from some physicians in the previous study, who reported that the adventure game narrative was distracting and that they preferred a more cognitive approach.

The second design challenge was identifying important, but potentially overlooked, features of trauma cases that the training would try to make part of severe injury archetypes, thereby leading to more valid judgment by representativeness (16). Our approach combined clinical evidence and experience. Specifically, we extracted decision principles from the triage guidelines of the American College of Surgeons (*SI Appendix, Table S1*), which we adjusted for recently published research on the effects of age and frailty on outcomes after injury (17, 18). A multidisciplinary group of five emergency medicine physicians and five trauma surgeons then identified cues that (*i*) should evoke each principle and (*ii*) were either readily seen (i.e., “representative”) or often overlooked (i.e., “nonrepresentative”), based on their clinical experience. We designed the games to communicate these nonrepresentative decision principles. The games provided both outcome feedback (how accurate the diagnostic judgments were) and process feedback (how poor judgments could be corrected). If successful, they should have the greatest impact with nonrepresentative cases. However, undertriage occurs with representative cases as well. For example, Lale et al. (9) found that 90% of Chicago-area patients with penetrating injuries (a feature representative of severe injury) taken to nontrauma centers were kept at that facility and not transferred to a higher level of care. As a result, successful interventions could improve performance all around.

We conducted a randomized trial comparing these two interventions with an “active” control condition, using widely accepted text-based educational materials, and with a “passive” control condition, with no intervention. Participants were a convenience sample of emergency medicine physicians, responsible for triage decisions at nontrauma centers, recruited at a national meeting. Those receiving one of the three interventions agreed to use it for 2 h and then complete the virtual simulation that served as our outcome assessment measure. Those in the passive control condition agreed to complete just the simulation, which had been validated in previous research (19). The simulation stimuli, procedures, and user interface were all different from those of the interventions, making it relatively free of shared method variance (20). We hypothesized that both games would be superior to the text-based training, which, in turn, would be better than no intervention at all. Secondarily, we hypothesized that the interventions would have greater impact

for the nonrepresentative cases, compared with the representative ones, and for participants who reported greater engagement.

Results

Participant Characteristics. We recruited 320 board-eligible or -certified emergency medicine physicians attending the annual scientific meeting of the American College of Emergency Physicians on October 29, 2017. Eligible physicians made transfer triage decisions for adult trauma patients in the United States, and therefore worked primarily at nontrauma or level III/IV trauma centers. We excluded physicians who worked only at level I/II trauma centers, managed only pediatric patients, or worked outside the United States. Among those recruited, 268 (84%) began and 257 (80%) completed the outcome assessment portion of the study protocol by November 30, 2017, when the study closed (Fig. 1). Those who completed the demographics section reported mean experience of 8 y (SD = 9.7), 239 (91%) had completed an emergency medicine residency, and 201 (75%) worked only at nontrauma centers (sample details are provided in *SI Appendix, Table S2*).

Use of the Interventions: Adherence, Usability, Enjoyment, and Engagement. We randomized eligible physicians to four equal groups: (*i*) adventure video game (narrative engagement intervention), (*ii*) puzzle video game (analogical reasoning intervention), (*iii*) text-based education (active control), or (*iv*) no intervention (passive control). The interventions’ content and development are described in *Materials and Methods*.

Physicians randomized to one of the three interventions received a mobile tablet (iPad; Apple, Inc.) with that intervention preloaded. They were asked to use it for at least 2 h. Members of all four groups completed a virtual simulation, used as outcome assessment, and some additional questions, provided online. Participants in all groups answered questions about demographics, educational background, and practice environment. Those who received an intervention also answered questions about how long they had used it, rated it in terms of how usable and enjoyable it was (on five-point Likert scales anchored at “not at all” and “very”), and completed a 12-item narrative engagement scale (21). Participants completed the study protocol at their convenience, within 1 mo of enrollment.

Physicians reported spending equal amounts of time on the three interventions. They rated the games as less usable than the text-based applications, but equally enjoyable. Those using the games reported less narrative understanding, but greater attentional focus and emotional engagement, than did those using the text-based applications.

Effect of the Interventions on Diagnosis. The primary trial outcome was performance on the virtual simulation, measured as the proportion of undertriaged cases (severely injured patients not transferred to a trauma center as recommended by clinical practice guidelines) (22). The simulation included 10 cases: four severely injured patients (two representative and two nonrepresentative), two minimally injured patients, and four critically ill nontrauma cases (*Materials and Methods*). Table 1 describes them. We scored participants’ disposition decision (transfer, admit, or discharge) for each of the four severely injured trauma cases, categorizing patients who died before a disposition decision as having been transferred to avoid penalizing physicians for decisions made during initial patient resuscitation. These decisions are shown in detail in *SI Appendix, Table S3*. Performance in the passive control group is roughly that observed in clinical settings in the United States, where undertriage ranges from 55 to 80% of severe injury cases (8, 10, 23).

We assessed the effect of the interventions on undertriage with linear regression models, using multiple imputation to account for missing responses. Physicians randomized to either game undertriaged a lower percentage of patients than did those

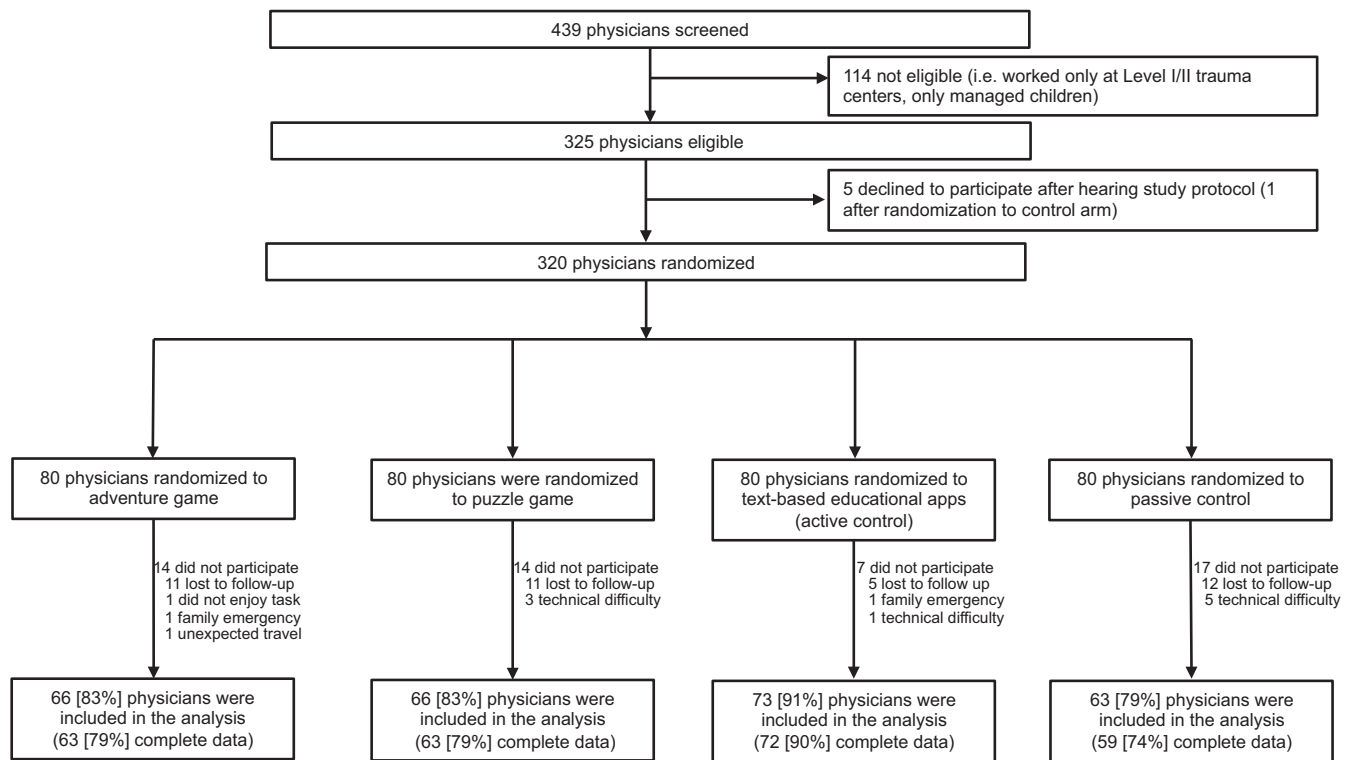


Fig. 1. Flow diagram of screening and allocation of study participants.

in the (no-intervention) passive control group [−18%, 95% CI: −30 to −6%, $P = 0.002$ (adventure); −17%, 95% CI: −28 to −6%, $P = 0.003$ (puzzle)]. Physicians who used the text-based applications performed similar to those in the passive control group (+8%, 95% CI: −3 to +19%, $P = 0.15$). Details are provided in *SI Appendix, Table S4*.

In planned secondary analyses, we analyzed undertriage separately for representative and nonrepresentative cases (those that did

and did not correspond to archetypes of severely injured patients). For nonrepresentative cases, exposure to both games reduced undertriage compared with passive control [−28%, 95% CI: −43 to −14%, $P < 0.001$ (adventure); −18%, 95% CI: −32 to −4%, $P = 0.01$ (puzzle)]. For representative cases, exposure to the puzzle game significantly reduced undertriage compared with passive control (−16%, 95% CI: −28 to −4%, $P = 0.007$), but exposure to the adventure game did not (−8%, 95% CI: −20 to +4%, $P = 0.21$).

Table 1. Description of cases included on virtual simulation

Case description	Class of case
Patient is a 65-y-old female s/p MVC with aortic transection and bilateral lower extremity fractures; presents with hypotension and arrests if not resuscitated	Severely injured: representative case
Patient is a 36-y-old male s/p GSW to abdomen with liver laceration; presents with hypotension and arrests if not resuscitated	Severely injured: representative case
Patient is an 80-y-old female s/p ground level fall with multiple rib fractures	Severely injured: nonrepresentative case
Patient is a 70 y-old male s/p fall down steps with pelvic fracture and intraparenchymal hemorrhage	Severely injured: nonrepresentative case
Patient is an 18-y-old female s/p bicycle collision with closed humerus fracture	Minimally injured
Patient is an 81-y-old male s/p MVC with no injuries but NSTEMI	Minimally injured
Patient is a 60-y-old female with subarachnoid hemorrhage and hypertensive emergency; develops altered mental status if not treated with anti-hypertensives within 5 min of arrival (game time)	Nontrauma case: critically ill
Patient is a 46-y-old female with sepsis; presents with hypotension and tachycardia on arrival, and develops worsening shock if not resuscitated within 5 min of arrival (game time)	Nontrauma case: critically ill
Patient is an 83-y-old male with CHF exacerbation and respiratory failure; arrests and dies if not intubated (or started on noninvasive ventilation) within 5 min of arrival (game time)	Nontrauma case: critically ill
Patient is a 69-y-old female with GI bleeding and hemorrhagic shock; arrests and dies if not resuscitated within 5 min (game time) of arrival	Nontrauma case: critically ill

CHF, congestive heart failure; GI, gastrointestinal; GSW, gunshot wound; MVC, motor vehicle collision; NSTEMI, nonST-segment elevation myocardial infarction; s/p, status post.

With the traditional text-based intervention, performance was similar to that of passive control for both types of cases. Responses to the narrative engagement scale were unrelated to performance.

Finally, we performed three sensitivity analyses testing the robustness of our estimates. We reran the regressions (*i*) excluding participants for whom we did not have complete case data to assess our imputation procedures; (*ii*) excluding participants who had previously worked at level I/II trauma centers, and hence might have been more proficient than physicians who had worked only at nontrauma centers; and (*iii*) reclassifying cases where patients died during the simulation as “not transferred,” thereby applying a more stringent evaluation criterion. All three sensitivity analyses produced results similar to those of the main analysis.

Discussion

Trauma triage occurs in conditions conducive to heuristic inference. Time is short, information is limited, and prognoses are uncertain. Despite concerted efforts to improve medical systems and training, between 55% and 80% of patients with severe injuries who present initially to nontrauma centers are not transferred to a higher level of care (8–11), contributing to an estimated 30,000 preventable deaths each year (24), disability, pain, and workforce dropout (4, 5, 25).

We developed and tested two interventions designed to reduce such undertriage by improving physicians’ use of a heuristic particularly suited to such situations: judgment by representativeness. In trauma triage, physicians relying on representativeness would judge an injury as severe (and requiring transfer) to the extent that it fit an archetype of such cases, given what physicians believe about the injury and the patient.

Drawing on medical records, professional guidelines, clinical experience, and a panel of experts, we identified cases where severe injury was representative of the underlying process (e.g., gunshot wounds) and where it was not (e.g., falls). We then created two “serious” video game interventions, each providing the kind of structured feedback and immersive experience needed for active mastery of complex principles. The two games embodied different design philosophies: narrative engagement and analogical reasoning. Their impact was compared with that of a traditional text-based intervention, based on material widely used in continuing medical education, and with that of a no-intervention control group, in a randomized controlled trial. Participants were physicians working in nontrauma centers, recruited at a major medical meeting. Their performance was assessed as undertriage on a validated simulation embodying another design philosophy. Participants were asked to spend at least 2 h on the interventions. The simulation took ~45 min.

We found that both game interventions reduced undertriage on the simulation, compared with the control condition, whereas the text-based intervention did not. The narrative game had greater improvement with the nonrepresentative cases than with the representative ones. The puzzle game was equally effective with both, and the text-based intervention was equally ineffective. The pattern of results was robust to several sensitivity analyses. The effect size for the adventure game was consistent with results from an earlier study (14), where it was compared with the text-based intervention (here: 27%, 95% CI: 16–38%, $P < 0.001$; there: 11%, 95% CI: 5–16%, $P < 0.001$). The earlier study asked participants to use their intervention for 1 h (rather than 2 h) and did not have a no-intervention (passive) control group.

Our text-based intervention was modeled on Advanced Trauma Life Support (ATLS), the gold standard for continuing medical education in trauma. Although well-tolerated by physicians, this active control showed no performance improvement, suggesting that users may have extracted as much as they could from that material already. We therefore speculate that physicians responsible for triage at nontrauma centers could benefit

from interventions grounded in other theoretical principles, like the serious games evaluated here, as adjuncts to existing quality improvement efforts.

Performance on even an intensive simulation is, of course, no guarantee of similar performance in clinical practice. Participants in the trial were a convenience sample of physicians, recruited at a national meeting. Although their demographic characteristics match those of practicing emergency medicine physicians (26), they may have disproportionately included physicians eager to learn (and perhaps able to do so). Another limitation introduced by our sampling frame was that it included some physicians who had experience working at level I/II trauma centers; hence, they might have been more proficient than those who had worked only at nontrauma centers. A post hoc sensitivity analysis found similar results when these participants were excluded. A third limitation of our study was that we relied on participants’ self-reports to measure time spent on the intervention. Because the commercial applications used in the active control did not allow assessing usage directly, self-reports were the only way to compare all groups. We had planned to measure actual usage on the two games to validate those participants’ self-reports. Unfortunately, a programming bug corrupted some of the data. Among participants for whom we had reliable data, we found a weak correlation ($r = 0.3$) between self-reported and objectively measured time spent on the game. However, we have no reason to believe that participants differed systematically across the groups in reporting their effort. A fourth limitation of the study was that the simulation included much higher rates of severe injury cases (and diagnostically difficult, nonrepresentative ones) than in actual practice (27), so as to have enough to observe. Fifth, the study was not designed to evaluate the effect of the interventions on overtriage (the proportion of patients with minor injuries transferred to trauma centers). Finally, we have no direct evidence regarding the durability of the treatment effect. However, in an earlier study comparing the adventure video game and the test-based educational applications, we found that performance differences persisted at six-month follow-up (17%, 95% CI: 9–25%, $P < 0.001$) (14).

Despite these limitations, there are theoretical reasons to hope that these results will generalize, whereas other interventions have not (28). One reason is that changes in heuristic reasoning may be more robust than changes that require physicians to retain new material in active memory (29, 30). A second reason is the independent support for the design philosophies underlying the two game interventions, as complements to text-based training (12, 13, 31). A third is our finding that physicians make similar judgments, in aggregate, on the simulation and in clinical practice (19). A fourth is that we operationalized the representativeness heuristic with a design process that drew on expert judgment, professional guidelines, and medical records.

Generalizing beyond the current context, theoretically based “serious games” might reduce costly errors in other settings where highly trained personnel work under intense pressure. Such interventions have the native advantages of games, including prompt and unambiguous feedback; feed-forward of explanatory principles, supporting users’ mental models and pattern recognition skills; scalability to large user populations; and adaptability to changing conditions (32–34). To realize that potential, we believe that all elements of the present design process are needed: subject matter expertise, familiarity with work environments, structured implementation of theoretical principles, professional design, and extensive pretesting. The potential return on that investment might be estimated by multiplying the reduction in undertriage found here by the number of such cases and their cost to patients and society.

Materials and Methods

Experimental Design. We conducted a randomized trial comparing two interventions designed to improve heuristic inference (narrative engagement

and analogical reasoning), an active control (text-based education), and a passive control (no intervention). Our primary hypothesis was that physicians exposed to either game-based intervention would undertriage fewer patients than would physicians exposed to the text-based applications, who would, in turn, undertriage fewer cases than physicians in the passive control arm.

Participants.

Randomization and blinding. Participants were recruited at a booth in the exhibition area of a national emergency medicine meeting. They were eligible if they worked in the EDs of nontrauma centers. We randomized them to one of four equal groups using a scheme built in Stata 13.0 (StataCorp), with random block sizes of four or eight: (i) adventure video game, (ii) puzzle video game, (iii) text-based education (active control), or (iv) no intervention (passive control). After registering a participant, study personnel obtained the intervention assignment from a central database. Although we could not maintain blindness after allocation, we masked condition assignment during the analyses. **Study protocol.** Participating physicians received written instructions. Those randomized to one of the three interventions received a mobile tablet (iPad; Apple, Inc.) with the intervention preloaded. They were asked to spend at least 2 h on it before completing the virtual simulation and answering final questions online, which they were told would take ~1 h. Passive control participants completed only the online outcome assessment and questions. Participants could complete the study at their convenience, within 1 mo of enrollment. They received reminder emails at weekly intervals and a telephone call in week 3 from research personnel blinded to group assignment.

We used a prepaid, wage-based financial incentive to increase response rates (35, 36). Physicians randomized to an intervention kept the iPad as an honorarium (approximate value of \$300); those in the passive control arm received a \$100 Amazon gift card when they enrolled.

Interventions.

Game-based interventions. We developed two interventions, drawing on different behavioral learning theories, to help players improve their use of heuristics in games designed to capture some of the intensity of clinical practice. Half of the patients with serious injury in both games had nonrepresentative severe injuries, defined as ones that do not fit an archetype for requiring transfer to a trauma center (22). An example of a representative case is a patient with penetrating wounds to the chest and torso. An example of a nonrepresentative case is a frail patient for whom complications are more likely.

Night Shift: Adventure video game based on narrative engagement. *Night Shift* relies on narrative engagement with compelling stories (13). Players take on the persona of Andy Jordan, a young emergency medicine physician who moves home after the disappearance of his estranged grandfather and takes a job in the local ED. The game centers on a series of trauma patients who arrive in the ED with severe injuries, some representative and some not. As players make their diagnoses, they gain experience with the consequences of undertriage, as nonrepresentative patients return with complications. Players must find solutions to these patients' deteriorating clinical condition and explain them to in-game characters (e.g., family members, consultants). These characters' responses sometimes draw attention to diagnostically relevant contextual cues. They also highlight the implications of undertriage, so as to evoke emotional responses making the feedback more memorable. Along the way, players learn more about Andy's grandfather.

The design incorporated three features intended to enhance players' engagement:

- i) The medical component was embedded in the mystery of Andy's grandfather's disappearance, hoping to increase players' empathy for Andy and the impact of the feedback he receives.
- ii) Patients arriving at the ED included several with representative severe injuries who decompensated shortly after arrival. Resuscitating them required players to participate in role-playing efforts designed to increase their immersion.
- iii) Several diagnostically challenging nontrauma cases (e.g., brucellosis) were meant to stimulate player interest, curiosity, and challenge.

Shift: The Next Generation: Puzzle video game based on analogical encoding. *Shift: The Next Generation* relies on analogical reasoning (15), delivered through a puzzle game, to help players identify relevant contextual cues. One author [D.M. (trauma surgery)] created a bank of cases, based on clinical data, exemplifying decision principles that physicians should use to triage patients (SI Appendix, Table S1). Two other authors [M.R.R. (trauma surgery) and D.J.W. (emergency medicine)] reviewed the cases for clarity and validity. Players compare cases, under time pressure, to identify relevant cues. We

iteratively refined the game based on feedback from emergency medicine physicians ($n = 36$), regarding its usability and enjoyability.

Shift: The Next Generation uses a four-step loop, exposing players to each decision principle. In step 1, players have 90 s to triage 10 patients. Each case has basic information (e.g., age, chief complaint). In each set, five cases reflect one principle and the remainder reflect other principles. Players can purchase information (e.g., physical examination findings, imaging results), using funds awarded at the start of a round, before making each triage decision (e.g., transfer, admit). Before the next case, they receive feedback on their decision. After each round, they receive a performance summary, highlighting cases reflecting the focal decision principle. In step 2, they must identify cues common to two of the five cases. If they identify the correct cues, they receive positive feedback. If not, an in-game character provides the relevant principle, framed as a useful mental shortcut. In step 3, players complete a "match 3" puzzle, intended as a "fun" break. It requires players to connect medical icons on a rectangular puzzle board, rewarding them with additional time on step 4, the final medical task. Step 4 is the same as step 1, with 10 new cases, offering players a chance to use what they have learned. It concludes with a summary of the principle and its rationale.

The design decisions for this game include the following:

- i) The decision principles reflect the American College of Surgeons' triage guidelines (adjusted for recent reviews of the effects of patient age and frailty) (17, 18, 22).
- ii) Players initially encounter representative cases, followed by less representative ones, with graded complexity meant to create a sense of flow and increasing challenge (21).
- iii) The match 3 game, adapted from casual games like *Candy Crush* and *Bejeweled*, provided a break before the final test.

Text-based educational interventions. We used two commercially available applications (*myATLS* and *Trauma Life Support MCQ Review*) to deliver a text-based educational program. Both are adjuncts to the ATLS course, the gold standard of continuing medical education in trauma. They reflect the two components of the ATLS course: a review of course content (*myATLS*) and a test of mastery, using multiple-choice questions (*Trauma Life Support MCQ*) (37).

Data Sources.

Outcome assessment. The primary outcome was performance on a virtual simulation, developed with a gaming company (Breakaway Ltd), which has been found to have both internal reliability and construct validity. At the group level, physicians make similar decisions for trauma patients on it as in clinical practice (19).

The simulation has 10 cases: four severely injured patients (two representative and two nonrepresentative), two minimally injured patients, and four critically ill nontrauma cases (Table 1). Users must evaluate and manage these cases over 42 min, simulating a busy 8-h ED shift. New patients arrive at prespecified (but unpredictable) intervals, so that physicians must manage multiple patients concurrently. Each case includes the patient, a chief complaint, vital signs that update every 30 s, a history, and a written description of the physical examination. Without appropriate clinical intervention by the user, severely injured trauma patients and critically ill patients decompensate and die.

Physicians manage patients by selecting orders from a list of 250 medications, studies, and procedures. Some orders affect patients' clinical status, with corresponding changes in vital signs and physical examinations. Other orders generate information, added to patients' charts. Each case ends when the player makes a disposition decision (admit, discharge, or transfer) or the patient dies.

Questionnaires. Physicians answered questions about their demographics, educational background, practice environment, and video game use. Those assigned to an intervention additionally reported how long they spent using it and how usable and enjoyable it was, on five-point Likert scales. Finally, they completed the narrative engagement scale, with four subscales: comprehension (narrative engagement), concentration (attentional focus), immersion (narrative presence), and emotion (emotional engagement) (21).

Analyses. We calculated the response rate as the proportion of participants who used the virtual simulation and the completion rate as the proportion who completed it.

Adherence, usability, enjoyment, and engagement. We summarized adherence, usability, enjoyment, and engagement across interventions using medians (interquartile ranges) and compared process measures across interventions using Kruskal-Wallis tests.

Physician performance. We scored participants' disposition decision (transfer, admit, or discharge) for each trauma case on the virtual simulation. We treated patients who died before a disposition decision as "transferred" to

avoid penalizing physicians for decisions made during the initial resuscitation. We calculated undertriage as the proportion of severely injured patients not transferred to a trauma center, as per American College of Surgeons guidelines (22).

We assessed the effects of the interventions on undertriage rates with linear regression. To account for missing responses, we performed multiple imputations by intervention assignment. We used the multivariable imputation by chained equations procedure (38), creating 20 imputed datasets and combining regression results following Rubin's rules (39).

We performed three sensitivity analyses. One used only cases for which we had full data. The second excluded participants who had ever worked at level I/II trauma centers. The third adopted a more stringent evaluation criterion, treating cases where patients died during the virtual simulation as not transferred (hence undertriaged).

In secondary analyses, we stratified cases as nonrepresentative and representative, and tested the association between the interventions and undertriage by type of case. We tested the association between measures of engagement, the interventions and undertriage. In response to the recommendation of an independent reviewer, we repeated the sensitivity and secondary analyses using the more stringent evaluation criterion, and present those results in *SI Appendix, Table S5*.

We performed multiple imputation using R 3.3.2 (The R Project for Statistical Computing) and all other analyses using Stata 13.0 (StataCorp).

Power Calculation. Using Cohen's method (40) and assuming an 80% completion rate, we estimated that recruiting 80 physicians per group ($n = 320$) would allow detecting a 15% greater reduction in undertriage with the game-based interventions than in the control arms, with $\alpha = 0.05$ and 80% power.

Human Subjects. The University of Pittsburgh hosted the simulation and questionnaires on a secure server, which participants could access online.

Upon enrollment, participants received a unique identifier for logging into the website, which they also used to access interventions. Only the principal investigator (D.M.) had access to the linkage file connecting these identifiers to physicians' name and contact information. This file was encrypted and stored on a secure server at the University of Pittsburgh.

The University of Pittsburgh Institutional Review Board approved the study (PRO17090094). All participants provided informed consent. We did not plan any interim analyses; hence, we had no stopping guidelines. However, we asked participants to communicate any adverse events or unintended effects of participation via email, which would have been relayed to the review board had any been reported. We registered the trial on <https://clinicaltrials.gov> (NCT03279575).

Data and Materials Availability. Access to the deidentified dataset will be made available upon written request to D.M. The games are available for download on the iPad App Store. The outcome assessment tool is available for use at <https://decisionmaking.ccm.pitt.edu>.

ACKNOWLEDGMENTS. We thank Michal Ksiazkiewicz, Harley Baldwin, and Jesse Schell at Schell Games and Tobi Saulnier, Matt Nolin, and other members of the team at 1st Playable Productions for developing *Shift: The Next Generation*. We thank the Biostatistics and Data Management Core, Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, Department of Critical Care Medicine, University of Pittsburgh for statistical support. Finally, we thank the many physicians at the University of Pittsburgh who participated in testing the video games. This work was supported by the NIH through Grant DP2 LM012339 (to D.M.), and Grant K08 HL122478 (to D.J.W.), by a grant from the UPMC Center for Clinical Decision Making (to D.M.), and by a grant from the Swedish Foundation for Humanities and Social Sciences (to B.F.). The funding agencies reviewed the study but played no role in its design, collection, analysis, or interpretation.

- National Academies of Sciences, Engineering, and Medicine (2016) *Improving Diagnosis in Health Care* (National Academies Press, Washington, DC).
- Singh H, Graber ML (2015) Improving diagnosis in health care—The next imperative for patient safety. *N Engl J Med* 373:2493–2495.
- Graber ML, Wachter RM, Cassel CK (2012) Bringing diagnosis into the quality and safety equations. *JAMA* 308:1211–1212.
- MacKenzie EJ, et al. (2006) A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med* 354:366–378.
- Mackenzie EJ, et al. (2007) The national study on costs and outcomes of trauma. *J Trauma* 63(Suppl):S54–S67, discussion S81–S86.
- Nathens AB, Jurkovich GJ, MacKenzie EJ, Rivara FP (2004) A resource-based assessment of trauma care in the United States. *J Trauma* 56:173–178, discussion 178.
- Mohan D, et al. (2013) Trauma triage in the emergency departments of nontrauma centers: An analysis of individual physician caseload on triage patterns. *J Trauma Acute Care Surg* 74:1541–1547.
- Delgado MK, et al. (2014) Factors associated with the disposition of severely injured patients initially seen at non-trauma center emergency departments: Disparities by insurance status. *JAMA Surg* 149:422–430.
- Lale A, Krajewski A, Friedman LS (2017) Undertriage of firearm-related injuries in a major metropolitan area. *JAMA Surg* 152:467–474.
- Gagliardi AR, Nathens AB (2015) Exploring the characteristics of high-performing hospitals that influence trauma triage and transfer. *J Trauma Acute Care Surg* 78:300–305.
- Mohan D, et al. (2015) Triage patterns for medicare patients presenting to nontrauma hospitals with moderate or severe injuries. *Ann Surg* 261:383–389.
- Kahneman D (2011) *Thinking, Fast and Slow* (Farrar, Straus, and Giroux, New York).
- Miller-Day M, Hecht ML (2013) Narrative means to preventative ends: A narrative engagement framework for designing prevention interventions. *Health Commun* 28:657–670.
- Mohan D, et al. (2017) Efficacy of educational video game versus traditional educational apps at improving physician decision making in trauma triage: Randomized controlled trial. *BMJ* 359:j5416.
- Idson LC, et al. (2004) Overcoming focusing failures in competitive environments. *J Behav Decis Making* 17:159–172.
- Mohan D, et al. (2016) Testing a videogame intervention to recalibrate physician heuristics in trauma triage: Study protocol for a randomized controlled trial. *BMC Emerg Med* 16:44.
- Goodmanson NW, et al. (2012) Defining geriatric trauma: When does age make a difference? *Surgery* 152:668–674, discussion 674–675.
- Joseph B, et al. (2014) Superiority of frailty over age in predicting outcomes among geriatric trauma patients: A prospective analysis. *JAMA Surg* 149:766–772.
- Mohan D, et al. (2014) Assessing the validity of using serious game technology to analyze physician decision making. *PLoS One* 9:e105445.
- Podsakoff PM, MacKenzie SB, Podsakoff NP (2012) Sources of method bias in social science research and recommendations on how to control it. *Annu Rev Psychol* 63:539–569.
- Busselle R, Bilandzic H (2009) Measuring narrative engagement. *Media Psychol* 12:321–347.
- Committee on Trauma, American College of Surgeons (2006) *Resources for Optimal Care of the Injured Patient 2006* (American College of Surgeons, Chicago), 1st Ed.
- Mohan D, et al. (2011) Assessing the feasibility of the American College of Surgeons' benchmarks for the triage of trauma patients. *Arch Surg* 146:786–792.
- National Academies of Sciences, Engineering, and Medicine (2016) *A National Trauma Care System: Integrating Military and Civilian Trauma Systems to Achieve Zero Preventable Deaths After Injury* (National Academies Press, Washington, DC).
- Rivara FP, et al. (2008) Outcomes of trauma patients after transfer to a level I trauma center. *J Trauma* 64:1594–1599.
- Counselman FL, et al. (2009) A study of the workforce in emergency medicine: 2007. *Am J Emerg Med* 27:691–700.
- Birnbaum MH (1983) Base rates in Bayesian inference: A signal detection analysis of the cab problem. *Am J Psych* 96:85–94.
- Graber ML, et al. (2012) Cognitive interventions to reduce diagnostic error: A narrative review. *BMJ Qual Saf* 21:535–557.
- Helfrich CD, et al. (2018) How the dual process model of human cognition can inform efforts to de-implement ineffective and harmful clinical practices: A preliminary model of unlearning and substitution. *J Eval Clin Pract* 24:198–205.
- Avorn J (2018) The psychology of clinical decision making—Implications for medication use. *N Engl J Med* 378:689–691.
- Morewedge CK, et al. (2015) Debiasing decisions. *Policy Insights Behav Brain Sci* 2:129–140.
- Ricciardi F, De Paolis LT (2014) A comprehensive review of serious games in health professions. *Int J Comput Games Technol* 2014:1–11.
- Graafland M, Schraagen JM, Schijven MP (2012) Systematic review of serious games for medical education and surgical skills training. *Br J Surg* 99:1322–1330.
- Madigan J (2016) *Getting Gamers: The Psychology of Video Games and Their Impact on the People Who Play Them* (Rowan & Littlefield, New York).
- Dickert N, Grady C (1999) What's the price of a research subject? Approaches to payment for research participation. *N Engl J Med* 341:198–203.
- Mohan D, et al. (2017) Using incentives to recruit physicians into behavioral trials: Lessons learned from four studies. *BMC Res Notes* 10:776.
- American College of Surgeons (2012) About advanced trauma life support. Available at <https://www.facs.org/quality-programs/trauma/atls/about>. Accessed August 16, 2016.
- van Buuren S, Boshuizen HC, Knook DL (1999) Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 18:681–694.
- Rubin DB (2009) *Multiple Imputation for Nonresponse in Surveys* (Wiley, New York).
- Cohen J (1992) A power primer. *Psychol Bull* 112:155–159.