Efficacy of educational video game versus traditional educational apps at improving physician decision making in trauma triage: randomized controlled trial

Deepika Mohan,[1] Coreen Farris,[2] Baruch Fischhoff,[3] Matthew R Rosengart,[4] Derek C Angus,[1] Donald M Yealy,[5] David J Wallace,[1] Amber E Barnato[6]

[1]Scaife Hall, 3550 Terrace St, University of Pittsburgh, Pittsburgh, PA 15261, USA

[2]4570 Fifth Avenue, Suite 600, RAND Corporation, Pittsburgh, PA 15213, USA

[3]Porter Hall 219E, 5000 Forbes Avenue, Carnegie Mellon University, Pittsburgh, PA 15213, USA

[4]F1266 Presbyterian Hospital, University of Pittsburgh, Pittsburgh, PA, 15213, USA

[5]3600 Meyran Avenue, University of Pittsburgh, Pittsburgh, PA 15260, USA

[6]The Dartmouth Institute, Williamson Translational Building, 5th Floor, One Medical Center Drive, Lebanon, NH 03756, USA

Correspondence to: D Mohan mohand@upmc.edu

Additional material is published online only. To view please visit the journal online.

ABSTRACT

OBJECTIVE
To determine whether a behavioral intervention delivered through a video game can improve the appropriateness of trauma triage decisions in the emergency department of non-trauma centers.

DESIGN
Randomized clinical trial.

SETTING
Online intervention in national sample of emergency medicine physicians who make triage decisions at US hospitals.

PARTICIPANTS
368 emergency medicine physicians primarily working at non-trauma centers. A random sample (n=200) of those with primary outcome data was reassessed at six months.

INTERVENTIONS
Physicians were randomized in a 1:1 ratio to one hour of exposure to an adventure video game (Night Shift) or apps based on traditional didactic education (myATLS and Trauma Life Support MCQ Review), both on iPads. Night Shift was developed to recalibrate the process of using pattern recognition to recognize moderate-severe injuries (representativeness heuristics) through the use of stories to promote behavior change (narrative engagement). Physicians were randomized with a 2×2 factorial design to intervention (game *v* traditional education apps) and then to the experimental condition under which they completed the outcome assessment tool (low *v* high cognitive load). Blinding could not be maintained after allocation but group assignment was masked during the analysis phase.

MAIN OUTCOME MEASURES
Outcomes of a virtual simulation that included 10 cases; in four of these the patients had severe injuries. Participants completed the simulation within four weeks of their intervention. Decisions to admit, discharge, or transfer were measured. The proportion of patients under-triaged (patients with severe injuries not transferred to a trauma center) was calculated then (primary outcome) and again six months later, with a different set of cases (primary outcome of follow-up study). The secondary outcome was effect of cognitive load on under-triage.

RESULTS
149 (81%) physicians in the game arm and 148 (80%) in the traditional education arm completed the trial. Of these, 64/100 (64%) and 58/100 (58%), respectively, completed reassessment at six months. The mean age was 40 (SD 8.9), 283 (96%) were trained in emergency medicine, and 207 (70%) were ATLS (advanced trauma life support) certified. Physicians exposed to the game under-triaged fewer severely injured patients than those exposed to didactic education (316/596 (0.53) *v* 377/592 (0.64), estimated difference 0.11, 95% confidence interval 0.05 to 0.16; P<0.001). Cognitive load did not influence under-triage (161/308 (0.53) *v* 155/288 (0.54) in the game arm; 197/300 (0.66) *v* 180/292 (0.62) in the traditional educational apps arm; P=0.66). At six months, physicians exposed to the game remained less likely to under-triage patients (146/256 (0.57) *v* 172/232 (0.74), estimated difference 0.17, 0.09 to 0.25; P<0.001). No physician reported side effects. The sample might not reflect all emergency medicine physicians, and a small set of cases was used to assess performance.

CONCLUSIONS
Compared with apps based on traditional didactic education, exposure of physicians to a theoretically grounded video game improved triage decision making in a validated virtual simulation. Though the observed effect was large, the wide confidence intervals include the possibility of a small benefit, and the real world efficacy of this intervention remains uncertain.

TRIAL REGISTRATION
clinicaltrials.gov; NCT02857348 (initial study)/ NCT03138304 (follow-up).

WHAT IS ALREADY KNOWN ON THIS TOPIC

Strategies designed to change physician decision making have had limited success

No interventions exist to improve physician heuristics—the intuitive judgments that drive much of medical decision making

WHAT THIS STUDY ADDS

In this randomized clinical trial, physicians exposed to a video game intervention were more likely to follow clinical practice guidelines in the triage of simulated trauma patients than physicians exposed to a traditional educational program

A theoretically grounded video game intervention has the potential to modify physician behavior, although the magnitude of the effect and real world effectiveness remain uncertain

Key limitations include our use of a convenience sample of physicians and the use of a virtual simulation as the outcome assessment tool

Introduction

Medical diagnosis often requires physicians to collect and integrate complex uncertain information from multiple sources.[1] Under normal conditions, that process requires reliance on heuristic cognitive processes.[2][3] Heuristics generate solutions to complex

problems through pattern recognition and simplifying assumptions. When calibrated well, heuristics allow people to function under conditions of time pressure and uncertainty.[4] When calibrated poorly, however, they result in predictable errors in judgment.[5] As a result, many interventions attempt to reduce reliance on heuristic cognitive processes.[6] Some of these interventions focus on increasing physicians' use of clinical practice guidelines through direct instruction (such as warning about the risks of relying on heuristics, checklists of necessary actions) or outcome feedback (such as telling physicians how they have done).[7-11] Others try to shift the locus of decision making from the bedside clinician to a third party such as a decision tool or an external consultant.[12 13] To date, these interventions have had mixed success, with limited transference across task domains.[6] More importantly, they do not directly deal with the long term need to improve the heuristic processes that underpin most physician decision making.[1 14-16]

Behavioral scientists agree that people develop well calibrated heuristics when the decisional context provides reliable valid cues to the problem and they have the opportunity to learn the relevant contextual cues.[17] Typically, calibration—the process of refining the accuracy of judgment—requires an experience-feedback loop. Researchers in other specialties (such as aviation, organizational science, threat detection), however, have experimented with different behavioral interventions that serve as surrogates for this process.[18-20] We created such an intervention for a medical diagnosis task that has proved difficult to improve: trauma triage decisions.

Trauma triage involves the identification and transfer of severely injured patients to trauma centers, either directly from the field or after evaluation at a non-trauma center. High levels of under-triage (the failure to transfer severely injured patients to trauma centers) persist, despite efforts to improve best practice. The problem is particularly acute at non-trauma centers, where fewer than 30% of severely injured patients are transferred as recommended by clinical practice guidelines.[21-25] Our prior experimental and observational work suggests that heuristics play an important role in under-triage.[26 27] We selected one promising method of recalibrating heuristics—narrative engagement—and developed a theoretically grounded intervention delivered through the platform of a video game. Narrative engagement is defined as the use of compelling stories to communicate and encode principles of best practice decision.

We compared the efficacy of a narrative engagement video game with that of a prominent education intervention for improving simulated decisions in trauma triage with emergency medicine physicians practicing at non-trauma centers as participants.

Methods
Overview
We have previously published the study protocol for this trial.[28] We developed a video game (Night Shift) in collaboration with Schell Games (Pittsburgh, PA) and conducted a randomized controlled trial of the effect of the game compared with traditional didactic education, administered through commercially available applications, on triage by US emergency physicians practicing at non-trauma centers. We hypothesized that physicians exposed to game based education would under-triage fewer patients on a virtual simulation than those exposed to the didactic education program (primary trial outcome) and that experimentally induced cognitive load would degrade triage performance less among physicians exposed to the game than among those exposed to the didactic program (secondary trial outcome). Process measures included adherence to the interventions, as well as their usability and likeability.

Subsequently, we assessed the duration of the treatment effect among a random sample of those who completed the outcome assessment tool by measuring physician performance again six months after the completion of the initial trial protocol (primary six month reassessment outcome). We hypothesized that physicians exposed to game based education during the initial trial would continue to under-triage fewer patients on the simulation than physicians exposed to the didactic program.

Conceptual model
In our conceptual model, grounded in behavioral decision research, physicians first judge the severity of the injury and then decide how to manage it.[3 29] Those judgments reflect the interaction between what are called system one processes, which are fast, automatic, and heuristic, and system two processes, which are slow, deliberate, and analytic.[2] Under time pressure, people increasingly default to heuristics (or mental shortcuts) that can produce accurate answers but also predictable errors.[30] As time pressure decreases, people are better able to synthesize the complex uncertain elements of difficult decisions—assuming that they have the training to do so.[2] The decisions themselves also reflect the influence of other variables.[31] In triage, these could include physicians' attitudes towards guidelines, institutional norms, resource constraints, and patient preferences.

Our intervention sought to modify system one processes so as to improve heuristic thinking in decision making in trauma triage. Our specific strategy was based on clinical experience and experimental observations.[27] For example, we observed that patients with gunshot wounds were far more likely to be transferred to a trauma center than patients who had fallen, even when they had similar injury severity scores. That pattern is consistent with judgment by representativeness: physicians have an archetype (a pattern) of how severely injured patients present and then transfer patients who match ("representative" cases) but admit or discharge those who do not ("non-representative" cases). Crucially, representativeness does not depend on how often a case occurs but reflects an associative judgment, informed by experience and

training.[32] As a result, physicians can systematically make correct decisions for cases they deem "obvious" (representative) but make errors with those that are less obvious (non-representative), regardless of frequency.[3] Furthermore, physicians with different backgrounds (such as trauma surgeons versus emergency medicine physicians) might make different judgments by representativeness. Heuristics are well calibrated if they align with the reference standard for the specific decision context. We developed Night Shift to recalibrate the heuristics of emergency medicine physicians in trauma triage.

Participants

The triage of trauma patients by physicians occurs at non-trauma centers and level III/IV centers. The designated trauma level of a hospital reflects its ability to manage injuries definitively, based on an accreditation process conducted by the American College of Surgeons and state authorities. The scale ranges from I (fully resourced hospital; serves as regional referral center) to IV (minimally resourced hospital; capable of stabilizing patients but must refer severe injuries to a higher level of care). Hospitals that have not applied for accreditation are referred to as non-trauma centers. Our goal was to recruit a national sample of emergency medicine physicians who make triage decisions. To that end, we recruited board ertified and board eligible physicians working primarily outside level I/II trauma centers in the US at the 2016 annual scientific meeting of the American College of Emergency Physicians (October 16-18).

We randomized physicians using a 2×2 factorial design, with a 1:1:1:1 allocation to complete either game based or didactic educational applications and to complete the outcome assessment tool under conditions of low or high cognitive load. We anticipated that variation in the cognitive load would amplify the effect of heuristics on performance, thereby allowing us to isolate the mechanism by which the game influenced performance.[30] Our randomization scheme was generated in Stata 13.0 with block sizes of four and eight. After participants registered, study personnel obtained their assignment to the intervention and outcome assessment condition from a central database. Although we could not maintain blindness after allocation, we masked group assignment during the analysis phase.

Study protocol

At enrollment, physicians received an iPad mini 2 loaded with their intervention. We asked them to spend at least an hour with the intervention and then log onto a secure website that hosted a questionnaire to assess demographics and personal characteristics; a questionnaire to assess use of the interventions (adherence, usability, likeability); and a virtual simulation that served as the outcome assessment tool. Responding to the questionnaires and simulation took about 60 minutes. Participants completed the protocol at their convenience and could keep the iPad (worth

about $260 (£195, €218)). They received weekly email reminders until the study closed on 14 November 2016 or until they completed the protocol.

Six months after the completion of the initial trial (May 2017), we emailed a random sample of 100 physicians from each group who had completed all primary study procedures to ask if they would participate in a second assessment. We asked respondents to complete the outcome assessment tool a second time (with a different set of cases) and offered a $100 Amazon gift card on completion of the task. Physicians who agreed to participate received weekly email reminders until the study closed on 15 June 2017.

Questionnaire to assess demographics

Each physician completed a questionnaire with items on age, sex, race, ethnicity, educational background (board certification, ATLS (advanced trauma life support) certification, years since completion of residency), and practice environment (trauma designation of their hospital, affiliation with a level I/II trauma center, affiliation with an emergency medicine residency program). We used the Big Five Inventory-10 for an exploratory analysis of personality traits that might influence the efficacy of the different interventions.[33]

Questionnaire to assess physicians' use of interventions

To measure adherence, we asked physicians to report how long they spent using their interventions. To measure the usability and likeability of the interventions, we asked physicians to provide qualitative feedback about the experience.

Game based education: Night Shift

Based on the input of an expert panel of seven trauma surgeons, we distilled the clinical practice guidelines for the triage of trauma patients into three simple principles. The following types of patients have severe injuries until proved otherwise: elderly (>70) and frail patients; patients with injuries affecting more than one body region; and patients with rib fractures or open long bone fractures. We built Night Shift, a two dimensional adventure video game that relies on narrative engagement (that is, the use of compelling stories to promote behavior change) to disseminate these principles. Three research threads support the potential for narrative engagement to alter judgment. One research thread finds that stories facilitate the processing and retention of new data.[34 35] The second body of research finds that practicing desired behaviors in a safe environment helps people to gain warranted feelings of self efficacy, providing the confidence needed to deploy newly acquired skills.[36] The third body of research finds that stories can engage players cognitively and emotionally in ways that transcend traditional forms of education.[37]

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 emergency department. They are given the dual objective of solving the mystery of the grandfather's absence and of interacting with the patients who present to the department. These patients have various traumatic and non-traumatic complaints, ranging from the obscure (such as foul smelling body odor and fever after exposure to camel's milk) to the common (such as low speed motor vehicle collision with minimal injuries). The game centers on a series of trauma patients who arrive with "non-representative" severe injuries—cases in which the injury complex does not fit the popular archetype for the problem. As players interact with these patients, they gain experience with the consequences of under-triage. Specifically, these patients return with complications from their injury. Not only do players have to find a solution for the patient's deteriorating clinical condition, they also have to explain the outcome to characters in the game (such as family members, consultants). In their responses, the characters provide didactic information about relevant contextual cues for the evaluation of the trauma patients. At the same time, they highlight the repercussions of under-triage (such as preventable disability from a delay in treatment) to evoke an emotional response that would make the feedback memorable (fig 1).

We made three additional design decisions to enhance players' emotional and cognitive engagement with the game. First, we embedded the medical component of the game within an overarching mystery. During the quest to find Andy's grandfather, the player uncovers Andy's background and motivation. The process allows the player to gain empathy for Andy, which in turn makes the feedback provided by characters in the game feel personally relevant. Second, we included patients with "representative" severe injuries—injuries that do fit the popular archetype of the problem—in the mix of cases that arrive at the emergency department. These patients decompensate shortly after arrival. To salvage them, players have to participate in various team based resuscitation scenarios. In play tests, we found that even a small amount of structured role playing embedded in the adventure increased immersion in the story. Third, we incorporated a puzzle solving mechanic into the medical portion of the game to increase its cognitive challenge. We based the non-trauma patients on abstracted versions of clinical case challenges published in the *New England Journal of Medicine*.[38-42] Not all the relevant information is provided, forcing players to draw connections between associated pieces of data. Only if they make the right connections, do patients offer the information required to make the correct diagnosis, allowing the player to initiate the correct treatment.

Didactic education apps: myATLS and Trauma Life Support MCQ Review

The ideal standard educational strategy in trauma is Advanced Trauma Life Support (ATLS)—a two day seminar designed to teach participants to resuscitate and stabilize trauma patients and to determine if

Fig 1 | Selected screenshots from interaction between Andy and his boss, the department chair. In this instance, Andy has failed to transfer a patient (Benjamin) with a cervical spine fracture to a trauma center, and Benjamin has returned with a central cord syndrome. During the conversation, players can choose how they want to respond to the department chair's criticism of their performance, either accepting responsibility or arguing that the complication represents the natural evolution of the disease process rather than a diagnostic error

the patient's needs exceeds the capabilities of their facility.[43] Participants must complete a multiple choice test before and after the course to receive certification. As a surrogate, we provided physicians with two educational software apps: myATLS and Trauma Life Support MCQ Review. The former contains a summary of all the content provided in the ATLS course.[44] The latter is designed to help users prepare for the ATLS exam and contains 550 multiple choice questions. We asked participants to use their discretion in deciding how to allocate the hour they spent reviewing the two apps.

Outcome assessment

We developed a virtual simulation to provide a high fidelity replication of the emergency department environment so that we could assess trauma triage decision making in a controlled environment.[45] When designing the simulation, we designated certain cases as "representative" (fitting the popular archetype for severe or minor injuries) and others as "non-representative" (not fitting the popular archetype for severe or minor injuries) based on clinical experience and experimental observations. We previously established the simulations' internal reliability and construct validity. In prior research, we found that physicians, as a group, make similar diagnostic (such as acquisition of computed tomogram) and triage decisions (such as transfer) for trauma patients on the simulation as in clinical practice.[45]

The simulation presents 10 cases over 42 minutes, representing a busy eight hour shift. It includes four patients with severe injuries, two with minimal injuries, and four with non-traumatic complaints (see appendix for details). Each case includes a 2D rendering of the patient, a chief complaint, vital signs that update every 30 seconds, a history, and a written description of the physical exam (fig 2). Users evaluate and manage patients by selecting from a prespecified list of 250 medications, studies, and procedures. Some orders affect a patient's clinical status, leading to corresponding changes in their vital signs and findings on physical exam. For example, a blood transfusion given to a patient in hemorrhagic shock will stabilize his/her blood pressure. Other orders generate additional information, presented as reports added to the patients' charts. The cases end when physicians either make a disposition decision (admit, discharge, transfer) or the patient dies.

New patients arrive at prespecified (but unpredictable) intervals so that physicians manage multiple patients concurrently. In addition to their clinical responsibilities, participants also have to respond to various audiovisual distractors, including nursing requests for help with disruptive patients, interruptions by families asking for information, and paging alerts.

During the initial study, we randomized physicians to complete the simulation under conditions of low or high cognitive load to test the mechanism of the treatment effect. We manipulated cognitive load in two ways. First, we varied the complexity of the non-trauma cases. In the low load arm, non-trauma patients
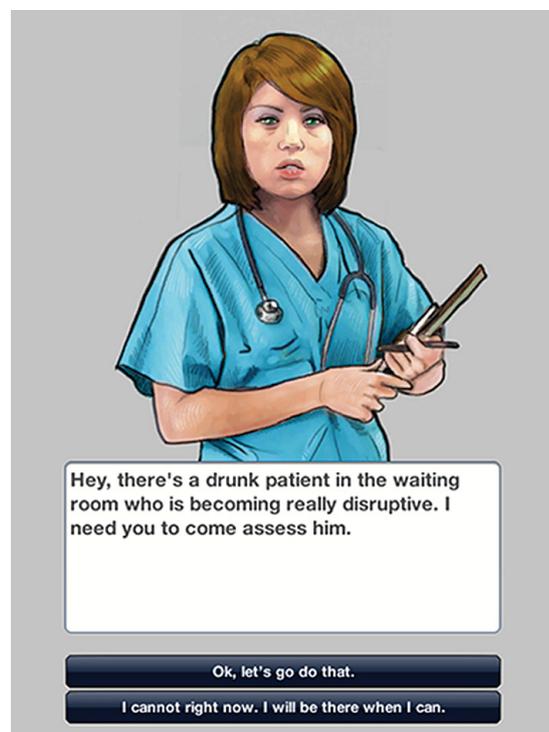

Fig 2 | Screenshot from the virtual simulation (outcome assessment tool). Audiovisual distractors were included, such as nursing requests for help with disruptive patients, to increase the verisimilitude of the experience

had routine complaints (such as appendicitis), arrived hemodynamically stable, and did not deteriorate over the course of the simulation. In the high load arm, non-trauma patients were critically ill (such as septic), arrived hemodynamically unstable, and deteriorated without adequate management. Second, we reduced the number of rooms that physicians could use to evaluate patients from eight in the low load arm to four in the high load arm, which increased the number of distractors that they received.

Our primary outcome for the trial was physicians' performance on the simulation as measured by the proportion of under-triaged severely injured patients. The secondary outcome of the trial was the effect of cognitive load on these simulated triage decisions.

Assessment of duration of treatment effect

Physicians who participated in the follow-up study completed the virtual simulation a second time after six months, with a different set of trauma cases (see appendix). Given the limited cohort size, we standardized the cognitive load manipulation for all participants, exposing all participants to high load conditions. The primary outcome of this follow-up study was the proportion of severely injured patients under-triaged.

Analyses

We calculated the response rate as the proportion of enrolled physicians who logged into the website and

the completion rate as the proportion who finished the virtual simulation. We conducted our primary analysis using an intention to treat approach. Specifically, we included physicians who did and did not adhere to the requirement to spend an hour on their assigned intervention but restricted our analysis to physicians with outcome data (that is, those who completed the virtual simulation). We assumed that physicians who did not complete would not differ substantially from those who did but then performed a sensitivity analysis in which we explored the effect of departures from that assumption.

We summarized physician characteristics using means (and SD) for continuous variables and proportions (%) for categorical variables.

Adherence, usability, and likeability

We measured adherence as self reported minutes spent on the intervention, summarized using medians and interquartile ranges. We categorized qualitative feedback about the usability and likeability of the intervention as positive or negative. We compared adherence across interventions using a Kruskal-Wallis test and the usability and likeability across interventions using χ^2 tests.

Outcome assessment

We first evaluated participants' performance on each trauma case based on a review of their disposition decisions (transfer, admit, discharge). We categorized patients who died before a disposition decision as "transferred" as we could not predict what the physician would have done given a successful resuscitation and wanted to give him/her the benefit of the doubt. We then calculated each group's proportion of under-triage (defined as the number of patients not transferred to a trauma center divided by total number of severely injured patients who should have been transferred to a trauma center).[46] To be consistent with our statistical analysis plan, we treated the proportion of under-triage as continuous and compared the effects of the intervention (primary outcome), cognitive load (secondary outcome), and the interaction of these factors on under-triage using a two way analysis of variance. In response to the recommendation of an independent statistical reviewer, we completed a post hoc analysis in which we treated under-triage as binomial and used a Poisson regression model with robust standard errors to test these effects.

We included all participants who had outcome data (that is, they completed the outcome assessment tool (simulation) within four weeks of exposure to their assigned treatment). In sensitivity analyses, we tested three different imputation scenarios to explore the potential bias introduced in our effect estimates based on types of non-random missingness. In scenario 1, we imputed worse than cohort average scores for missing physician outcomes in the game arm, assuming they performed like physicians in the education arm. In scenario 2, we imputed better than cohort average scores for missing physician outcomes in the educational arm, assuming they performed like physicians in the game arm. In scenario 3, we imputed worse than cohort average scores for missing physician outcomes in the game arm and better than cohort average scores for missing physician outcomes in the education arm.

We also performed sensitivity analyses to determine the effect of excluding participants who worked at both level I/II trauma centers and non-trauma centers, because they might have different heuristics to physicians who only ever worked at non-trauma centers, and participants who experienced usability issues with the interventions. Finally, we excluded from the analysis cases in which the patient died in the emergency department to see if changing our definition of the outcome would affect our estimate of the effects of the interventions.

In exploratory analyses, defined post hoc, we further assessed the relation between triage decisions and patient representativeness, defined as injuries fitting or not fitting the archetype; adherence, measured as time spent on the intervention (in thirds); and likeability, measured as whether the participant reported liking the intervention. Again we used analysis of variance (ANOVA) and Poisson regressions to test the associations between predictors and outcome measures.

Duration of treatment effect

As during the main trial, we scored each participant's responses to the simulation, summarized triage decisions at the group level, and compared the effects of the intervention on under-triage using both ANOVA (primary six month reassessment outcome) and Poisson regression analyses (post hoc).

All statistical analyses were conducted with Stata 13.0 (Statacorp, TX).

Human subjects and power calculation

We registered the trial and the follow-up study on clinicaltrials.gov (NCT02857348; NCT03138304). We planned the six month reassessment after initiating the trial, based on our receipt of supplemental funding, and therefore registered the follow-up study as a second trial.

We used Cohen's method of estimating power for behavioral trials and assumed a 70% completion rate.[47] We planned to perform a two way analysis of variance and predicted the mean proportion of under-triage and standard deviation based on results from prior work. This calculation resulted in a plan to recruit 368 physicians, which would give us 80% power to detect an 8-12% (moderate-large) difference in performance between the two intervention groups at a significance level of 0.05.

For the six month outcome study, we used a similar strategy to plan the sample size (albeit assuming a 60% response rate) and estimated that recruiting 200 physicians would give us 80% power to detect an 8% (moderate) difference in performance between the two intervention groups at a significance level of 0.05.
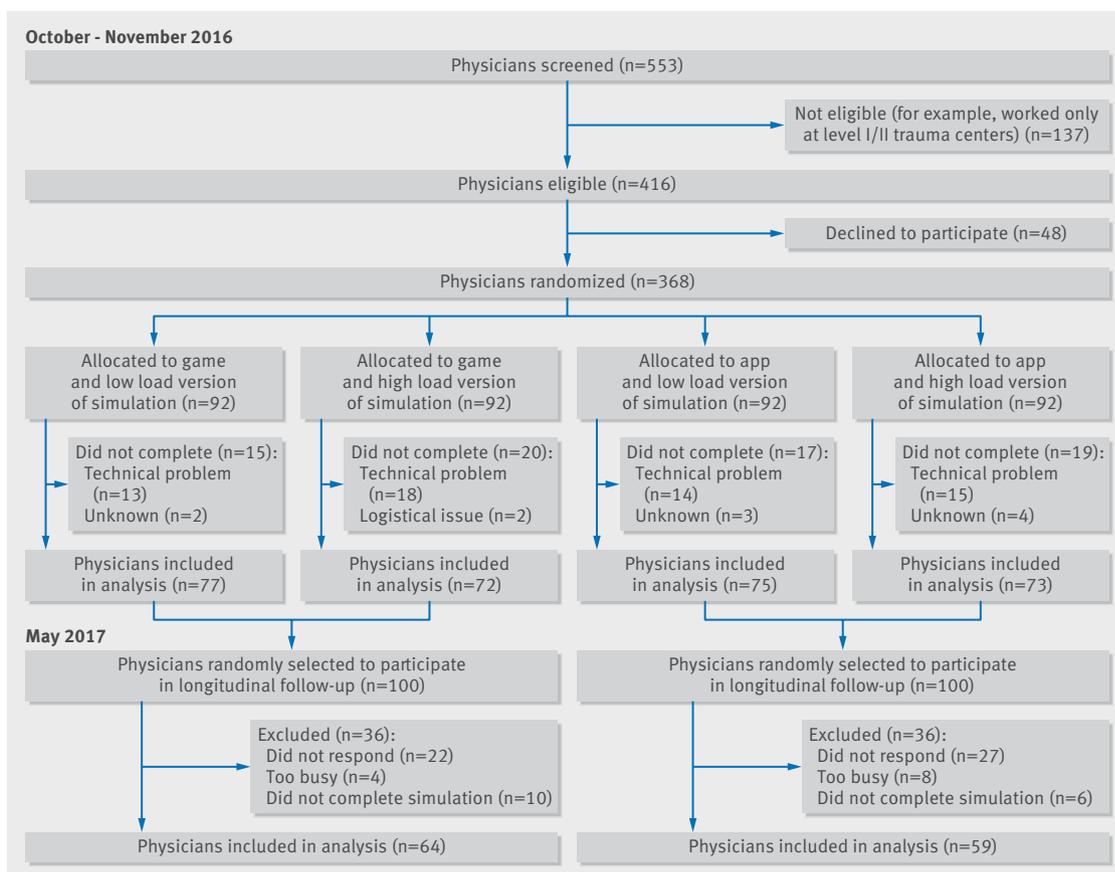
Fig 3 | Screening, randomization, and analysis. In total, 297 (81%) physicians completed the simulations during the initial trial and 122 (61%) completed the simulations during the follow-up study

Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for design or implementation of the study. No patients were asked to advise on interpretation or writing up of results. Results of the trial will be made available to all participants via clinicaltrials.gov as well as by email notification.

Results

Participant characteristics

We enrolled 368 physicians in the trial between 16-17 October 2016. Of these, 324 physicians logged into the website (88%), and 297 (81%) completed the outcome assessment portion of the study protocol by 14 November 2016, when the study closed (fig 3). The mean age of physicians completing the protocol was 39.9 (SD 8.9). Of those who took part, 283 (96%) had completed a residency in emergency medicine, 207 (70%) had received certification in advanced trauma life support (table 1), and 36 (12%) currently worked in a level I/II trauma center as well as at a non-trauma center.

In May 2017, we recruited a random sample of 100 physicians from each intervention arm from the 297 who completed the trial and enrolled 142 (71%) in the six month outcome assessment. Of these, 122 (61%) completed the outcome assessment tool for a

second time by the time that study closed on 15 June 2017 (fig 3).

Triage decision making

Table 2 shows the effect sizes with confidence intervals and significance levels, and table 3 shows the relative risks. Physicians randomized to receive game based education (n=149) under-triaged fewer severely injured patients than physicians exposed to the didactic educational program (n=148) (316/596 (0.53) v 377/592 (0.64); mean difference 0.11, 95% confidence interval 0.05 to 0.16; P<0.001). The main effect of cognitive load was not significantly related to under-triage and did not interact significantly with intervention assignment. The effect of the intervention on performance did not change when we excluded from the analysis physicians who worked at level I/II trauma centers or cases in which the patient died.

When we limited the analysis to the cases most likely to resemble the physicians' archetype of a severely injured patient (those we had designated as "representative"), exposure to the intervention did not affect triage. When we analyzed decisions for cases less likely to resemble the physicians' archetype (those we had designated as "non-representative"), however, we found that exposure to game based training did reduce the rate of under-triage compared with exposure to

Table 1 | Characteristics of participating physicians in study of effect of video game versus traditional educational apps on triage decisions in simulated trauma cases. Figures are numbers (percentage) unless stated otherwise

Characteristic	Game		Educational apps	
	Initial study (n=149)	Follow-up study (n=64)	Initial study (n=148)	Follow-up study (n=58)
Mean (SD) age (years)	40 (9.4)	38 (7.5)	40 (8.4)	38 (7.3)
Female	47 (32)	28 (44)	54 (37)	18 (31)
Mean (SD) years of experience	9.1 (9.0)	8.2 (7.4)	8.0 (8.1)	7.7 (7.4)
Race:				
White (non-Hispanic)	104 (70)	45 (70)	97 (66)	37 (63)
White (Hispanic)	11 (7)	4 (6)	12 (8)	4 (7)
Black	4 (3)	2 (3)	6 (4)	1 (2)
Asian	25 (17)	13 (20)	25 (17)	14 (24)
American Indian	2 (1)	0 (0)	3 (2)	1 (2)
Other	3 (2)	0 (0)	5 (3)	2 (4)
Primary board certification:				
Emergency medicine	141 (95)	61 (95)	142 (96)	58 (98)
Internal medicine/family practice	7 (5)	3 (5)	3 (2)	1 (2)
Other	1 (1)	0 (0)	2 (1)	0 (0)
ATLS certified	105 (71)	49 (77)	102 (69)	41 (71)
Practice at trauma center:				
Level I	15 (10)	6 (9)	5 (3)	2 (3)
Level II	9 (6)	3 (5)	7 (5)	3 (5)
Level III	19 (13)	7 (11)	23 (16)	10 (17)
Level IV	5 (3)	3 (5)	2 (1)	1 (2)
None	101 (68)	45 (70)	110 (75)	42 (72)
Means (SD) score for personality traits*:				
Extraversion	3.5 (1.0)	3.6 (0.97)	3.6 (1.0)	3.5 (1.01)
Agreeableness	3.6 (0.79)	3.5 (0.78)	3.6 (0.81)	3.6 (0.9)
Conscientiousness	4.3 (0.57)	4.3 (0.6)	4.3 (0.68)	4.3 (0.7)
Neuroticism	2.2 (0.83)	2.3 (0.89)	2.1 (0.82)	2.1 (0.9)
Openness	3.4 (0.89)	3.2 (0.83)	3.6 (0.91)	3.6 (0.96)

*Assessed with Big Five Inventory-10.[33]

didactic education (186/298 (0.63) *v* 239/296 (0.81); mean difference 0.18, 95% confidence interval 0.11 to 0.25; P<0.001).

Table 4 shows the sensitivity analysis in which we tested the possible influence of missing data on our results. Regardless of imputation assumption, performance among physicians exposed to the game based education remained significantly higher than among physicians exposed to didactic education.

Adherence, usability, and likeability of the interventions

Physicians reported spending similar amounts of time with the game (90 minutes; range 30-240; interquartile range 60-120) as on the educational apps (90 minutes; range 45-300; interquartile range 65-120; P=0.06).

Physicians in the game arm more often noted usability problems (30%) than physicians in the didactic education arm (8%, P<0.001). One specific problem, experienced by several participants during

Table 2 | Assessment of triage decision making by physicians randomized to video game versus traditional educational apps based on educational programs on simulated trauma cases with analyses of variance

	Proportion under-triaged (No)		Estimated difference (95% CI)	F statistic	P value
	Video game	Educational apps			
Main model	0.53 (149)	0.64 (148)	0.11 (0.05 to 0.16)	4.91	<0.001
Sensitivity analyses					
Excluding physicians who work at trauma centers	0.56 (125)	0.65 (135)	0.09 (0.03 to 0.15)	0.29	0.002
Excluding physicians who experienced usability issues	0.53 (105)	0.64 (136)	0.11 (0.05 to 0.17)	12.53	<0.001
Excluding cases in which patient died	0.64 (149)	0.76 (148)	0.12 (0.06 to 0.18)	17.23	<0.001
Post hoc analyses					
Types of cases:					
Representative cases	0.44 (149)	0.47 (148)	0.03 (−0.05 to 0.11)	0.58	0.45
Non-representative cases	0.63 (149)	0.81 (148)	0.18 (0.11 to 0.25)	24.81	<0.001
Adherence:					
Exposure <75 min	0.59 (51)	0.67 (33)	0.08 (−0.03 to 0.19)	2.25	0.13
Exposure 75-105 min	0.53 (85)	0.63 (91)	0.09 (0.02 to 0.17)	6.61	0.01
Exposure >105 minutes	0.36 (13)	0.65 (24)	0.29 (0.13 to 0.45)	13.32	<0.001
Likeability:					
Did not report enjoying intervention	0.56 (149)	0.60 (148)	0.04 (−0.11 to 0.18)	0.28	0.60
Six month follow-up:					
Duration of treatment effect	0.57 (64)	0.74 (59)	0.17 (0.09 to 0.25)	16.14	<0.001

Table 3 | Assessment of triage decision making by physicians randomized to video game versus traditional educational apps based on educational programs on simulated trauma cases with Poisson regression models

	Relative risk (95% CI)	P value
Main model (n=297)		
Exposure to video game (reference: educational program)	0.86 (0.75 to 0.99)	0.04
Completion of outcome assessment under conditions of high cognitive load (reference: low load)	1.06 (0.94 to 1.20)	0.34
Sensitivity analyses		
Excluding physicians who work at level I/II trauma center (n=260):		
Exposure to video game (reference: educational program)	0.86 (0.78 to 0.95)	0.003
Completion of outcome assessment under conditions of high cognitive load (reference: low load)	1.02 (0.92 to 1.12)	0.73
Excluding physicians who experienced usability issues with interventions (n=241):		
Exposure to video game (reference: educational program)	0.82 (0.74 to 0.92)	0.001
Completion of outcome assessment under conditions of high cognitive load (reference: low load)	0.99 (0.90 to 1.10)	0.92
Excluding cases in which patients died (n=297):		
Exposure to video game (reference: educational program)	0.84 (0.78 to 0.91)	<0.001
Completion of outcome assessment under conditions of high cognitive load (reference: low load)	1.02 (0.94 to 1.10)	0.66
Post hoc analyses		
Cases with representative severe injuries (n=297):		
Exposure to video game (reference: educational program)	0.93 (0.78 to 1.11)	0.45
Cases with non-representative severe injuries (n=297)		
Exposure to video game (reference: educational program)	0.78 (0.70 to 0.86)	<0.001
Adherence (n=297):		
Exposure to video game (reference: educational program)	0.82 (0.75 to 0.90)	<0.001
Time spent on intervention: 75-105 min (reference <75 min)	0.92 (0.83 to 1.01)	0.11
Time spent on intervention: >105 minutes (reference <75 min)	0.84 (0.71 to 0.99)	0.04
Likeability (n=297)		
Exposure to video game (referent: educational program)	0.81 (0.72 to 0.91)	<0.001
Did not describe liking intervention (reference: liked intervention)	1.04 (0.92 to 1.19)	0.45
6 month follow-up study		
Duration of treatment effect (n=122)		
Exposure to video game (reference: educational program)	0.77 (0.67 to 0.88)	<0.001

the first week of the trial, was a programming error in Night Shift that prevented play after about 75-90 minutes. The gaming company provided an update, via the Apple Store, for the second week of the trial. Many physicians, however, did not download the update and therefore encountered the error.

Physicians who used the game were also less likely to describe their intervention as enjoyable (40%) than physicians who used the educational apps (91%, P<0.001). Physicians who provided positive feedback about the game described the adventure as engaging; those who provided negative feedback described it distracting or annoying (table 5). In contrast, physicians who provided positive feedback about the didactic educational intervention described the apps as useful

and accessible; those who provided negative feedback described them as "superficial" or "remedial."

Physicians who spent more time on their assigned intervention had better performance on the triage simulation (tables 2 and 3). Exclusion of physicians who reported usability issues did not alter the effect of the intervention on performance. Reported enjoyment of the intervention was unrelated to performance on the simulation.

Duration of treatment effect
As shown in tables 2 and 3, six months after completing their intervention, physicians who used the video game under-triaged fewer patients on the virtual simulation compared with physicians who used the didactic

Table 4 | Sensitivity analysis to test effect of missing outcome data in study of effect of video game versus traditional educational apps based on educational programs with analyses of variance

	Proportion of under-triage		Mean difference (95% CI)	P value
	Video game	Educational apps		
Main analysis (n=297)	0.53	0.64	0.11 (0.05 to 0.16)	<0.001
Assumptions for scenario 1 (n=368)				
Missing physicians in game arm would have performed similarly to those observed in educational arm and missing physicians in educational arm would have performed similarly to those observed in their cohort	0.55	0.64	0.09 (0.03 to 0.14)	0.003
Assumptions for scenario 2 (n=368)				
Missing physicians in game arm would have performed similarly to those observed in their cohort and missing physicians in educational arm would have performed similarly to those observed in game arm	0.53	0.62	0.09 (0.03 to 0.14)	<0.002
Assumptions for scenario 3 (n=368)				
Missing physicians in game arm would have performed similarly those observed in educational arm and missing physicians in educational arm would have performed similarly those observed in game arm	0.55	0.62	0.07 (0.01 to 0.12)	0.02

Table 5 | Adherence, usability, and likeability of video game versus traditional educational apps. Figures are numbers (percentage) unless stated otherwise

Measure	Game		Educational apps	
	Data	Example of qualitative feedback	Data	Example of qualitative feedback
Median (IQR) time spent on intervention (min)*	90 (60-120)	NA	90 (65-120)	NA
Described problem with usability of apps	44 (30)	Not much of a gamer but enjoyed the app. Loved the mystery solving aspect of the game - found it annoying that if you accidentally clicked on "discharge" a "multi-trauma patient" for example, you could not go back and edit your mistake	12 (8)	The myATLS app is useful but the user interface could be improved…A better interface would make the otherwise useful info better. The review app has great questions, but is also in need of a better user interface. The questions are great and it's very responsive, but could be improved to the standard of other quiz apps
Commented positively about likeability of intervention†	60 (40)	The game, Night Shift, was a really fun experience. The story was interesting and had my attention to details as I wanted to solve the mystery and to connect the dots. It felt realistic and I could put myself in a position of the main character	135 (91)	The iPad apps were very user friendly. I completed questions while I was on a flight, and it was actually entertaining and made the time go by quickly. If I wanted to review a specific topic from a question, I could easily pull up the chapter
Commented negatively about likeability of intervention	89 (60)	The overall effect seemed more of a distraction than a help. I am not sure if I am to be more concerned with the "US NAVY" or the fact that elderly with trauma do better at a trauma hospital. The entire time I was playing the game I kept focusing on who wrote this the program and what their goal might be	13 (9)	myATLS was too superficial. More of an outline rather than substantive reading

NA=not applicable.
*Participants were asked to use their intervention(s) for minimum of one hour and to report their usage.
†Participants could provide both positive and negative feedback about their interventions.

education program (146/256 (0.57) v 172/232 (0.74); mean difference 0.17, 95% confidence interval 0.09 to 0.25; P<0.001).

Discussion
Principal findings
Exposure to a theoretically based video game changed the behavior of a convenience sample of physicians compared with those exposed to a traditional educational program. The change in behavior was concentrated on the cases we had designated as "non-representative" or least likely to evoke the common archetypes used to classify patients. The game exerted an effect on behavior despite the fact that physicians in the game arm were less likely to describe the intervention as usable or enjoyable.

We hypothesized that the game would recalibrate physicians' representativeness heuristics by changing their archetypes of patients with minor and severe injuries. We predicted that this change would manifest as a differential response to cognitive load. We found, however, that cognitive load did not affect physicians in either arm of our study. One explanation is that the experimental manipulation of cognitive load did not work. Given that the same manipulation affected performance in prior research, we speculate that the interventions interacted with the load manipulation, inducing a ceiling effect. An alternate explanation is that the game changed behavior through a mechanism other than heuristics. For example, the format could have convinced physicians of the need to transfer injured patients to trauma centers.

Players exposed to the game did not uniformly enjoy the experience. Physician dissatisfaction could reflect the well described observation from the video game literature that people have preferences for different genres of game (such as puzzle games).[48] Age could also play a role in physicians' reactions to Night Shift. Younger physicians, with greater tolerance for games used for training purposes, might represent a more suitable target population for this type of intervention.[49] Finally, usability issues with the game probably also affected enjoyment. The game itself was a proof of concept prototype, without the full production values of commercially available games that participants might have come to expect.

Design decisions for the intervention reflected our belief that effective interventions compensate for deficiencies in the experience-feedback loop by immersing or engaging the user in the training task.[17] We speculated that enjoyment would offer a means to that end and chose our mechanism (narrative) and delivery platform (video games) accordingly. We found that while adherence influenced performance, however, enjoyment did not. We therefore conclude that though enjoyment is one pathway to engagement, it is not the only one.

Strengths and limitations
Our study had several limitations. First, we selected participants attending a national conference; if they differ from the overall population of practicing emergency medicine physicians this would affect external validity. Use of a convenience sample would have affected both arms equally, however, rendering the results internally valid. Second, the simulation included only 10 cases with an enriched base rate of severe injuries, potentially introducing bias and precluding the assessment of individual physician performance.[3 50] Case volume imposes a well known barrier to reliable and valid estimation of individual physician performance.[51] One solution is to create instruments that focus on conditions of interest.[52] Another is to aggregate responses to assess group level performance.[53] We have previously validated our use

of simulation to measure performance by comparing the responses of emergency medicine physicians on the simulation with their practice patterns, finding that key decisions (such as acquisition of radiologic studies, disposition) match.[45] Third, physicians exposed the video game might have had an unfair advantage when completing the virtual simulation. We designed the game and simulation, however, with different objectives in mind (engagement versus assessment). As a result, the two products included different mechanics and interfaces (video 1 and video 2). In addition, both groups of physicians reviewed the same tutorial and completed a non-trauma case before beginning the trauma cases, further reducing any carryover effects. Readers who review the trial registration website will note a difference in our terminology regarding the primary outcome measure. In the trial registration, we used "under-triage rate" whereas here we use "proportion of under-triage." Though the literature on trauma triage uses these terms interchangeably, in the clinical trials literature, "rate" connotes events per time period, whereas proportion does not. Our use of rate and proportion interchangeably reflects imprecision in our language and not an attempt to manipulate outcome reporting.

Conclusions and policy implications

Trauma triage exemplifies the complexity and importance of diagnostic decisions made under time pressure and uncertainty. Severely injured patients treated at trauma centers have better outcomes than patients treated at non-trauma centers, including a 25% reduction in mortality, less disability at discharge, less pain at one year, and increased rates of returning to work.[54-56] About 55-80% of patients with severe injuries who present initially to non-trauma centers, however, are not transferred to a higher level of care, contributing to 30 000 preventable deaths each year.[21 23 57 58] Clinical practice guidelines in trauma instruct physicians to triage patients based on a history, physical exam, and chest and pelvic radiographs—ideally as rapidly as possible.[46 59] In other words, physicians must make their decisions quickly and with incomplete information. Additionally, most physicians have relatively little experience with severely injured patients: physicians working at a non-trauma center evaluate 1000 patients for every one with severe injuries.[60] These conditions make it extremely challenging to learn appropriate triage. Existing interventions, which emphasize physician knowledge of and attitudes towards the clinical practice guidelines, do not adequately deal with the challenges faced by physicians at non-trauma centers.

To address this, we developed a novel intervention that combined video game technology with narrative engagement to recalibrate physician heuristics and tested its efficacy in reducing diagnostic errors in simulated trauma triage. Our results suggest that narrative based video games have the potential to influence physician behavior, although the real world implications remain unclear.

1   National Academies of Sciences. *Engineering and Medicine. Improving diagnosis in health care.* The National Academies Press, 2015.
2   Kahneman D. *Thinking fast and slow.* Farrar, Straus, and Giroux, 2011.
3   Kahneman D, Tversky A. Judgment under uncertainty: Heuristics and biases. In: Kahneman D, Slovic P, Tversky A, eds. *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press, 2003: 3-20.
4   Gigerenzer G, Todd PM. ABC Research Group. *Simple heuristics that make us smart.* Oxford University Press, 1999.
5   Kahneman D, Frederick S. Representativeness revisited: attribute substitution in intuitive judgment. In: Gilovich T, Griffin D, Kahneman D, eds. *Heuristics and Biases: the Psychology of Intuitive Judgment.* Cambridge University Press, 2002doi:10.1017/CBO9780511808098.004.
6   Graber ML, Kissam S, Payne VL. Cognitive interventions to reduce diagnostic error: a narrative review. *BMJ Qual Saf* 2012;21:535-57. doi:10.1136/bmjqs-2011-000149.
7   Mamede S, van Gog T, van den Berge K. Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA* 2010;304:1198-203. doi:10.1001/jama.2010.1276.
8   Tudor GR, Finlay DB. Error review: can this improve reporting performance? *Clin Radiol* 2001;56:751-4. doi:10.1053/crad.2001.0760.
9   Wolpaw T, Papp KK, Bordage G. Using SNAPPS to facilitate the expression of clinical reasoning and uncertainties: a randomized comparison group trial. *Acad Med* 2009;84:517-24. doi:10.1097/ACM.0b013e31819a8cbf.
10 Raab SS, Grzybicki DM, Mahood LK, Parwani AV, Kuan SF, Rao UN. Effectiveness of random and focused review in detecting surgical pathology error. *Am J Clin Pathol* 2008;130:905-12. doi:10.1309/AJCPPIA5D7MYKDWF.
11 Ramnarayan P, Winrow A, Coren M. Diagnostic omission errors in acute paediatric practice: impact of a reminder system on decision-making. *BMC Med Inform Decis Mak* 2006;6:37. doi:10.1186/1472-6947-6-37.

12 Elstein AS. Thinking about diagnostic thinking: a 30-year perspective. *Adv Health Sci Educ Theory Pract* 2009;14(Suppl. 1):7-18. doi:10.1007/s10459-009-9184-0.

13 Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA* 1998;280:1339-46. doi:10.1001/jama.280.15.1339

14 Arkes HR. Costs and benefits of judgment errors: Implications for debiasing. *Psychol Bull* 1991;110:486-98. doi:10.1037/0033-2909.110.3.486.

15 Lichtenstein S, Fischhoff B. Training for calibration. *Organ Behav Hum Perform* 1991;26:149-71. doi:10.1016/0030-5073(80)90052-5.

16 Milkman KL, Chugh D, Bazerman MH. How Can Decision Making Be Improved? *Perspect Psychol Sci* 2009;4:379-83. doi:10.1111/j.1745-6924.2009.01142.x.

17 Kahneman D, Klein G. Conditions for intuitive expertise: a failure to disagree. *Am Psychol* 2009;64:515-26. doi:10.1037/a0016755.

18 Morewedge CK, Yoon H, Scopelliti I. Debiasing Decisions. *Policy Insights from the Behavioral and Brain Sciences* 2015;2:129-40. doi:10.1177/2372732215600886.

19 Anguera JA, Boccanfuso J, Rintoul JL. Video game training enhances cognitive control in older adults. *Nature* 2013;501:97-101. doi:10.1038/nature12486.

20 Idson LC, Chugh D, Bereby Meyer Y. Overcoming focusing failures in competitive environments. *J Behav Decis Making* 2004;17:159-72. doi:10.1002/bdm.467.

21 Delgado MK, Yokell MA, Staudenmayer KL, Spain DA, Hernandez-Boussard T, Wang NE. Factors associated with the disposition of severely injured patients initially seen at non–trauma center emergency departments: disparities by insurance status. *JAMA Surg* 2014;149:422-30. doi:10.1001/jamasurg.2013.4398.

22 Utter GH, Maier RV, Rivara FP, Mock CN, Jurkovich GJ, Nathens AB. Inclusive trauma systems: do they improve triage or outcomes of the severely injured? *J Trauma* 2006;60:529-37. doi:10.1097/01.ta.0000204022.36214.9e.

23 Zhou Q, Rosengart MR, Billiar TR, Peitzman AB, Sperry JL, Brown JB. Factors Associated With Nontransfer in Trauma Patients Meeting American College of Surgeons' Criteria for Transfer at Nontertiary Centers. *JAMA Surg* 2017;152:369-76. doi:10.1001/jamasurg.2016.4976.

24 Gomez D, Haas B, de Mestral C. Institutional and provider factors impeding access to trauma center care: an analysis of transfer practices in a regional trauma system. *J Trauma Acute Care Surg* 2012;73:1288-93. doi:10.1097/TA.0b013e318265cec2.

25 Mohan D, Barnato AE, Rosengart MR, Angus DC, Wallace DJ, Kahn JM. Triage patterns for medicare patients presenting to nontrauma hospitals with moderate or severe injuries. *Ann Surg* 2015;261:383-9. doi:10.1097/SLA.0000000000000603.

26 Mohan D, Barnato AE, Angus DC, Rosengart MR. Determinants of compliance with transfer guidelines for trauma patients: a retrospective analysis of CT scans acquired prior to transfer to a Level I Trauma Center. *Ann Surg* 2010;251:946-51. doi:10.1097/SLA.0b013e3181d76cb5.

27 Mohan D, Rosengart MR, Farris C, Fischhoff B, Angus DC, Barnato AE. Sources of non-compliance with clinical practice guidelines in trauma triage: a decision science study. *Implement Sci* 2012;7:103. doi:10.1186/1748-5908-7-103.

28 Mohan D, Rosengart MR, Fischhoff B. Testing a videogame intervention to recalibrate physician heuristics in trauma triage: study protocol for a randomized controlled trial. *BMC Emerg Med* 2016;16:44. doi:10.1186/s12873-016-0108-z.

29 Sanfey AG, Loewenstein G, McClure SM, Cohen JD. Neuroeconomics: cross-currents in research on decision-making. *Trends Cogn Sci* 2006;10:108-16. doi:10.1016/j.tics.2006.01.009.

30 Gillard E, Van Dooren W, Schaeken W, Verschaffel L. Proportional reasoning as a heuristic-based process: time constraint and dual task considerations. *Exp Psychol* 2009;56:92-9. doi:10.1027/1618-3169.56.2.92.

31 Cabana MD, Rand CS, Powe NR. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999;282:1458-65. doi:10.1001/jama.282.15.1458

32 Kahneman D. A perspective on judgment and choice: mapping bounded rationality. *Am Psychol* 2003;58:697-720. doi:10.1037/0003-066X.58.9.697.

33 Rammstedt B, John OP. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *J Res Pers* 2007;41:203-12. doi:10.1016/j.jrp.2006.02.001.

34 Nabi RL, Green MC. The Role of a Narrative's Emotional Flow in Promoting Persuasive Outcomes. *Media Psychol* 2014;18:137-62. doi:10.1080/15213269.2014.912585.

35 Moyer Gusé E, Nabi RL. Explaining the Effects of Narrative in an Entertainment Television Program: Overcoming Resistance to Persuasion. *Hum Commun Res* 2010;36:26-52. doi:10.1111/j.1468-2958.2009.01367.x.

36 Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. *Br J Surg* 2012;99:1322-30. doi:10.1002/bjs.8819.

37 Miller-Day M, Hecht ML. Narrative means to preventative ends: a narrative engagement framework for designing prevention interventions. *Health Commun* 2013;28:657-70. doi:10.1080/10410236.2012.762861.

38 Kapoor R, Saint S, Kapoor JR, Johnson RA, Dhaliwal G. Clinical problem-solving. D is for delay. *N Engl J Med* 2014;371:2218-23. doi:10.1056/NEJMcps1212211.

39 Tsai TT, Nallamothu BK, Prasad A, Saint S, Bates ER. Clinical problem-solving. A change of heart. *N Engl J Med* 2009;361:1010-6. doi:10.1056/NEJMcps0903023.

40 Safdar N, Odden A, Abad CL, Thapa R, Saint S. Clinical problem-solving. In sight and out of mind. *N Engl J Med* 2015;372:2218-23. doi:10.1056/NEJMcps1413402.

41 Bliss SJ, Flanders SA, Saint S. Clinical problem-solving. A pain in the neck. *N Engl J Med* 2004;350:1037-42. doi:10.1056/NEJMcps032253.

42 Drapkin MS, Kamath RS, Kim JY. Case records of the Massachusetts General Hospital. Case 26-2012. A 70-year-old woman with fever and back pain. *N Engl J Med* 2012;367:754-62. doi:10.1056/NEJMcpc1200087.

43 About Advanced Trauma Life Support. https://www.facs.org/quality%20programs/trauma/atls/about.

44 American College of Surgeons - Committee on Trauma. myATLS. http://www.myatls.com.

45 Mohan D, Angus DC, Ricketts D. Assessing the validity of using serious game technology to analyze physician decision making. *PLoS One* 2014;9:e105445. doi:10.1371/journal.pone.0105445.

46 Committee on Trauma - American College of Surgeons. Resources for the optimal care of the injured patient. https://www.facs.org/~/media/files/quality%20programs/trauma/vrc%20resources/resources%20for%20optimal%20care.ashx

47 Cohen J. A power primer. *Psychol Bull* 1992;112:155-9. doi:10.1037/0033-2909.112.1.155

48 Madigan J. *The psychology of immersion in video games*. The Psychology of Video Games 2016.

49 Entertainment Software Association. Essential facts about the computer and video game industry. http://www.theesa.com/wp-content/uploads/2017/04/EF2017_FinalDigital.pdf.

50 Birnbaum MH. Base Rates in Bayesian Inference: Signal Detection Analysis of the Cab Problem. *Am J Psychol* 1983;96:85. doi:10.2307/1422211.

51 Nyweide DJ, Weeks WB, Gottlieb DJ, Casalino LP, Fisher ES. Relationship of primary care physicians' patient caseload with measurement of quality and cost performance. *JAMA* 2009;302:2444-50. doi:10.1001/jama.2009.1810.

52 Sirovich B, Gallagher PM, Wennberg DE, Fisher ES. Discretionary decision making by primary care physicians and the cost of U.S. Health care. *Health Aff (Millwood)* 2008;27:813-23. doi:10.1377/hlthaff.27.3.813.

53 Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA* 2000;283:1715-22. doi:10.1001/jama.283.13.1715.

54 Mackenzie EJ, Rivara FP, Jurkovich GJ. The National Study on Costs and Outcomes of Trauma. *J Trauma* 2007;63(Suppl):S54-67, discussion S81-6. doi:10.1097/TA.0b013e31815acb09.

55 MacKenzie EJ, Weir S, Rivara FP. The value of trauma center care. *J Trauma* 2010;69:1-10. doi:10.1097/TA.0b013e3181e03a21.

56 Celso B, Tepas J, Langland-Orban B. A systematic review and meta-analysis comparing outcome of severely injured patients treated in trauma centers following the establishment of trauma systems. *J Trauma* 2006;60:371-8, discussion 378. doi:10.1097/01.ta.0000197916.99629.eb.

57 National Academies of Sciences, Engineering, and Medicine. *A national trauma care system: integrating military and civilian trauma systems to achieve zero preventable deaths after injury.* The National Academies Press, 2016.

58 Lale A, Krajewski A, Friedman LS. Undertriage of Firearm-Related Injuries in a Major Metropolitan Area. *JAMA Surg* 2017;152:467-74. doi:10.1001/jamasurg.2016.5049.

59 American College of Surgeons – Committee on Trauma. *Advanced Trauma Life Support for Doctors.* American College of Surgeons, 2008.

60 Mohan D, Barnato AE, Rosengart MR. Trauma triage in the emergency departments of nontrauma centers: an analysis of individual physician caseload on triage patterns. *J Trauma Acute Care Surg* 2013;74:1541-7. doi:10.1097/TA.0b013e31828c3f75.

Appendix 1: Supplementary information