Statistical Analysis Plan

Developing a deliberate practice intervention to recalibrate physician heuristics in trauma triage
NIA R21AG072072
Clinicaltrials.gov identifier #NCT05168579
November 25, 2022

Assessment framework

We used Proctor's Framework of Implementation Outcomes in Research to structure our process assessment of the novel intervention. We defined the implementation outcomes as feasibility, fidelity, acceptability, adoption, and appropriateness. Using the NIA's Stage Model of Intervention Development which recommends assessment of efficacy in the laboratory before moving to real-world testing, we defined the service outcome (efficacy) as compliance with clinical practice guidelines, measured using simulation.

Data sources and management

Screening questionnaire and tracking database

Each respondent completed a questionnaire with items on their age, sex, race, ethnicity, and educational background at the time of enrollment. We maintained a database with a list of scheduled coaching sessions, which was updated daily with the status of the sessions.

Coaching sessions

We recorded all the coaching sessions and automatically uploaded them to a secure server hosted by the University of Pittsburgh. Two members of the study team (KJR, JLB) developed a codebook to assess the delivery of core session components, refined it until they achieved acceptable inter-rater reliability (Cohen's kappa 0.84), and independently applied it to the recordings. Coding discrepancies were resolved through consensus (KJR, JLB, DM). We used NVivo qualitative analysis software (QSR International, Melbourne, Australia) for data management.

Post-intervention debriefing materials

Participants in the intervention group provided structured assessments of the acceptability of the intervention using the User Engagement Scale–Short Form to evaluate the video game (a validated 12-item instrument with a 5-point Likert scale), and the Wisconsin Surgical Coaching Rubric to evaluate the quality of the coaching (a 4-item instrument with a 5-point scale). They also participated in semi-structured debriefing interviews after the final coaching session, during which they discussed their perception of the acceptability, adoption, and appropriateness of the intervention. Two members of the study team (KJR, JLB) developed a codebook, refined it until they achieved an acceptable inter-rater reliability (kappa 0.84), and independently applied it to transcripts of the interviews. Any coding discrepancies were resolved through consensus (KJR, JLB, DM).

Simulation to measure efficacy

We used a validated 2-D simulation to assess changes in physician behavior after exposure to the intervention (i.e., compliance with clinical practice guidelines). The simulation required participants to respond to 10 cases over 42 minutes: 4 severely injured patients, 2 minimally injured patients, and 4 critically ill non-trauma patients. New patients arrived at pre-specified but unpredictable intervals, so that users managed multiple patients concurrently. Without appropriate clinical intervention by the player, severely injured patients and critically ill distractor patients decompensated and died over the course of the simulation. Each case included a 2-D

rendering of the patient, a chief complaint, vital signs which updated every 30 seconds, a history, and a written description of the physical exam. Users could request more information by selecting from a pre-specified list of 250 medications, studies, and procedures. They could place orders, and consult specialty services. Each case ended when the player either made a disposition decision (admit, discharge, transfer) or the patient died.

We asked all trial participants to complete the simulation online, and responses were uploaded and stored on a secure server hosted by the University of Pittsburgh.

Analyses

We summarized physician characteristics using means (standard deviations [SD]) for continuous variables and number (%) for categorical variables. We analyzed implementation outcomes using an intention-to-treat approach, but excluded from the efficacy analysis participants who did not use the simulation. We had two criteria for the success of the trial: efficacy and feasibility. Our primary hypothesis was that physicians exposed to the intervention would under-triage ≥25% fewer patients on the simulation than physicians in the control arm. Our secondary hypothesis was that ≥90% of participants would receive all three coaching sessions.

Implementation outcomes

We quantified the proportion of coach-participant dyads that completed three thirty-minute sessions (to measure feasibility), and summarized the proportion of core components delivered to participants (to measure fidelity). We summarized participant responses to the User Engagement Scale-Short Form and to the Wisconsin Surgical Coaching Rubric (to measure acceptability). We also summarized themes that arose during the semi-structured interviews (to further assess acceptability and to assess appropriateness and adoption).

Efficacy

We summarized time spent and decisions made for each severely injured trauma case on the simulation (e.g., diagnostic testing, administration of blood products), and scored disposition decisions as consistent with the American College of Surgeons guidelines or not. To compare differences between the intervention and control groups, we used generalized linear models, clustered at the participant level. In a post-hoc sensitivity analysis, we excluded physicians who had previously participated in our research.

Human subjects and power calculation

We designed the experiment to detect a 25% (large effect size) reduction in under-triage between physicians in the intervention and control groups, with an alpha of 0.05 and a power of 80%, using Cohen's method of estimating power for behavioral trials. Based on these estimates, and anticipating a 67% retention rate in the control arm, we planned to recruit 30 physicians for each arm.