STATISTICAL ANALYSIS PLAN

Using a theory-based, customized video game as an educational tool to improve the trauma triage decisions of emergency department physicians: a randomized clinical trial

---

Statistical Analysis Plan: Version 1.5

Updated March 6, 2026

DSMB Protocol: Version 1.5 (October 3, 2025)

ACRONYM:	**Night Shift**
VERSION:	**1.0 (11/13/2023)** **1.1 (12/7/2023)** **1.2 (2/7/2024)** **1.3 (4/30/2024)** **1.4 (9/29/2025)** **1.5 (10/3/2025)**
PROTOCOL CHAIR:	**Deepika Mohan, MD, MPH** Associate Professor of Surgery and Critical Care Medicine University of Pittsburgh School of Medicine
FUNDED BY:	**National Institute on Aging (R01AG076499)**
DATE:	**March 6, 2026**
NATIONAL CLINICAL TRIAL (NCT) NO.	NCT06063434

TRANSPARENCY NOTE

This Statistical Analysis Plan (SAP) was reconstructed on 6 March 2026, after trial completion, from analytical content embedded in the DSMB Protocol submitted to the National Institute on Aging, and updated over time. This reconstruction was undertaken in response to peer review feedback noting that a standalone SAP was required by internationally accepted clinical research standards.

The content of this SAP reflects only analyses described in edited DSMB Protocol. No new analyses have been introduced during reconstruction. It should be read together with the table in the Appendix, comparing analytic elements across study documents.

Revision History Table

Version	Date	Section	Description of Change	Rationale
1.0	11/11/23	–	Initial protocol with statistical analysis plan approved by DSMB	-
1.1	12/7/23	3.1.1	Inclusion criteria broadened to include all participants who agree to participate without consideration of demographic characteristics	Slow rate of enrollment
		3.1.1	Inclusion criteria broadened to include physicians employed at Level 3–5 trauma centers	Expert advisory panel recommendation to address slow rate of enrollment
		3.2.1	Inclusion criteria changed to define eligible patients as those who present initially to Level 3–5 trauma centers (and not just non-designated acute care hospitals)	See above
1.2	2/7/24	3.2.1	Inclusion criteria changed to define eligible patients as treated within 365 days of the physicians' enrollment, instead of between 1 January 2024–31 December 2024.	Delay in enrollment
1.3	4/30/25	3.2.1	Inclusion criteria changed to exclude patients with Medicare Advantage coverage	Cost of data purchase
		5.2.1	Secondary outcome measure of functional independence delayed	Cost of data purchase
1.4	9/29/25	7.5	Primary analysis will not include baseline covariates or random effects for the hospital	Parsimony; model stability
		7.7	Add a complier average causal effect analysis (CACE) to per protocol analysis to test the effect of study protocol compliance	Robustness
		7.10	One-sided alpha changed to a two-sided test	Improve rigor of interpretation
DATA LOCK OCCURS – October 1, 2025				
1.5	10/3/25	5.3	Re-defined dose as game usage during each intervention period instead of based on time spent.	Improve understanding of mediation
		7.1	Remove exploratory analysis of fidelity of intervention receipt; remove mediation analysis of clinical outcomes	Inability to acquire correct statistical package on CMS Virtual Desktop; results preclude necessity

SUMMARY

Study Population: The primary study population consists of 800 board-certified emergency physicians employed at non-trauma centers in the continental US. Outcomes are assessed using Medicare claims from patients aged 65 or older treated by trial physicians during the one-year post-enrollment period.

Design: This is a Phase 3, Type 1 hybrid effectiveness-implementation randomized clinical trial in which physicians are randomized 1:1 to play *Night Shift 2024* (game-based training) or to receive usual education (control). The primary outcome is assessed using Medicare claims data collected for 13 months after the close of enrollment.

Sample Size: A total of 800 physicians (400 per arm) provides 80% power to detect a 7.4% to 10% absolute reduction in under-triage at a one-sided significance level of 0.05, assuming an intraclass correlation coefficient of 0.45 and a median of 1-2 eligible patients per physician.

TABLE OF CONTENTS

1. ABBREVIATIONS 7
2. BACKGROUND AND RATIONALE 8
3. ENROLLMENT CRITERIA 8
3.1. Physician Participants 8
3.1.1. Inclusion criteria 9
3.1.2. Exclusion criteria 9
3.2. Patient Participants 9
3.2.1. Inclusion criteria 9
3.2.2. Exclusion criteria (primary analysis) 9
3.3. Hospital Characteristics 9
3.4. Key Decision Makers 10
3.4.1. Inclusion criteria (must meet at least one of the following) 10
3.4.2. Exclusion criteria 10
4. INTERVENTIONS 10
4.1. Intervention Arm (Game-Based Training) 10
4.1.1. Dosing and administration 10
4.1.2. Measurement of fidelity of treatment delivery 11
4.1.3. Measurement of fidelity of treatment receipt 11
4.2. Control Arm (Usual Education) 11
4.3. Concomitant Interventions 11
5. ENDPOINTS 11
5.1. Primary Endpoint 11
5.2. Secondary Endpoints 12
5.2.1. Patient outcomes 12
5.2.2. Safety outcome 12
5.3. Exploratory Endpoints 12
5.3.1. Heterogeneity of treatment effect 12
5.3.2. Mechanisms of action 12
5.4. Implementation Endpoints 13
6. TRIAL CONDUCT 13
6.1. Randomization 13
6.2. Blinding 13
6.3. Data Collection 13
6.4. Interim Analyses 14
6.5. Stopping Rules 14
6.6. Protocol Deviations 14
6.7. Participant Withdrawal 14
7. STATISTICAL ANALYSIS PLAN 14
7.1. Statistical Objectives 14
7.2. Sample Size 15
7.3. Analysis Populations 15
7.3.1. Intention-to-treat population 15
7.3.2. Per-protocol population 15
7.3.3. Implementation analysis population 15
7.4. General Analytic Principles 16
7.5. Primary Endpoint Analysis 16
7.6. Secondary Endpoint Analysis 17

7.6.1. Patient outcomes 17
7.6.2. Safety outcomes 17
7.7. Exploratory Analyses 17
7.7.1. Protocol Deviations 17
7.7.2. Heterogeneity of treatment effect 17
7.7.3. Mediation of the game's effect on under-triage 18
7.7.4. Mediation of the game's effect on clinical outcomes 18
7.8. Implementation Analyses 18
7.9. Handling of Missing Data 18
7.10. Type I Error Rate 19
8. ETHICAL CONSIDERATIONS 19
8.1. Regulatory and Ethical Oversight 19
8.2. Informed Consent 19
8.3. Data Safety and Monitoring Board 19
8.4. Adverse Events 20
8.5. Confidentiality and Data Protection 20
8.6. Conflict of Interest 20
9. FUNDING 20
10. APPENDIX 21
10.1. Comparison of analytic elements across study documents 21

1. ABBREVIATIONS

AE	Adverse Event
ATLS	Advanced Trauma Life Support
CACE	Complier Average Causal Effect
CFR	Code of Federal Regulations
CI	Confidence Interval
CMS	Centers for Medicare and Medicaid
CONSORT	Consolidated Standards of Reporting Trials
DSMB	Data and Safety Monitoring Board
ED	Emergency Department
GLMM	Generalized Linear Mixed Model
HCRIS	Healthcare Cost Report Information System
HIPAA	Health Insurance Portability and Accountability Act
HRPO	Human Research Protection Office
ICC	Intra-class Correlation Coefficient
ICH GCP	International Council on Harmonization Good Clinical Practice
ICPSR	Inter-university Consortium for Political and Social Research
ISS	Injury Severity Score
ITT	Intention-to-Treat
MAR	Missing at Random
MCAR	Missing Completely At Random
MICE	Multivariate Imputation by Chained Equations
MNAR	Missing Not At Random
NIA	National Institute on Aging
NIH	National Institutes of Health
NPI	National Provider Identifier
PP	Per Protocol
SAE	Serious Adverse Event
SAP	Statistical Analysis Protocol
TIEP	Trauma Information Exchange Program
US	United States
VPN	Virtual Private Network

2. BACKGROUND AND RATIONALE

Trauma disproportionately affects older adults, who account for nearly half of all injury-related hospital admissions and two-thirds of in-hospital injury-related deaths in the United States. The cornerstone of trauma care is timely transfer of severely injured patients to trauma centers, where mortality is reduced by 10-25% and rates of return to independent living are substantially improved. Guidelines from the American College of Surgeons specify the injury patterns that warrant transfer. Despite four decades of quality improvement efforts, under-triage (the failure to transfer severely injured patients) remains common, affecting approximately 50% of patients at non-trauma centers overall and more than 80% of patients over the age of 65.

Physicians represent the largest identifiable source of variation in triage practices, accounting for up to 40% of the variance in transfer decisions. A critical but underappreciated driver of this variation is diagnostic error rooted in poorly calibrated clinical heuristics. Older trauma patients frequently lack the classic physiologic markers of severe injury (hypotension, tachycardia, penetrating wounds) that trigger pattern recognition in physicians trained on younger populations. Compounding this, ED physicians at non-trauma centers rarely encounter severely injured patients (approximately 1 in every 1,000 presentations), receive delayed or no feedback on their triage decisions, and are therefore unable to accumulate the experience needed to recalibrate their intuitive judgments. Existing educational interventions, including Advanced Trauma Life Support (ATLS), are designed around a rational actor model of decision making and have had limited impact on real-world triage behavior.

This trial is grounded in dual-process theory, which distinguishes between fast, heuristic-based (System 1) and deliberate, rule-based (System 2) cognitive processes. Rather than dismissing heuristics as flawed, the intervention is designed to recalibrate them by bringing intuitive judgments into alignment with clinical practice guidelines through structured surrogate experience and feedback. *Night Shift 2024* is a theory-based adventure video game developed iteratively by a team of game designers, behavioral scientists, and clinicians. In prior randomized trials conducted in laboratory settings (n = 368 and n = 320), physicians randomized to the game demonstrated 11-18% lower rates of under-triage on a validated virtual simulation compared to those receiving text-based education, with effects persisting through six-month follow-up. A preliminary analysis using Medicare claims data further suggested that game-exposed physicians under-triaged fewer patients in the year following the trial, though the difference-in-difference estimate did not reach statistical significance due to limited sample size.

The present trial is the next logical step in this program of research: a pragmatic, real-world test of whether the intervention changes physician triage behavior in practice, reduces adverse patient outcomes, and can be implemented sustainably at scale. The trial has three objectives: (1) to compare the effectiveness of *Night Shift 2024* versus usual education on real-world triage decisions and patient outcomes; (2) to compare the mechanisms by which the game influences physician behavior; and (3) to identify contextual factors that influence the implementation and maintenance of clinical practice guidelines in trauma triage.

3. ENROLLMENT CRITERIA

3.1. Physician Participants

Physicians are the unit of randomization and the primary study participants.

3.1.1. Inclusion criteria

- Board-certified physician
- Holds a National Provider Identifier (NPI)
- Works exclusively in the emergency department of a non-trauma center in the continental United States (US)
- Responsible for the triage of adult injured patients
- Attests to willingness to adhere to the trial protocol and, if assigned to the intervention arm, play the video game for the required duration
- Provides digital informed consent.

3.1.2. Exclusion criteria

- Not board-certified
- Not a physician (e.g., nurse practitioner, physician assistant)
- Currently providing services at both trauma and non-trauma centers
- Works outside the continental US (e.g., Puerto Rico, Alaska, Hawaii).

Protocol Deviation (Version 1.1): Inclusion criteria broadened to include all respondents who agreed to participate, instead of stratifying enrollment to increase the proportion of minority (women and/or non-white) physicians, and physicians working at Level 3–5 trauma centers. We therefore defined a "non-trauma" center as any non-Level 1 or 2 trauma center.

3.2. Patient Participants

Patients are not enrolled as active participants but are identified retrospectively through Medicare claims for the purpose of outcome assessment.

3.2.1. Inclusion criteria

- Medicare Fee-for-Service or Advantage beneficiaries
- Patients ≥65 years old
- Any ICD-10 code for physical injury (S00-T99)
- ICD-10 code indicates initial visit (suffix "A" or "B")
- Received treatment at a non-trauma center in the continental US
- Treated by a trial physician between January 1, 2024 – December 31, 2024.

3.2.2. Exclusion criteria (primary analysis)

- Severe injury (ISS ≥ 16) with death on the day of admission.

Protocol Deviation (Versions 1.3 and 1.2): We did not obtain Medicare Advantage claims, and so changed our inclusion criteria to patients 65 years or older with Medicare Fee-for-Service only. We also included patients treated in the 365 days after enrollment, given delays in meeting our sample size target. Finally, we included patients who presented to Level 3–5 trauma centers initially (as discussed above).

3.3. Hospital Characteristics

Hospitals are not enrolled as participants but are characterized using the CMS Healthcare Cost Report Information System (HCRIS) and the Trauma Information Exchange Program (TIEP).

Relevant variables include geographic location, hospital network participation, total and ICU bed count, ownership, teaching status, and trauma center designation. These data are used to describe the practice environments of trial physicians and to inform understanding of contextual factors influencing treatment decisions.

3.4. Key Decision Makers

Key decision makers are enrolled separately for the implementation sub-study.

3.4.1. Inclusion criteria (must meet at least one of the following)

- Paramedic who treats adult injured patients
- ED director at a non-trauma center
- Trauma director at a trauma center
- Hospitalized for a severe injury within one year of recruitment
- Primary caregiver of someone hospitalized with a severe injury within one year of recruitment
- Participant in the intervention arm of the effectiveness trial.

3.4.2. Exclusion criteria

- None.

4. INTERVENTIONS

4.1. Intervention Arm (Game-Based Training)

Participants randomized to the intervention receive *Night Shift 2024*, a theory-based video game developed to recalibrate physician heuristics in trauma triage. The game is preloaded on a new iPad, which is mailed to participants upon enrollment and kept as a fixed honorarium. The game has two components. During the adventure game, *Night Shift*, the player follows Andy Jordan, a young emergency physician who takes night shifts at a community hospital while investigating his grandfather's disappearance. Through this storyline, players manage a series of trauma cases and receive feedback on their triage decisions, designed to evoke emotional engagement and reinforce evidence-based decision principles. Night Shift 2024 also includes a mini puzzle game, called *Graveyard Shift. Graveyard Shift* has 10 levels, each of which takes 5-6 minutes to complete. During each level, the player must triage 10 cases over 90 seconds, 5 of which exemplify a single triage decision principle, then review 2 of the cases to derive the principle independently, and finally receive summative feedback. We will release different levels of the game through the year, to keep participants engaged in game-based training.

4.1.1. Dosing and administration

Participants are asked to complete a minimum of 2 hours of gameplay within 2 weeks of receiving their device, followed by three 20-minute booster sessions at approximately 4, 7, and 10 months post-enrollment. Participants receive three email reminders and a phone call during the first two weeks, followed by quarterly reminders thereafter. A conditional honorarium of $20 per completed booster session is provided. Participants also keep the trial iPad at the end of the year, as part of their honorarium.

4.1.2. Measurement of fidelity of treatment delivery

We define fidelity of treatment delivery as adherence to the study protocol. The application automatically records time spent, number of visits, and progress through the game. These data are uploaded to a secure University of Pittsburgh server each time the device connects to a wireless network. Participants are additionally asked to self-report game usage and complete the User Engagement Scale Short Form (a validated 12-item instrument) after completing the first 2 hour game-play session.

4.1.3. Measurement of fidelity of treatment receipt

We define intervention receipt as evidence that participants understand and can use the skills or knowledge learned during the intervention, and will evaluate it in three ways. First, we will ask participants (intervention and control) to describe their understanding of the central principle of trauma triage on the first post-enrollment survey. Second, we will ask them to complete an online tool that assesses physician behavior in trauma triage (SONAR). SONAR is a 2D web-based serious game, during which physicians must triage 30 trauma patients (15 with severe injuries; 15 with minor injuries). The game evaluates decisions using the guidelines published by the American College of Surgeons – Committee on Trauma, and produces two sets of metrics: compliance with guidelines (e.g., under-triage) and signal detection parameters. Signal detection theory came to prominence during World War II and allows inferences about the sources of non-compliance with clinical practice guidelines by parsing the influence of perceptual sensitivity (the ability to distinguish between minimally and severely injured patients) and decisional thresholds (preferences to err on the side of under- or over-triage) on decisions.

4.2. Control Arm (Usual Education)

Participants randomized to the control receive on usual education, consisting of continuing medical education mandated by state medical licensing boards, professional organizations, and hospital credentialing authorities. We will not provide any supplemental educational materials. Control participants complete online assessment tools and receive a conditional wage-based honorarium of $100 per hour spent on study tasks.

4.3. Concomitant Interventions

There are no restrictions on concomitant educational activities for either arm. Participants in both arms are expected to continue their usual professional development activities throughout the trial period.

5. ENDPOINTS

5.1. Primary Endpoint

The primary endpoint is **under-triage**, defined as 1 minus the proportion of severely injured patients (ISS $\geq$ 16) treated by trial physicians who were transferred to a trauma center. This endpoint is assessed using Medicare claims data over the one-year period following physician enrollment and captures the core behavioral target of the intervention.

5.2. Secondary Endpoints

5.2.1. Patient outcomes

- **30-day mortality and readmission**s: A composite outcome capturing death or unplanned hospital readmission within 30 days of the index visit, assessed among severely injured patients treated by trial physicians.
- **Functional dependence:** The proportion of patients with a pre-injury residence at home (identified via a 90-day lookback in claims) who are discharged to a rehabilitation facility or skilled nursing facility, used as a proxy for loss of functional independence.

5.2.2. Safety outcome

- **Over-triage:** The proportion of patients with minimal injuries who are transferred to a trauma center, assessed to evaluate potential harm associated with the intervention.

Protocol Deviation (Protocol Deviation 1.3): The ClinicalTrials.gov registration (NCT06063434) and original statistical analysis plan included functional independence as a secondary clinical outcome that that we did not measure, because we could not purchase Medicare Skilled Nursing Facility claims.

5.3. Exploratory Endpoints

5.3.1. Heterogeneity of treatment effect

We will evaluate heterogeneity of the treatment effect across the following physician characteristics:

- Sex (male vs. female physicians)
- Race/ethnicity (White vs. non-White physicians)
- Attitudes toward video games prior to enrollment (play games versus do not play games)

5.3.2. Mechanisms of action

We will evaluate mediators of the game effect:

- Fidelity of intervention delivery:
  - Dose of intervention (game usage greater than vs. less than 2 hours)
  - Timing between exposure – treatment of patient (first 30 days post-usage vs. days 30-89 versus 90+ days)
- Fidelity of intervention receipt – high versus low perceptual sensitivity; high versus low decisional threshold)

Protocol Deviation (Protocol Deviation 1.5): We changed our classification of "dose" for this analysis from time ≥2 hours or <2 hours to participation in game usage during each intervention period for any amount of time (4 total). We deferred evaluation of mediation by fidelity of intervention receipt because we could not download the needed statistical package to the CMS Virtual Desktop.

5.4. Implementation Endpoints

Barriers and facilitators of adoption, implementation, and maintenance of guideline-concordant trauma triage are assessed through semi-structured interviews with a subset of intervention-arm physicians (n = 20) and key decision makers (n = 80-100). These endpoints are analyzed using qualitative methods and are reported separately from the quantitative effectiveness analyses.

6. TRIAL CONDUCT

6.1. Randomization

Physicians are randomized 1:1 to the intervention or control arm using a permuted block design with variable block sizes of 4 or 6, generated by the study statistician in Stata 17.0 (StataCorp, Texas). Trial coordinators link the randomization schema to enrollment data and assign participants to groups. To address equity, minority physicians (female, non-White) are oversampled and may comprise up to 50% of the final cohort.

The use of variable block sizes of 4 or 6 is disclosed in this public protocol, which creates a theoretical risk that future allocation sequences could be predicted by individuals with access to sequential enrollment data. This risk is mitigated by the following four features of the trial design:

- enrollment is conducted entirely remotely and centrally via an online platform, with no face-to-face interaction between recruiting staff and participants
- allocation assignments are masked to all trial staff until after the completion of data cleaning
- use of variable rather than fixed block sizes reduces predictability even when the range of block sizes is publicly known
- primary outcome is assessed using Medicare Fee-for-Service claims, an administrative data source entirely independent of the enrollment and allocation process.

Protocol Deviation (Protocol Deviation 1.1): We included all eligible respondents without attempting to oversample minority physicians because of difficulty finding willing participants.

6.2. Blinding

Full blinding is not possible in this trial given the nature of the intervention. Participants are aware of their assigned arm following allocation. However, data analysts are kept masked to treatment assignment throughout the analysis period to minimize the risk of bias in outcome assessment.

6.3. Data Collection

Four types of source data are collected:

- **Physician self-reported data:** Baseline demographic and professional characteristics are collected via an online Qualtrics survey at enrollment. Game usage statistics are collected from the *Night Shift 2024* application. Responses to the User Engagement Scale Short Form are collected from a trial portal.
- **Physician experimental data**: Physicians completed a virtual simulation (SONAR) in which they responded to 36 cases (26 with injury), allowing us to calculate signal

detection theory metrics: perceptual sensitivity and decisional threshold. These metrics allow understanding of fidelity of intervention receipt.

- **Medicare claims data:** Inpatient, Outpatient, Skilled Nursing Facility, and Professional Claims filed with Medicare Fee-for-Service and Advantage for beneficiaries aged 65 or older with ICD-10 codes associated with injuries are obtained from CMS for the trial period (2024), with flanking quarters (Q4 2023 and Q1 2025) to capture preceding and follow-up care. Claims are used to ascertain patient eligibility, injury severity, triage decisions, and patient outcomes.
- **Qualitative data:** Audio recordings and transcripts of semi-structured interviews with trial physician participants and key decision makers are collected via Pitt Zoom.

All primary quantitative and qualitative trial data are stored on the University of Pittsburgh Sharepoint. All CMS data are stored on the Virtual Desktop. Participants are assigned unique study identifiers at enrollment; identifying information is stored separately from research data.

Protocol Deviation (Protocol Deviation 1.3): As noted above, we did not obtain Medicare Advantage claims, nor did we obtain Medicare Skilled Nursing Facility claims.

6.4. Interim Analyses

No interim analysis is planned for this trial.

6.5. Stopping Rules

Given the minimal risk nature of the trial, no pre-specified stopping rules are defined. The Data and Safety Monitoring Board (DSMB) meets every nine months to review safety and efficacy data and retains the authority to recommend early termination if unanticipated safety concerns arise.

6.6. Protocol Deviations

Protocol deviations are defined as any noncompliance with the study protocol. The impact of protocol deviations on the primary analysis will be evaluated through a per protocol and a complier average causal effect (CACE) analysis, described in Section 7.7.

6.7. Participant Withdrawal

Participants may withdraw from the trial at any time. Upon withdrawal, self-reported data are excluded from analysis unless the participant consents to their continued use. The primary analysis uses an intention-to-treat approach and includes all randomized physicians regardless of their degree of participation, provided they filed Medicare claims during the follow-up period, and did not ask for withdrawal.

7. STATISTICAL ANALYSIS PLAN

7.1. Statistical Objectives

The primary statistical objective is to determine whether physicians randomized to *Night Shift 2024* under-triage a smaller proportion of severely injured older patients compared to those receiving usual education, as assessed using Medicare claims data over the one-year post-

enrollment period. Secondary objectives are to evaluate the effect of the intervention on patient outcomes (30-day mortality and readmissions, functional dependence) and to assess safety (over-triage).

Exploratory objectives include evaluating: 1) the heterogeneity of treatment effect across pre-specified physician-level subgroups; 2) the mediation of the intervention's effect by dose, time between game usage and treatment of patient, and fidelity of intervention receipt; 3) the mediation of clinical outcomes by the intervention

Protocol Deviation (Version 1.3 and 1.5): We did not assess the effect of the game on functional dependence, mediation by fidelity of intervention receipt, or mediation by the game.

7.2. Sample Size

Based on prior studies, we assume that ≥75% of enrolled physicians will encounter at least one eligible patient (i.e., enrolled in Medicare Fee-for-Service, severely injured, age ≥65 years), with a median number of patients per physician of 1-2. In this clustered randomization trial, physicians serve as the randomization unit. Based on data from a prior study, we assume an intraclass correlation coefficient (ICC) of 0.45, indicating the correlation in outcomes within the same physician. Under these assumptions, with 400 physicians per group (N=800), we can detect between a 7.4–10% difference in under-triage between the intervention and control groups with 80% power and a significance level of 0.05, using a one-sided hypothesis test. We have chosen a one-sided hypothesis test to boost statistical efficiency and because we prioritize identification of a positive effect of the intervention. A negative effect would produce the same outcome as a null: the decision to pause further efforts to disseminate the intervention.

The sample size estimate is conservative, as it is based on a simple comparison of proportions. The mixed-effects model used in the primary analysis is expected to yield greater effective power by accounting for additional sources of variance. Sample size calculations were performed using PASS 2023.

7.3. Analysis Populations

7.3.1. Intention-to-treat population

The intention-to-treat (ITT) population includes all randomized physicians, regardless of their adherence to the intervention, who filed at least one Medicare claim during the follow-up period and who did not ask to have their data withdrawn from the trial. This is the primary analysis population.

7.3.2. Per-protocol population

The per-protocol (PP) population includes only those physicians in the intervention arm who completed at least 2 hours of gameplay, and all physicians in the control arm who completed study tasks. The PP analysis is a pre-specified sensitivity analysis to assess the robustness of the primary findings.

7.3.3. Implementation analysis population

The implementation analysis population includes all key decision makers and trial physician participants who completed at least one semi-structured interview. No formal statistical comparisons are made in this population.

7.4. General Analytic Principles

All quantitative analyses are conducted using Stata 17.0 (StataCorp, TX). Baseline characteristics of physician participants are summarized descriptively by arm, using means and standard deviations for continuous variables and frequencies and proportions for categorical variables. No inferential statistics are used to compare baseline characteristics between arms. Patient-level and hospital-level variables are summarized similarly. Patterns of missingness are described prior to analysis to identify variables with non-random or high (≥ 10%) missingness.

7.5. Primary Endpoint Analysis

The primary endpoint (under-triage) is analyzed using a generalized linear mixed model (GLMM) with a binomial error distribution, a logit link function, and an exchangeable correlation structure. The outcome is modeled at the patient level. The fixed effect is treatment arm (game-based training vs. usual education). Random effects are included at the physician and hospital levels to account for clustering of patients within physicians and physicians within hospitals. Random effects are specified as random intercepts only, with an assumption of no covariance between the physician-level and hospital-level random intercepts. This specification was selected as the most parsimonious structure consistent with the nested clustering of patients within physicians and physicians within hospitals, and to minimize the risk of convergence failure given the number of hospital-level clusters in the trial. A random intercepts-only specification is appropriate given that the primary quantity of interest is the fixed treatment effect, and there is no a priori expectation that the treatment effect varies across physicians or hospitals in a way that requires random slopes. If the model does not converge (due to the scarcity of cases within hospitals), we will remove the hospital-level random effect preferentially.

Model estimation is performed in Stata 17.0 using the melogit command adaptive Gauss-Hermite quadrature with the intpoints(7) option. Convergence is assessed using the following diagnostics: (1) the final convergence code, which should indicate successful convergence (code = 0 in Stata or equivalent); (2) the number of iterations required to reach convergence; and (3) the scaled gradient at convergence, which should fall below the conventional threshold of 0.001. If convergence is not achieved with the default optimizer and settings, the following pre-specified remediation steps are taken in sequence: (a) increase the number of integration points using the intpoints() option; and (b) rescale continuous covariates to have mean 0 and standard deviation 1.

The primary model will the following baseline covariates: patient-level characteristics (age, sex, age-adjusted Charlson Comorbidity Index score); physician-level characteristics (sex, years of experience, number of shifts per month; and hospital-level characteristics (resource availability, bed size, teaching status). The analysis is conducted on the ITT population. A one-sided hypothesis test is used at a significance level of 0.05.

The test statistic follows a chi-square distribution with degrees of freedom equal to the number of variance parameters removed. Variance components for both random effects, including estimated variances and corresponding intraclass correlation coefficients at the physician and hospital levels will be reported. These results are interpreted as evidence of the degree of clustering of patients within physicians and physicians within hospitals, respectively, and provide empirical justification for the multilevel modeling approach used in this trial.

Protocol Deviation (Version 1.4): We decided to use unadjusted primary model with no baseline covariate adjustment and clustering only at the physician level to ensure transparency and to maximize chances for model stability. In sensitivity analyses we did test the effect of

clustering at the hospital level (patients nested within patients within hospitals), and then assessed variance to determine whether to retain the clustering or not.

Protocol Deviation (Version 1.5): We had pre-specified the use of a one-sided significance test of 0.05. However, during analysis, consistent with the CONSORT guidelines, we used a two-sided test, which corresponds to a one-sided test of 0.025. We disclose this deviation in the Methods of the paper.

7.6. Secondary Endpoint Analysis

7.6.1. Patient outcomes

The composite 30-day mortality and readmission outcome and functional dependence are each analyzed using the same GLMM framework as the primary endpoint, with the respective binary outcome as the dependent variable.

7.6.2. Safety outcomes

Over-triage is analyzed using the same GLMM framework, with the proportion of minimally injured patients transferred to a trauma center as the dependent variable.

Protocol Deviation (Version 1.3): We did not test functional dependence as a secondary clinical outcome.

7.7. Exploratory Analyses

These analyses are hypothesis-generating and are not adjusted for multiple comparisons. Full results of these analyses, including interaction term estimates, standard errors, and 95% confidence intervals will be reported.

7.7.1. Protocol Deviations

For the per protocol analysis, we restrict physicians in the game-based training group to those who completed at least 2 hours of game play and include all the physicians in the usual education group. We will then repeat the primary analysis, with the endpoint of under-triage.

Protocol Deviation (Version 1.4). We add a CACE analysis to test the effect of study compliance on the magnitude of the treatment effect. We reclassify physicians in the game-based training group who never access the game and physicians in the usual education group who access the game as non-compliers, and then estimate the effect of the intervention among physicians who comply with their assigned treatment using an instrumental variable approach, with treatment assignment as the instrument.

7.7.2. Heterogeneity of treatment effect

Heterogeneity of treatment effect is evaluated through a series of pre-specified subgroup analyses. Each analysis adds an interaction term between the treatment arm and the moderator of interest to the primary GLMM. The moderators examined are sex (male vs. female), race/ethnicity (White vs. non-White), pre-enrollment attitudes toward game-based learning (play games for fun versus not).

7.7.3. Mediation of the game's effect on under-triage

We test 4 mediators of the treatment effect: intervention dose (≥2 hours versus <2 hours), timing of game use–treatment of patient (≤30 days versus 39-90 days versus >90 days versus none), and fidelity of intervention receipt (perceptual sensitivity and decisional threshold). To evaluate exposure–effect relationships, we model the probability of under-triage and included an interaction term between the mediator and patient injury severity. To evaluate fidelity of intervention receipt relationships, we conduct a causal mediation analysis to assess whether the effect of the intervention on under-triage is mediated through perceptual sensitivity (i.e., diagnostic ability – targeted by the game) or by decisional threshold (i.e., willingness to transfer). The total effect of the intervention is decomposed into natural direct and natural indirect effects representing pathways independent of and operating through the mediator.

Protocol Deviation (Version 1.5): We deferred the mediation analysis testing the influence of fidelity of intervention receipt because we could not obtain the correct statistical package to run the analysis on the CMS Virtual Desktop. We also reconceptualized dose as the number of intervention periods during which game play occurred for any amount of time (out of 4 total). Dose therefore represented actual treatment and not assignment.

7.7.4. Mediation of the game's effect on clinical outcomes

We test whether the game exerts its effect on clinical outcomes through the pathway of under-triage, using a causal mediation analysis as above. Again, we decompose the effect of the game on clinical outcomes into direct and indirect effects representing pathways independent of and operating through the mediator.

Protocol Deviation (Version 1.5): We eliminated this mediation analysis testing once we recognized that the game did not influence clinical outcomes.

7.8. Implementation Analyses

Qualitative data from semi-structured interviews are analyzed using constant comparative methods and thematic content analysis, guided by a hybrid deductive-inductive approach informed by the Consolidated Framework for Implementation Research (CFIR). Two coders independently code transcripts using a pre-specified codebook, with discrepancies resolved by consensus. Interrater reliability is assessed using the kappa statistic, with a target of 0.80 or higher. Findings are validated through member checking with additional key decision makers.

7.9. Handling of Missing Data

Patterns of missingness in physician-level, patient-level, and hospital-level variables are described prior to analysis using a systematic audit of all key variables. Missingness proportions are reported overall and separately by treatment arm to identify any evidence of differential missingness. Variables with non-random or high (≥ 10%) missingness are identified and handled on a case-by-case basis.

The likely missing data mechanism is assessed based on the overall level of missingness, its distribution across arms, and its concentration by variable type. Where missingness is low, non-differential between arms, and concentrated in self-reported variables rather than outcome or claims-derived variables, a missing at random (MAR) or missing completely at random (MCAR) mechanism is considered most plausible. A missing not at random (MNAR) mechanism cannot be formally excluded and is acknowledged as a limitation where relevant.

The ITT approach mitigates the impact of missing outcome data by including all randomized physicians who filed Medicare claims, regardless of intervention adherence. Complete case analysis is used for the primary effectiveness analysis. If missingness in key covariates is substantial (≥ 10%), sensitivity analyses using multiple imputation by chained equations (MICE) will be considered as a pre-specified contingency.

7.10. Type I Error Rate

A one-sided significance level of 0.05 is used for the primary hypothesis test to boost statistical efficiency and because we prioritize identification of a positive effect of the intervention. A negative effect would produce the same outcome as a null: the decision to pause further efforts to disseminate the intervention.

Protocol deviation (Version 1.4): After database lock and prior to unmasking, the study statistician determined that a two-sided test at $\alpha = 0.05$ was more appropriate given the standard implementation of the mixed effects logistic regression framework used and CONSORT reporting requirements. This is equivalent to a one-sided threshold of $\alpha = 0.025$ and represents a more conservative standard. This deviation is disclosed in the manuscript Limitations.

8. ETHICAL CONSIDERATIONS

8.1. Regulatory and Ethical Oversight

This trial is conducted in accordance with the International Council on Harmonization Good Clinical Practice (ICH GCP) guidelines, applicable US federal regulations (45 CFR Part 46, 21 CFR Parts 50 and 56), and NIH requirements for the conduct of clinical trials. The protocol, informed consent forms, and all recruitment materials have been reviewed and approved by the University of Pittsburgh Human Research Protection Office (HRPO). Any amendments to the protocol require HRPO review and approval prior to implementation.

8.2. Informed Consent

Digital informed consent is obtained from all physician participants and key decision makers prior to enrollment. Consent forms emphasize that participation is voluntary, that withdrawal is permitted at any time without penalty, and that study participation is not a condition of employment with the participating staffing organizations. Patient participants treated by trial physicians are not consented individually; a waiver of informed consent has been granted by the HRPO on the grounds that the research involves no more than minimal risk, could not practicably be conducted without the waiver, and will not adversely affect patient rights or welfare. A waiver of HIPAA authorization has additionally been granted for the use of Medicare claims data.

8.3. Data Safety and Monitoring Board

An independent Data and Safety Monitoring Board (DSMB) composed of experts in trauma surgery, statistics, and clinical trial conduct oversees the safety and progress of the trial. The DSMB meets at least every nine months, reviews safety and efficacy data from both arms, and provides recommendations to the NIA. DSMB members are free of conflicts of interest and independent from the study team. The DSMB operates under an approved charter established at its first organizational meeting.

8.4. Adverse Events

An adverse event (AE) is defined as any untoward or unfavorable occurrence temporally associated with study participation, whether or not considered related to the intervention. Expected AEs in this trial are limited to reports of stress experienced by physician participants, given the voluntary nature of participation and the non-clinical nature of the intervention. A serious adverse event (SAE) is defined as any AE resulting in death, life-threatening condition, hospitalization, persistent disability, or congenital anomaly. All AEs are documented using the NIA Severe Adverse Event Form and reported to the IRB within 10 days; SAEs are reported to the NIH and IRB within 48 hours of the investigator becoming aware of the event. A summary of all AEs is provided to the NIA Program Officer and the DSMB at each scheduled meeting. In the event of an AE or SAE, all trial participants are notified by email within 7 days.

8.5. Confidentiality and Data Protection

All participants are assigned unique study identifiers at enrollment. Identifying information is stored separately from research data on secure University of Pittsburgh servers and is accessible only to authorized study team members. Medicare claims data are stored on the CMS Virtual Desktop and accessed via VPN with two-factor authentication. Data are not permitted to leave the CMS servers. Upon completion of the study, de-identified primary trial data and code for cleaning the Medicare files will be transmitted to the Inter-university Consortium for Political and Social Research (ICPSR) for archiving and future use by other researchers, consistent with NIH data sharing policies.

8.6. Conflict of Interest

No members of the investigative team have a financial interest related to this research. Conflicts of interest for all study personnel are disclosed and managed in accordance with University of Pittsburgh and NIA policies. DSMB members are similarly required to disclose and manage any actual or perceived conflicts of interest prior to their appointment.

9. FUNDING

This trial is funded by the National Institute on Aging (NIA) of the National Institutes of Health (NIH) under grant number R01 AG076499, awarded to the University of Pittsburgh. The sponsor and funder had no role in the design of the statistical analysis plan. Results will be disseminated in accordance with the NIH Public Access Policy, which requires submission of peer-reviewed manuscripts to PubMed Central upon acceptance for publication. The trial is registered at ClinicalTrials.gov (NCT06063434).

10. APPENDIX

10.1. Comparison of analytic elements across study documents.

Domain	DSMB protocol	Manuscript	Status	Rationale
Core Design Elements				
Primary outcome	Under-triage of severely injured patients (i.e., those with an ISS ≥16)	Under-triage of severely injured patients (i.e., those with an ISS ≥16)	Consistent	No change
Secondary outcomes	Over-triage; Mortality/readmission; Functional dependence*	Over-triage Mortality/readmission	***Deferred**	SNF claims not available
Follow up	01/01–12/31/2024 (V1) 365 days (V1.2)	365 days	Consistent	No change
Analysis population	Intention-to-treat	Intention-to-treat	Consistent	No change
Unit of analysis	Patient	Patient	Consistent	No change
Statistical model	Mixed-effects logistic regression	Mixed-effects logistic regression	Consistent	No Change
Clustering	Physician-level and hospital-level random effects (V 1) Physician (V 1.4)	Physician-level	Consistent	No change
Covariate adjustment	Patient, physician, hospital (V1) None (V1.4)	None	Consistent	No change
Repeated measures	Sensitivity analysis	Sensitivity analysis	Consistent	N/A
Power calculation	Detailed	Overview only	Consistent	N/A
Alpha level	0.05	0.025	**Deviation**	Discussed in the methods
Sidedness of tests	One-sided	Two-sided	**Deviation**	Discussed in Methods
Multiplicity control	Not done	Not done	Consistent	N/A
Reporting and Secondary Analyses				
Effect modification	Sex, race, attitudes to video games	Sex, race, attitudes to video	Consistent	Simplify manuscript
Mediation analysis	Dose, game use–time to treatment of patients	Dose, game use–time to treatment of patients	Consistent	No change
Mediation analysis	Fidelity of intervention receipt	Not done	Not performed	Statistical package not available
Mediation analysis	Effect of under-triage on clinical outcomes	Not done	**Not performed**	No difference in clinical outcomes

Domain	DSMB protocol	Manuscript	Status	Rationale
Sensitivity analyses	None (version 1) Multiple (version 1.7)	Multiple	**Expanded**	Increase rigor