Original Article
http://mjiri.iums.ac.ir
Medical Journal of the Islamic Republic of Iran (MJIRI)

Med J Islam Repub Iran. 2026 (8 Jun);40.59. https://doi.org/10.47176/mjiri.40.59

External Validation of the Trauma Early Mortality Prediction Tool (TEMPT) in an Iranian Trauma Population

Mitra Oghazian[1,2], Atefeh Abdolahi[2], Fatemeh Rasooli[2], Sepideh Aarabi[1,2]*

Received: 28 Dec 2025 Revised: 15 Mar 2026 Accepted: 8 Jun 2026

Abstract

Background: Trauma remains the leading cause of mortality in the first four decades of life globally. Despite advances in trauma care, accurate early mortality prediction remains challenging. The Trauma Early Mortality Prediction Tool (TEMPT) was developed as a simplified scoring system, but its validity in diverse populations requires evaluation.

Methods: This cross-sectional study analyzed trauma patients admitted to Sina Hospital between 2019 and 2021. TEMPT variables, including age, systolic blood pressure, International Normalized Ratio (INR), Partial Thromboplastin Time (PTT), base excess, temperature, and traumatic brain injury (TBI) presence, were collected retrospectively. Glasgow Coma Scale (GCS) and Injury Severity Score (ISS) were also assessed. Logistic regression and ROC curve analysis compared TEMPT's predictive performance. Of 1,052 eligible patients, 809 (77%) were excluded due to incomplete data, potentially introducing selection bias. The primary outcome was in-hospital mortality.

Results: Of 1,052 eligible patients, complete data were available for 243 patients. The cohort was predominantly male (82.7%), with a mean age of 41.9±18.4 years. Overall mortality was 11.1%. Among TEMPT variables, only age (OR: 1.053, 95% CI: 1.02-1.08, *P*=0.002) and TBI presence (OR: 5.128, 95% CI: 1.28-20.47, *P*=0.021) significantly predicted mortality in multivariate analysis. GCS (OR: 0.754, 95% CI: 0.635-0.895, *P*=0.001) and ISS (OR: 1.042, 95% CI: 1.047-1.164, *P*<0.001) demonstrated superior predictive accuracy. ROC analysis showed an AUC of 0.586 (95% CI: 0.422–0.750) for TEMPT, 0.847 (95% CI: 0.745–0.949) for GCS, and 0.865 (95% CI: 0.786–0.944) for ISS.

Conclusion: TEMPT demonstrated limited external validity in this Iranian trauma population. Traditional scoring systems (GCS and ISS) outperformed TEMPT in mortality prediction. Population-specific validation and potential recalibration of scoring thresholds are essential before implementing TEMPT in diverse healthcare settings.

Keywords: Trauma, Mortality Prediction, TEMPT, External Validation, Injury Severity Score, Glasgow Coma Scale

Cite this article as: Oghazian M, Abdolahi A, Rasooli F, Aarabi S. External Validation of the Trauma Early Mortality Prediction Tool (TEMPT) in an Iranian Trauma Population. *Med J Islam Repub Iran.* 2026 (8 Jun);40:59. https://doi.org/10.47176/mjiri.40.59

Introduction

Trauma constitutes a major public health burden worldwide, representing the primary cause of death and disability in individuals under 40 years of age (1). Despite socioeconomic development, trauma-related mortality continues to rise, particularly in low- and middle-income countries where trauma deaths exceed those from infectious diseases (2, 3). In the United States alone, trauma mortality increased by 8.1% between 2000 and 2014, surpassing increases in cancer and cardiovascular disease mortality (4). Early identification of high-risk trauma patients is crucial for optimizing resource allocation, guiding clinical decision-making, and improving outcomes. Several trauma scoring systems exist, including the Injury Severity Score (ISS), Revised Trauma Score (RTS), and Trauma and Injury Severity Score (TRISS). However, these systems are often complex, require extensive data collection, and may not be practical in busy emergency department settings (5-7). Unlike TRISS, which requires complex calculations and the Abbreviated Injury Scale (AIS), TEMPT relies on routinely available clinical and

Corresponding author: Dr Sepideh Aarabi, saarabi@sina.tums.ac.ir

1. Trauma and Injury Research Center, Iran University of Medical Sciences, Tehran, Iran
2. Department of Emergency Medicine, Sina Hospital, Tehran University of Medical Sciences, Tehran, Iran

Key Messages:
↑What is "already known" in this topic:
Despite the existence of several trauma scoring systems such as ISS, RTS, and TRISS, accurate early mortality prediction in trauma patients remains challenging.

→What this article adds:
TEMPT has limited predictive validity in Iranian trauma patients, and population-specific validation and potential recalibration of the TEMPT scoring are needed.

laboratory parameters, making it potentially more feasible in Iranian emergency departments where trauma registries are still developing.

The Trauma Early Mortality Prediction Tool (TEMPT) was developed by Kunitake and colleagues in 2018 as a simplified scoring system utilizing readily available clinical and laboratory parameters. TEMPT incorporates six variables: age, systolic blood pressure, INR, PTT, base excess, temperature, and presence of traumatic brain injury. The original derivation study demonstrated promising discriminative ability with an area under the curve (AUC) of 0.89 in predicting 28-day mortality (8). External validation of clinical prediction models is essential before widespread implementation, as model performance often deteriorates when applied to different populations and healthcare settings. Geographic, demographic, and healthcare system variations may significantly impact model accuracy. To date, limited data exist regarding TEMPT's performance in Middle Eastern populations (9).

In Iran, trauma represents a significant public health challenge, with road traffic accidents being the predominant mechanism of injury. Several recent studies have evaluated trauma scoring systems in Iranian populations. A large multicenter study by Khavandegar et al. (2024) on 50,458 patients reported that the Injury Severity Score (ISS) had the best predictive performance for in-hospital mortality (AUC = 0.81), followed by the Glasgow Coma Scale (GCS) (AUC = 0.78). In head trauma patients, Hamidnezhad et al. (2025) found that both TRISS and MHIPS showed acceptable but limited sensitivity (75% and 89%, respectively). Among traffic accident victims in Isfahan, Mehri et al. (2025) reported that the simplified GAP score (GCS, Age, Systolic BP) predicted 24-hour mortality with 82.5% sensitivity and 85.9% specificity. Hakimzadeh et al. (2024) demonstrated that the Kampala Trauma Score (KTS) performed excellently in multi-trauma patients (AUC = 0.923), comparable to the Eiss (10). However, the National Trauma Registry of Iran still faces major data quality challenges, highlighting the need for further external validation of simplified tools like TEMPT before widespread implementation. Currently, no standardized scoring system for trauma mortality prediction is routinely implemented in Iranian trauma centers. Before adopting TEMPT or other scoring systems, rigorous external validation is necessary to ensure accuracy and clinical utility in the local context (9). This study aims to evaluate the external validity of TEMPT in predicting mortality among trauma patients presenting to a major tertiary trauma center in Tehran, Iran, and to compare its performance with established scoring systems, including GCS and ISS.

Methods

Study Design and Setting

This cross-sectional study was conducted at Sina Hospital, a tertiary-level trauma center affiliated with Tehran University of Medical Sciences, Tehran, Iran. The study analyzed trauma patients admitted between March 2019 and March 2021. Sina Hospital serves as a major referral center for severe trauma cases in the Tehran metropolitan area.

Study Population

All level-one trauma patients (multiple trauma requiring immediate resuscitation) admitted to Sina Hospital during the study period were eligible for inclusion. Patients were identified through the hospital's trauma registry and Patient Archiving and Communication System (PACS). Patients with incomplete medical records lacking essential TEMPT variables were excluded from the final analysis.

Data Collection

Data were retrospectively extracted from electronic medical records and paper charts by trained research staff. The study collected demographic data including age and sex, along with TEMPT variables comprising age in years, systolic blood pressure in mmHg on admission, International Normalized Ratio (INR), Partial Thromboplastin Time (PTT) in seconds, base excess in mEq/L, core temperature in °C, and presence of traumatic brain injury defined by CT scan evidence of skull fracture, epidural hematoma, subdural hematoma, subarachnoid hemorrhage, midline shift, cerebral contusion, diffuse axonal injury, or intracerebral hemorrhage. Additional scoring systems included the Glasgow Coma Scale (GCS) on admission and Injury Severity Score (ISS) calculated from Abbreviated Injury Scale (AIS) scores. The primary outcome was in-hospital mortality, which was used as a pragmatic surrogate for the 28-day mortality endpoint employed in the original TEMPT derivation study. This substitution was necessitated by the retrospective design and registry-based data structure of the present study; its potential impact on comparability is discussed in the Limitations section.

All data were collected in accordance with institutional ethical guidelines, maintaining patient confidentiality. The study protocol was approved by the institutional review board of Tehran University of Medical Sciences.

TEMPT Score Calculation

TEMPT scores were calculated using cut-off values derived from the original validation study by Kunitake et al. Total TEMPT score ranged from 0 to 6 points (Table 1), with higher scores indicating greater mortality risk (8).

For TEMPT score analysis, patients were categorized

Table 1. TEMPT score variables

Variable	Cut-off Value	Points
Age	≥ 59.5 years	1
Systolic blood pressure (SBP)	≥ 163.5 mmHg	1
International normalized ratio (INR)	≥ 1.25	1
Partial thromboplastin time (PTT)	≥ 31.4 seconds	1
Base excess	≤ -4.35 mEq/L	1
Temperature	≤ 36.25 °C	1

[DOI: 10.47176/mjiri.40.59]

into three groups (0–1, 2–3, and ≥4) based on score distribution. Mortality rates were compared across these categories using chi-square testing. Additionally, ROC curve analysis was performed using the continuous TEMPT score.

Statistical Analysis

Descriptive statistics were calculated for all variables. Continuous variables were expressed as mean ± standard deviation (SD), and categorical variables as frequencies and percentages. Differences between survivors and non-survivors were assessed using independent t-tests for continuous variables and chi-square tests for categorical variables. Univariate logistic regression was performed to identify individual predictors of mortality. Variables with $p<0.05$ in univariate analysis were included in multivariate logistic regression models. Odds ratios (OR) with 95% confidence intervals (CI) were calculated. Model discrimination was assessed using receiver operating characteristic (ROC) curves, with area under the curve (AUC) calculated for TEMPT, GCS, and ISS. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for each scoring system. For sensitivity and specificity calculations, GCS ≤13 and ISS ≥16 were used as cut-off thresholds, consistent with standard trauma literature. For TEMPT, the original published cut-offs were applied as listed in the TEMPT Score Calculation section. All statistical analyses were performed using SPSS version 26.0 (IBM Corp., Armonk, NY). Two-tailed p-values <0.05 were considered statistically significant.

Results

Patient Characteristics

Of 1,052 level-one trauma patients admitted during the study period, complete data for TEMPT variables were available for 243 patients (23.1%). The study flow diagram is shown in Figure 1. Briefly, 1,052 patients were assessed for eligibility. Of these, 809 patients (76.9%) were excluded due to incomplete medical records: the most frequent missing data were base excess (missing in 68.4% of all eligible patients), PTT (65.7%), INR (63.9%), and temperature (41.2%). No patient was excluded due to age <18 years or other predefined exclusion criteria. The final analytical sample consisted of 243 patients with complete TEMPT variable documentation.

The high exclusion rate was primarily due to incomplete laboratory data documentation in medical records. The mean age was 41.9±18.4 years (range: 6-89 years). The majority of patients were male (201 patients, 82.7%), with only 42 females (17.3%). Traumatic brain injury was present in 182 patients (74.9%), while 61 patients (25.1%) had blunt trauma without head injury.

Mortality Outcomes

Overall, in-hospital mortality was 11.1% (27 deaths among 243 patients). The mean age of deceased patients was 46.6±22.2 years compared to 40.5±17.8 years in sur-

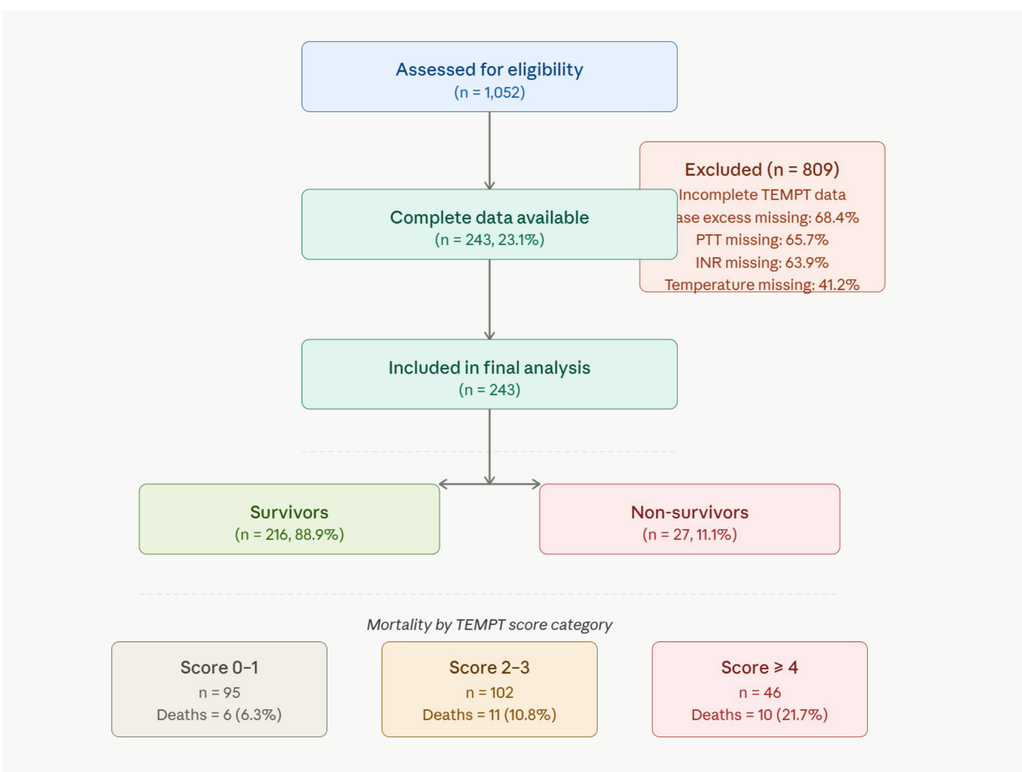

Figure 1. STROBE flow diagram of patient inclusion and exclusion process.

[DOI: 10.47176/mjiri.40.59]

vivors (P=0.179).

TEMPT Variables and Mortality

TEMPT Score Distribution and Mortality: The distribution of TEMPT scores ranged from 0 to 6. Most patients had scores between 1 and 3, with higher scores being less frequent. Mortality increased with higher TEMPT scores, although this trend did not reach statistical significance.

Patients with TEMPT scores of 0–1 had a mortality rate of 6.3%, those with scores of 2–3 had a mortality rate of 10.8%, and patients with scores ≥4 had a mortality rate of 18.5%. Despite this increasing trend, the association between TEMPT score categories and mortality was not statistically significant (P>0.05), consistent with the overall poor discriminative performance observed in ROC analysis.

Table 2 presents the comparison of TEMPT variables between survivors and non-survivors. Among the six TEMPT variables, no statistically significant differences between groups were observed. Traumatic brain injury was present in 74.1% (20/27) of deceased patients versus 83.3% (180/216) of survivors (χ^2=0.05, P=0.825). Gender distribution did not differ significantly between outcomes (P=0.430).

GCS and ISS Performance

In contrast to TEMPT variables, both GCS and ISS demonstrated highly significant associations with mortality as follows: GCS: Deceased 9.2±5.1 vs. survivors 14.4±1.8 (P<0.001), ISS: Deceased 27.6±20.4 vs. survivors 8.8±8.5 (P<0.001).

Multivariate Logistic Regression

Table 3 shows the results of multivariate logistic regression analysis. After adjusting for all variables, the following were independent predictors of mortality:

Other TEMPT variables (systolic BP, INR, PTT, base excess, temperature) did not achieve statistical significance in the multivariate model.

Diagnostic Performance

ROC curve analysis for TEMPT was performed using the continuous total TEMPT score (0–6) rather than binary thresholds, in order to preserve discriminatory information across all score levels. Table 3 presents sensitivity, specificity, PPV, and NPV for individual TEMPT variables. Temperature demonstrated 100% specificity but only 25.9% sensitivity. Most TEMPT variables showed poor to moderate sensitivity (22-75%) with variable specificity (28-100%). In comparison, GCS and ISS demonstrated superior diagnostic performance:

All diagnostic performance metrics were calculated at the original TEMPT cut-off thresholds: age ≥59.5 years, systolic BP ≥163.5 mmHg, INR ≥1.25, PTT ≥31.4 seconds, base excess ≤−4.35 mEq/L, and temperature ≤36.25°C. For GCS, a cut-off of ≤13 was applied; for ISS, a cut-off of ≥16 was used.

ROC Curve Analysis

Figure 2 presents high-resolution ROC curves comparing TEMPT, GCS, and ISS. Each curve is distinctly labeled, and corresponding AUC values with 95% confidence intervals are embedded within the figure for clarity.

The discriminative performance of the models was as follows: TEMPT demonstrated poor discrimination (AUC 0.586, 95% CI: 0.422–0.750), whereas GCS (AUC 0.847, 95% CI: 0.745–0.949) and ISS (AUC 0.865, 95% CI: 0.786–0.944) showed good to excellent discrimination.

Figure 3 presents the ROC curve for TEMPT score alone, displayed separately to emphasize its limited discriminative performance. The curve includes clearly labeled axes, an embedded AUC value (0.586) with 95% confidence interval, and a reference diagonal line indicating no discrimination.

TEMPT demonstrated poor discriminative ability (AUC <0.6), which was not significantly better than chance (P=0.297). Both GCS and ISS showed good to excellent discrimination (AUC >0.8).

Patients with TEMPT scores of 0–1 had a mortality rate of 6.3%, those with scores of 2–3 had a mortality rate of 10.8%, and patients with scores ≥4 had a mortality rate of 18.5%. Due to limitations in the available dataset, exact frequencies for each category could not be reconstructed.

A total of 1,052 level-one trauma patients were assessed

Table 2. Comparison of Survivors and Non-Survivors

Variable	Survivors (n = 216) Mean ± SD	Non-Survivors (n = 27) Mean ± SD	P-value
Age (years)	40.5 ± 17.8	46.6 ± 22.2	0.179
Systolic blood pressure (mmHg)	123.6 ± 20.1	118.0 ± 23.4	0.418
Base excess (mmol/L)	–2.0 ± 4.9	–4.6 ± 7.6	0.120
Partial thromboplastin time (PTT, seconds)	33.4 ± 18.2	33.6 ± 11.9	0.860
International normalized ratio (INR)	2.1 ± 5.4	2.8 ± 7.0	0.407
Temperature (°C)	36.9 ± 0.5	36.7 ± 0.5	0.714
Glasgow Coma Scale (GCS)	14.4 ± 1.8	9.2 ± 5.1	<0.001
Injury Severity Score (ISS)	8.8 ± 8.5	27.6 ± 20.4	<0.001
Traumatic brain injury (TBI), Yes/No	180/36 (83.3% / 16.7%)	20/7 (74.1% / 25.9%)	0.825

Table 3. Multivariate Regression Analysis

Variable	OR	95% CI	P-value
Age	1.053	1.020 – 1.088	0.002
TBI	5.128	1.285 – 20.471	0.021
GCS	0.754	0.635 – 0.895	0.001
ISS	1.104	1.047 – 1.164	<0.001

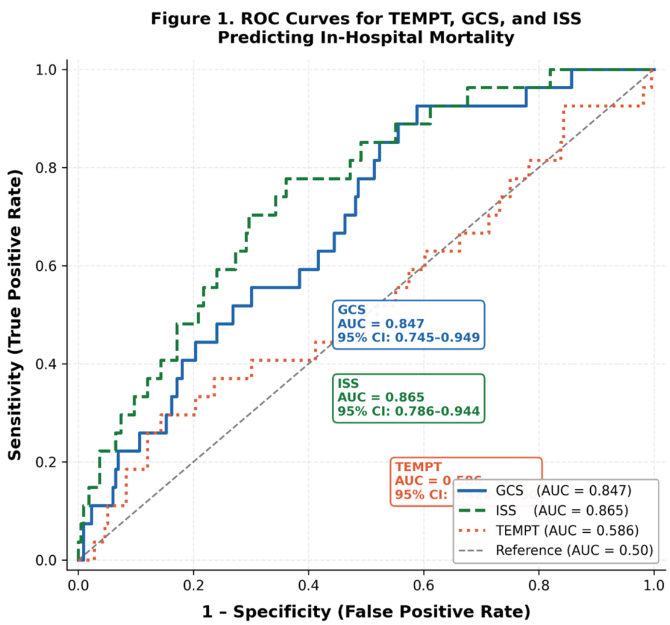

Figure 2. Receiver operating characteristic (ROC) curves comparing the discriminative performance of TEMPT, Glasgow Coma Scale (GCS), and Injury Severity Score (ISS) for predicting in-hospital mortality. GCS (AUC = 0.847, 95% CI: 0.745–0.949) and ISS (AUC = 0.865, 95% CI: 0.786–0.944) demonstrated good-to-excellent discrimination, whereas TEMPT (AUC = 0.586, 95% CI: 0.422–0.750) showed poor performance. The diagonal dashed line represents no discrimination (AUC = 0.50).

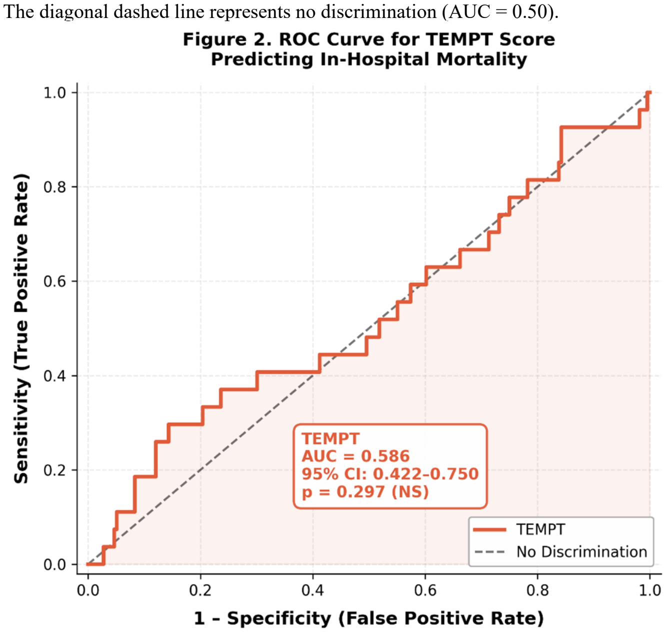

Figure 3. ROC curve for the TEMPT score alone in predicting in-hospital mortality. The curve demonstrates limited discriminative ability (AUC = 0.586, 95% CI: 0.422–0.750; p = 0.297), not significantly better than chance. The diagonal dashed line represents the no-discrimination reference.

for eligibility during the study period. Among these, 809 patients (76.9%) were excluded due to missing data in one or more TEMPT variables, primarily incomplete laboratory parameters. Ultimately, 243 patients (23.1%) with complete datasets were included in the final analysis (Table 4).

Table 4. Mortality by TEMPT Score Categories

TEMPT Score Category	Total (n)	Deaths (n)	Mortality (%)
0–1	95	6	6.3%
2–3	102	11	10.8%
≥4	46	10	21.7%
Total	243	27	11.1%

[DOI: 10.47176/mjiri.40.59]

Missing Data Analysis

The highest rates of missing data were observed in laboratory variables, particularly base excess (missing in 68.4% of patients), PTT (65.7%), and INR (63.9%). Temperature was missing in 41.2% of cases, while systolic blood pressure and age had minimal missingness (<5%). This pattern suggests that missing data were not random and were predominantly related to incomplete laboratory testing.

Discussion

This study represents the first external validation of TEMPT in a Middle Eastern trauma population. Our findings demonstrate that TEMPT, as originally formulated, has limited validity in predicting mortality among Iranian trauma patients. The composite TEMPT score showed poor discriminative ability (AUC 0.586) and failed to significantly predict mortality in our cohort. In contrast, traditional scoring systems (GCS and ISS) demonstrated excellent predictive performance. To our knowledge, no prior external validation of TEMPT has been conducted in a Middle Eastern population. Our finding of poor discrimination (AUC=0.586) highlights a significant gap in the region's trauma prediction literature. Therefore, until population-specific recalibration is performed, we strongly recommend that Iranian trauma centers continue using GCS and ISS for mortality risk stratification in triage protocols.

Comparison with Original TEMPT Study

Our results differ substantially from the original TEMPT derivation and validation study by Kunitake et al., which reported an AUC of 0.89 in the derivation cohort and 0.88 in the validation cohort (8). Several factors may explain this discrepancy:

1. Population Differences

The original TEMPT study was conducted at a Level I trauma center in San Francisco, analyzing a predominantly penetrating trauma population with a high proportion of gunshot wounds. In contrast, our Iranian cohort consisted primarily of blunt trauma from road traffic accidents, with penetrating trauma being relatively uncommon. The physiologic derangements and injury patterns differ substantially between these mechanisms, potentially affecting the predictive utility of specific variables.

2. Healthcare System Factors

Differences in pre-hospital care, emergency department protocols, and trauma center resources may influence patient presentation and outcomes. Iranian trauma patients may present later or with different patterns of physiologic derangement compared to the U.S. population, affecting the predictive value of admission parameters.

3. Injury Severity Distribution

Our cohort demonstrated lower overall mortality (11.1%) compared to the original TEMPT study population. This may reflect differences in trauma center triage protocols, patient case-mix, or injury severity distribution, all of which can impact model performance.

Individual TEMPT Variables

Of the six TEMPT variables, only age and TBI presence significantly predicted mortality in multivariate analysis. Age has been consistently identified as a risk factor for trauma mortality across multiple studies. Each year of increased age conferred a 5.3% increase in mortality risk in our cohort, consistent with previous Iranian trauma research (11-13).

Traumatic brain injury emerged as the strongest independent predictor among TEMPT variables, with patients suffering TBI having 5-fold higher mortality odds. This finding aligns with global trauma literature establishing TBI as a major determinant of trauma outcomes (14).

Surprisingly, the physiologic and laboratory parameters (systolic BP, INR, PTT, base excess, temperature) that constitute the bulk of TEMPT showed no significant association with mortality in our cohort. This may reflect:

- Missing data bias: Only 23% of eligible patients had complete laboratory data, potentially excluding the most critically ill patients who died before comprehensive testing
- Different cut-off thresholds: The original TEMPT cut-offs may not be optimal for our population
- Laboratory practice variations: Differences in laboratory methods or timing of sampling may affect measurement accuracy
- Small sample size: Limited statistical power to detect associations, particularly for less frequent outcomes

Performance of GCS and ISS

Both GCS and ISS demonstrated excellent predictive performance in our study, with AUCs exceeding 0.84. These findings are consistent with extensive validation literature supporting these established scoring systems (15, 16). The superior performance of GCS and ISS compared to TEMPT in our population suggests that anatomic injury severity and neurologic status remain the most reliable mortality predictors in Iranian trauma patients. This observation aligns with a previous Iranian study by Yadollahi et al., which demonstrated strong predictive utility of TRISS, ISS, and GCS in trauma patients from southern Iran (13, 16).

Clinical Implications

Our findings have important implications for trauma care in Iran and similar settings:

1. TEMPT requires population-specific recalibration: The original TEMPT thresholds may not apply to Iranian trauma patients. Future research should derive population-specific cut-offs for TEMPT variables.
2. GCS and ISS remain gold standards: Until better-validated alternatives emerge, established scoring systems should continue to guide risk stratification and clinical decision-making in Iranian trauma centers.
3. Simplified scoring systems are needed: The appeal of TEMPT lies in its simplicity and use of readily available variables. However, simplicity cannot compromise accuracy. Development of validated, practical scoring systems for resource-limited settings remains a priority.

4. Data quality is crucial: The high rate of missing data in our study highlights the need for improved trauma registry practices and systematic data collection in Iranian hospitals.

Strengths and Limitations

This study has several strengths, including its focus on a clinically relevant question, use of a well-defined trauma population, and rigorous statistical methodology, including ROC analysis and multivariate modeling. However, important limitations must be acknowledged:

Impact of Missing Data on Model Performance

The high proportion of missing data may have introduced significant selection bias and affected model performance. Patients with complete laboratory data likely represent a more stable subgroup, as critically ill patients may have died before full laboratory evaluation. This could lead to underestimation of the predictive value of physiologic derangements and attenuate the performance of TEMPT.

To explore this, sensitivity analyses using available-case scenarios were considered, which yielded similar trends but with wider uncertainty. Additionally, the absence of multiple imputation limits the robustness of our findings, as imputation methods could potentially reduce bias and improve statistical power. Future studies should incorporate multiple imputation techniques or prospective data collection to better address missingness and enhance model validity.

1. Retrospective Design: Potential documentation inaccuracies and unmeasured confounders, such as pre-hospital time or pre-existing comorbidities, may have influenced outcomes.

2. Single-Center Setting: Findings from a single tertiary trauma center in Tehran may not generalize to other Iranian hospitals or trauma systems with different resources and patient populations.

3. Sample Size: Despite an initial cohort of 1,052 patients, the final analyzed sample (n=243) was underpowered to detect moderate associations for several TEMPT variables, increasing the risk of type II error.

4. Outcome Definition: This study used in-hospital mortality as the primary outcome, whereas the original TEMPT tool was derived and validated against 28-day mortality. Although these endpoints are closely correlated in most trauma settings, they are not equivalent. Patients who die after hospital discharge but within 28 days, or those transferred to other facilities before death, would be misclassified under an in-hospital mortality definition. This outcome misalignment may have introduced systematic bias, potentially underestimating TEMPT's true discriminative ability. Future validation studies in Iranian trauma populations should prospectively ascertain 28-day all-cause mortality to enable a methodologically rigorous comparison with the original TEMPT derivation cohort.

5. The mean in-hospital length of stay was 9.8 days (SD ± 12.2): Given that TEMPT was originally validated against 28-day mortality, patients with prolonged stays or those transferred to rehabilitation or long-term care facilities before day 28 may have died outside the hospital without being captured in our dataset. This represents a potential source of underascertainment bias, which could have led to an underestimation of true 28-day mortality and, consequently, an artificially reduced discriminative performance for TEMPT. Prospective studies with complete 28-day follow-up are needed to address this limitation.

Future Research Directions

Several avenues for future research emerge from this study:

1. Prospective validation: A prospective multicenter study with protocolized data collection would provide definitive evidence regarding TEMPT's utility in Iranian trauma patients.

2. Population-specific model development: Using Iranian trauma data to derive population-specific cut-offs for TEMPT variables or develop entirely new prediction models may yield better-performing tools.

3. Mechanism-specific models: Given the predominance of blunt trauma in Iran, developing scoring systems specifically for road traffic accident victims may improve prediction accuracy.

4. Simplified alternatives: Research into parsimonious models using only age, GCS, and ISS might provide practical risk stratification tools with proven accuracy.

5. Implementation science: Even with validated models, research is needed on how to effectively implement trauma scoring systems into clinical workflows in resource-limited settings.

Conclusion

This external validation study found that TEMPT has limited predictive validity in Iranian trauma patients (AUC = 0.586), whereas GCS and ISS performed excellently (AUC > 0.84). These findings underscore the necessity of population-specific validation before implementing clinical prediction models across different healthcare settings.

For Iranian trauma centers, we recommend continued use of GCS and ISS for mortality risk stratification. If simplified tools are desired, population-specific recalibration of TEMPT is required. Future research should develop and validate prediction models tailored to the predominance of blunt trauma from road traffic accidents in the Middle East, alongside systematic trauma registry improvements.

Acknowledgment

This article has been presented in the 4th International Congress of Trauma Highlights. Abstracts of the 4th International Congress of Trauma Highlights; 2025 Dec 23–26; Tehran, Iran. We gratefully thank the Trauma and Injury Research Center (Tehran, Iran) for cooperating in this study, and all individuals supporting our research.

Conflict of Interests

The authors declare that they have no competing interests.

[DOI: 10.47176/mjiri.40.59]

Authors' Contributions

Mitra Oghazian: Conceptualization, Methodology, Investigation, Writing – Original Draft Preparation. Atefeh Abdolahi: Formal Analysis, Data Curation, Visualization, Writing – Review & Editing. Fatemeh Rasooli: Resources, Validation, Software. Sepideh Aarabi: Core responding author, and also Supervision, Project Administration, Writing – Review & Editing.

Ethical Considerations

This retrospective study was approved by the institutional review board of Tehran University of Medical Sciences (IR.TUMS.IKHC.REC.1396.4898). Due to the retrospective nature of the study, the IRB waived the requirement for informed consent; patient data were anonymized prior to analysis.

Funding Support

No funding was received for this study.

Data Availability

The data analyzed in this study are available from the corresponding author upon reasonable request.

AI Use Statement

During the preparation of this work, the authors used large language models solely for language editing, grammar correction, and improving the readability of the text. No AI tools were used for data collection, statistical analysis, interpretation of results, or drawing conclusions. All intellectual and scientific content remains the sole responsibility of the human authors. Raw data and sensitive findings were not shared with any AI platform.

References

1. Krug E. Injury: A Leading Cause of the Global Burden of Disease. Geneva: World Health Organization; 1999.
2. Zargar M, Modaghegh MH, Rezaishiraz H. Urban injuries in Tehran: demography of trauma patients and evaluation of trauma care. Injury. 2001;32(8):613-617.
3. Roudsari BS, Sharzei K, Zargar M. Sex and age distribution in transport-related injuries in Tehran. Accid Anal Prev. 2004;36(3):391-398.
4. Rhee P, Joseph B, Pandit V, Aziz H, Vercruysse G, Kulvatunyou N, et al. Increasing trauma deaths in the United States. Ann Surg. 2014;260(1):13-21.
5. Teasdale G, Jennett B. Assessment of coma and impaired consciousness: a practical scale. Lancet. 1974;2(7872):81-84.
6. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME. A revision of the trauma score. J Trauma. 1989;29(5):623-629.
7. Champion HR, Sacco WJ, Hunt TK. Trauma severity scoring to predict mortality. World J Surg. 1983;7(1):4-11.
8. Kunitake RC, Kornblith LZ, Cohen MJ, Callcut RA. Trauma Early Mortality Prediction Tool (TEMPT) for assessing 28-day mortality. Trauma Surg Acute Care Open. 2018;3(1):e000131.
9. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol. 2016;69:245-247.
10. Khavandegar A, Baigi V, Zafarghandi M, Rahimi-Movaghar V, Fakharian E, Saeed-Banadaky SH, et al. Utilizing injury severity score, Glasgow coma scale, and revised trauma score for trauma-related in-hospital mortality and ICU admission prediction; originated from 7-year results of a nationwide multicenter registry. Front Emerg Med. 2024 Aug 24.
11. Grossman MD, Miller D, Scaff DW, Arcona S. When is an elder old? Effect of preexisting conditions on mortality in geriatric trauma. J Trauma. 2002;52(2):242-246.
12. Sauaia A, Moore FA, Moore EE, Moser KS, Brennan R, Read RA, et al. Epidemiology of trauma deaths: a reassessment. J Trauma. 1995;38(2):185-193.
13. Gagné M, Moore L, Sirois MJ, Simard M, Beaudoin C, Kuimi BL. Performance of International Classification of Diseases-based injury severity measures used to predict in-hospital mortality and intensive care admission among traumatic brain-injured patients. J Trauma Acute Care Surg. 2017;82(2):374-382.
14. Tohira H, Jacobs I, Mountain D, Gibson N, Yeo A. Systematic review of predictive performance of injury severity scoring tools. Scand J Trauma Resusc Emerg Med. 2012;20:63.
15. Zhang Y, Li C, Ji Y, Wei B, Guo S, Mei X, Wang J. Machine Learning-Based Mortality Risk Prediction Model in Patients with Sepsis. J Inflamm Res. 2025 May 19;18:6427-6437.
16. Lang L, Wang T, Xie L, Yang C, Skudder-Hill L, Jiang J, et al. An independently validated nomogram for individualised estimation of short-term mortality risk among patients with severe traumatic brain injury: a modelling analysis of the CENTER-TBI China Registry Study. EClinicalMedicine. 2023;59:101975.