Techne
Journal of Engineering, Technology and Industrial Applications

Healthcare Emergency Triage: Quantifying Safety, Throughput, and Equity Risk Under Uncertainty, Crowding, and Decision Latency

Sreyneang Keo

Author Affiliation:
Faculty of Electrical Engineering and Computing, Cambodia University of Technology and Science (CamTech), Phnom Penh 12000, Cambodia

***Corresponding Author**
Faculty of Electrical Engineering and Computing, CamTech, Phnom Penh 12000, Cambodia
Email: sreyneang.keo@camtech.edu.kh

ABSTRACT

This article proposes an engineering-oriented reliability framework for emergency triage that models end-to-end uncertainty propagation from initial symptoms and vital signs through decision rules, reassessment intervals, queueing dynamics, and escalation policies into distributional outcomes that matter for safety and service performance, including probability of under-triage, time-to-provider exceedance probability, adverse clinical event risk before definitive evaluation, over-triage burden, and an operational cost and harm index that integrates patient risk with resource strain. A scenario-based quantitative study is developed for a generic high-volume emergency department serving mixed-acuity adult patients, comparing four triage architectures: baseline scale-based triage with static rules and discretionary reassessment, expanded screening with more data but without governance, calibrated risk scoring with static thresholds, and a governance-optimized two-tier system that combines calibrated risk scoring, explicit uncertainty handling, capacity-aware dynamic thresholds, staged reassessment with trigger-based escalation, and safety-bounded operational controls. Results show that adding data or algorithms without governance can increase volatility and over-triage under crowding, that calibrated scoring improves stability but is fragile when queueing regimes shift, and that the two-tier governed architecture reduces under-triage and time-to-provider exceedances while stabilizing resource use and reducing inequitable failure patterns under documentation noise and surge conditions. Three copy-ready tables and complete prompts for data-driven figures are provided for Techne submission.

Keywords: Emergency Department, Triage Reliability, Patient Safety, Under-Triage, Over-Triage, Crowding, Decision Latency.

1. INTRODUCTION

Emergency triage is frequently presented as a categorization task, yet in practical engineering terms it is a reliability decision pipeline that must convert imperfect, early, and often ambiguous evidence into time-critical actions that allocate scarce clinical resources, and the quality of this pipeline is measured not by how neatly patients fit into categories but by whether the right patients receive timely evaluation, escalation, and treatment when delays meaningfully change clinical outcomes (Karkkainen, 2019; Vasilakis & Wood, 2020). When triage fails, the failure typically manifests as a delay rather than a wrong label in a dataset, because a patient who should have been prioritized is often still eventually seen, but the time window in which intervention would have prevented deterioration may have been missed, leading to adverse events that are clinically and ethically unacceptable and operationally costly. If triage prioritizes too aggressively, the system generates over-triage that can reduce throughput and create crowding, and crowding itself increases the probability of missed deterioration, medication errors, and left-without-being-seen events, which implies that triage reliability cannot be evaluated in isolation from the queueing dynamics and operational regimes that triage decisions help shape (Akter et al., 2023; Sultana, 2022; Tsao et al., 2016).

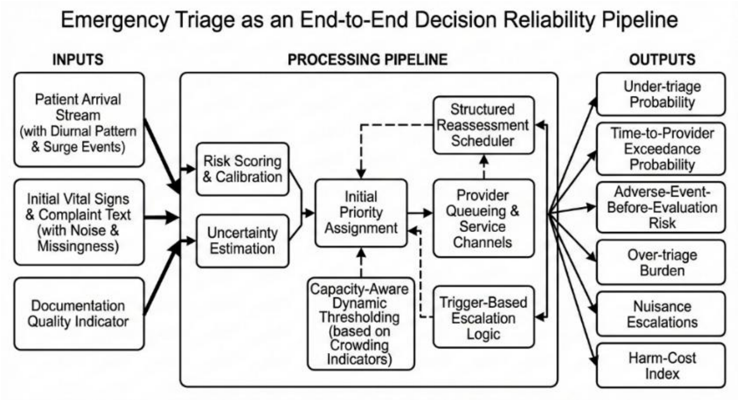

Figure 1. Triage as an end-to-end reliability decision pipeline

Three features of the emergency department environment make reliability analysis necessary rather than optional. The first is uncertainty in early evidence, because initial triage relies on limited information: a brief complaint description, a small set of vital signs that may be noisy or transient, and a rapid assessment that is constrained by time pressure and variable patient communication. Many high-risk presentations, such as early sepsis, occult bleeding, atypical myocardial ischemia, or silent hypoxia, can resemble benign conditions in the first minutes, and the statistical overlap between serious and non-serious cases is large, meaning that perfect classification is impossible and reliability must be managed through staged evidence collection and reassessment rather than through one-shot judgement alone (Druetto et al., 2024; Griffin et al., 2016; Sahlaoui et al., 2023).

The second feature is decision latency, because triage is not only about initial assignment but about whether the system updates decisions when new evidence appears, and in crowded environments, reassessment can be delayed or inconsistent, allowing deterioration to proceed unnoticed. The third feature is system coupling through crowding and capacity constraints, because triage decisions influence queue composition, which influences time-to-provider for all patients, which in turn influences future triage performance by increasing stress and reducing observation quality; this coupling creates feedback loops where small policy changes can have nonlinear effects on safety and throughput (Bhuiyan, 2025; Lameesa et al., 2024; O'Hara et al., 2014).

Traditional triage scales and clinician judgement have been essential in structuring emergency operations, but in many settings the reliability challenge is evolving because patient volumes increase, case-mix complexity grows, and expectations for rapid evaluation and equitable outcomes intensify. Operational crowding has become a defining characteristic of many emergency departments, and crowding shifts the system into regimes where service time distributions and waiting time distributions become heavy-tailed, meaning that average wait time is less informative than the probability of extreme delays, and extreme delays are precisely the events most associated with harm for time-sensitive conditions (Hao et al., 2020; Jaghni, 2025). A triage method that performs adequately in calm conditions can become unreliable in stress conditions if it lacks explicit governance for reassessment, escalation, and capacity-aware prioritization. The increasing use of digital systems, risk scoring, and algorithmic triage introduces new failure modes, because models trained on historical data may not remain calibrated under drift, because data quality can degrade during peaks, and because overreliance on algorithmic outputs can reduce the vigilance that traditionally compensates for instrument noise and documentation gaps (Ortíz-Barrios & Alfaro-Saíz, 2020; Youn et al., 2022).

Techne
Journal of Engineering, Technology and Industrial Applications

From an engineering and applied technology standpoint, the most useful question is therefore not whether a particular triage tool is "better" in a static sense, but how a triage architecture manages uncertainty and operational constraints end-to-end, and how it performs in distributional terms under both normal and surge conditions. A reliability approach treats triage as a decision system with staged actions, where the system must allocate attention efficiently while ensuring that the probability of catastrophic under-triage remains below an acceptable bound, and where operational efficiency is treated as a constraint rather than the only objective (Fan et al., 2025; Machireddy, 2023). This approach aligns with reliability engineering in other domains, such as process control and grid protection, where rare high-consequence failures dominate risk and where system design focuses on detection, escalation, and robust governance rather than on average-case performance.

This article develops an engineering-oriented reliability framework for emergency triage and demonstrates it through a scenario-based quantitative study that compares four architectures under realistic uncertainty and capacity constraints. Architecture A represents baseline scale-based triage with static rules and discretionary reassessment, reflecting common practice where an initial triage category is assigned and re-triage occurs informally when staff are available or when a patient actively deteriorates. Architecture B represents expanded screening where additional measurements and prompts are added, such as more frequent vital sign capture and broader symptom checklists, but without drift-aware governance or reassessment discipline, reflecting the assumption that more data automatically improves reliability. Architecture C represents calibrated risk scoring, where a probabilistic risk score is computed using observed features and is mapped to priority categories with cost-sensitive thresholds, but thresholds remain static and reassessment is not explicitly governed. Architecture D represents a governance-optimized two-tier system, where triage is treated as staged decision-making: the initial stage assigns a risk and uncertainty estimate, medium-risk or high-uncertainty cases are routed to structured reassessment and low-friction evidence collection pathways, high-risk cases trigger immediate escalation, and the entire policy adapts to operational regime through capacity-aware thresholding that preserves safety bounds while controlling over-triage and queue saturation.

The article addresses three research questions that are directly relevant to emergency department engineering and to Techne's applied scope. First, how do measurement noise, documentation incompleteness, and clinical heterogeneity propagate into under-triage probability and time-to-provider exceedance risk, and which factors dominate reliability tails under crowding? Second, how do different triage architectures trade off safety (under-triage and adverse event risk) against throughput (time-to-provider and length-of-stay proxies) and friction (over-triage and resource strain), especially under surge conditions where queueing dynamics shift regimes? Third, what governance mechanisms, such as uncertainty-aware routing, dynamic thresholds, and reassessment protocols, provide robust reliability improvements without creating unsustainable operational burden or inequitable patterns across patient subgroups?

The remainder of the article is organized to facilitate practical implementation. The literature review synthesizes reliability-relevant concepts from triage practice and decision systems, emphasizing uncertainty, drift, and operational coupling rather than purely descriptive perspectives. The methodology section defines the scenario design, uncertainty and drift models, architecture definitions, and evaluation metrics, with explicit focus on distributional exceedance probabilities and harm-oriented indices. The results and discussion present comparative outcomes in three copy-ready tables and interpret the mechanisms behind the differences, highlighting why governance changes outcomes even when the underlying scoring is similar. The conclusion consolidates engineering implications and provides complete prompts for data-driven figures that can be generated from emergency department operational datasets, enabling replication and adaptation.

2. LITERATURE REVIEW

Triage as a Staged Decision Pipeline under Uncertainty

Triage decisions are made early, quickly, and with incomplete information, and the clinical state of a patient can evolve between initial assessment and definitive evaluation, which implies that triage reliability depends on the design of the entire pipeline rather than the initial category assignment alone (Hao et al., 2025).

Triage consists of multiple stages, including initial assessment, prioritization, waiting, reassessment, and escalation, and failures often occur when the system assumes that initial triage is sufficient and does not enforce structured reassessment intervals or trigger-based escalation rules. This staged view is aligned with reliability engineering, where early signals are treated as uncertain and where the system design emphasizes detection and escalation pathways that convert uncertain early evidence into higher-confidence decisions over time, particularly for conditions with high time sensitivity (Adari, 2020; Joseph et al., 2022).

The overlap between serious and non-serious presentations at triage is often large, and therefore the objective is not perfect separation but controlled error rates under operational constraints. Under-triage is typically the critical error because it exposes high-risk patients to delay, while over-triage increases resource strain and can indirectly create harm by worsening crowding. Because under-triage and over-triage are coupled through crowding, a reliable system must manage both, which suggests that triage should be governed by explicit risk tolerances and capacity-aware controls rather than by static thresholds alone (Youssef et al., 2024).

The Role of Measurement Noise and Documentation Completeness

Vital signs, complaint descriptions, and early clinical cues are noisy, and their noise is not purely random because it is influenced by measurement conditions, patient behavior, and staff workload. For example, respiratory rate is commonly under-measured or rounded, oxygen saturation can be influenced by motion and perfusion, and temperature measurement can vary by device and technique, while complaint descriptions can be incomplete due to communication barriers or time pressure. Documentation quality often degrades under crowding, and missingness can become informative in a negative way because the absence of data may correlate with higher workload and therefore higher risk of missed deterioration (DeHollander et al., 2025). This introduces a reliability challenge for algorithmic triage systems because models may rely on features that are systematically missing or biased in high-stress periods, and a system that does not account for this can become less reliable precisely when it is most needed.

Reliability-oriented design therefore emphasizes robust features, plausibility checks, and uncertainty estimation that reflects data quality. Instead of treating missing data as neutral, a reliable system can treat it as uncertainty that triggers either conservative buffering, additional measurement, or structured reassessment, depending on the expected value of further evidence and on capacity constraints (Griffin et al., 2016; O'Hara et al., 2014).

Crowding, Queueing Dynamics, and Non-Linear Risk

Emergency department crowding changes the statistical structure of waiting times, because as utilization approaches capacity, queueing systems produce rapidly increasing waiting times and heavy tails, meaning that the probability of extreme delay becomes the dominant risk driver. Under these conditions, triage decisions that add marginally more over-triage can have disproportionate effects on the system by increasing queue lengths and pushing more patients into long-delay regimes, while triage decisions that reduce over-triage too aggressively can increase under-triage risk if high-risk cases are missed (Sultana, 2022; Vasilakis & Wood, 2020). Triage reliability must be evaluated under realistic crowding regimes rather than under average conditions, and policies must include safeguards that preserve safety while avoiding runaway queue growth.

Crowding also interacts with reassessment, because reassessment requires staff time and attention, and when staff are overwhelmed, reassessment intervals can increase, which increases risk for patients whose condition deteriorates while waiting. A reliable architecture must therefore allocate reassessment resources efficiently, often by focusing reassessment on uncertainty and on risk indicators rather than applying uniform reassessment, which is rarely feasible (O'Hara et al., 2014; Ortíz-Barrios & Alfaro-Saíz, 2020).

Drift, Changing Case-Mix, and Policy Fragility

Emergency departments experience drift in patient case-mix due to seasonal patterns, outbreaks, demographic shifts, and changes in access to primary care, and operational processes also change due to staffing changes, protocol updates, and new diagnostic pathways. Algorithmic triage systems trained on historical data can become miscalibrated under drift, and even non-algorithmic systems can become inconsistent as staff change or as local interpretations of triage categories evolve. Policy fragility arises when thresholds and rules are fixed while the distribution of risk changes, leading to either increased under-triage or increased over-triage, and both can manifest as reliability failures (Lameesa et al., 2024; Machireddy, 2023).

Reliability-oriented governance includes drift monitoring and controlled adaptation. Rather than frequent ad hoc changes, a system can use predefined triggers to adjust thresholds, reassessment intensity, or escalation pathways, and it can enforce safety bounds that protect against under-triage even when the system adapts for throughput.

Equity and Differential Error Risk

Triage reliability is not only an aggregate safety concern; it also includes fairness and equity, because documentation quality, communication barriers, and baseline vital sign differences can vary across patient groups, and because implicit bias can affect clinical judgement (Akter et al., 2023; Sahlaoui et al., 2023). A reliable architecture should therefore be evaluated for differential error rates across subgroups, and governance should include mechanisms to reduce inequitable failure patterns, such as structured symptom prompts that reduce reliance on subjective narratives, uncertainty routing that triggers additional evidence collection when communication is limited, and monitoring that detects subgroup-specific drift in error rates (Druetto et al., 2024; Griffin et al., 2016; Youn et al., 2022).

Although this article uses a generic scenario-based study rather than site-specific subgroup data, the framework explicitly models documentation noise and missingness that can act as a proxy for differential evidence quality, and it evaluates whether architectures remain stable under such degradation, which is a necessary step toward equitable reliability.

Gap Research

Many triage evaluations focus on predictive performance of scales or models in static datasets, yet operational reliability depends on end-to-end decision pipelines under crowding, uncertainty, and delayed escalation. There remains a need for engineering-oriented frameworks that quantify how evidence uncertainty and operational constraints propagate into exceedance probabilities for time-to-provider and into adverse event risk proxies, while also integrating reassessment governance and capacity-aware controls. This study addresses that gap by comparing architectures using distributional reliability metrics and by emphasizing governance mechanisms that are implementable in real emergency department workflows.

3. METHOD

Study Design and Scenario Scope

The study adopts a scenario-based quantitative design to evaluate triage architectures under controlled assumptions that represent typical emergency department conditions while allowing systematic variation of uncertainty, crowding, and drift. The objective is not to predict outcomes for a specific hospital but to model decision mechanisms and to compare architectures under consistent and transparent assumptions. The simulated emergency department is assumed to receive a high volume of adult patients with a mixed acuity distribution, including a small fraction of time-sensitive critical illness, a moderate fraction of urgent but stable illness, and a larger fraction of low-acuity conditions. Arrivals follow a time-varying stochastic process with diurnal patterns and occasional surge periods representing outbreaks, mass casualty spillover, or seasonal peaks.

Triage occurs at arrival and produces a priority assignment that influences queue position for provider evaluation. Provider capacity is represented as a set of service channels with variable staffing by hour, and service time distributions are heterogeneous, with higher acuity cases consuming longer provider time and more downstream resources. After provider evaluation, a subset of patients require diagnostics or admission, but for the purposes of triage reliability, the critical focus is the interval from arrival to provider evaluation and the probability of deterioration before evaluation, which is modeled as a function of patient risk and delay.

Patient Risk Model and Outcome Proxies

Each arriving patient is assigned a latent clinical risk state representing the probability of a time-sensitive adverse event if treatment is delayed beyond certain thresholds. This latent risk is not directly observable at triage and must be inferred from features. Observable features include vital sign vector (heart rate, blood pressure, respiratory rate, oxygen saturation, temperature), complaint category, age proxy, and simple risk flags such as immunocompromised status, though these flags may be missing. Features are generated with noise and missingness that increase under crowding, reflecting documentation degradation.

Adverse events are modeled as a hazard process that increases with delay for high-risk patients, allowing computation of a probability that an adverse event occurs before provider evaluation. This is not intended to replicate clinical reality in full detail but to provide an engineering-relevant proxy for the harm associated with under-triage and excessive delay. The model also includes a “deterioration signal” process where some patients’ vital signs worsen while waiting, which can be detected if reassessment occurs, enabling escalation.

Crowding and Queueing Model

The waiting process is modeled as a priority queue, where triage priority determines service order. Provider capacity varies by time, and service times depend on acuity. Under high utilization, waiting time distributions develop heavy tails, and this tail behavior is central to reliability analysis because it determines the exceedance probabilities of clinically meaningful time thresholds. The model includes a left-without-being-seen (LWBS) process for low-acuity patients where abandonment probability increases with waiting time, creating a throughput and safety risk of missed conditions in patients who leave, and it also affects queue composition and subsequent waiting times.

Metrics

The evaluation emphasizes distributional metrics aligned to reliability. Under-triage rate is defined as the fraction of high-risk patients assigned to low or intermediate priority such that their expected time-to-provider exceeds a clinically meaningful threshold. Over-triage rate is defined as the fraction of low-risk patients assigned high priority, creating resource strain. Time-to-provider exceedance probability is computed for multiple thresholds (for example, 30 minutes for high-priority and 120 minutes for intermediate), focusing on tail probabilities rather than mean times. Adverse event before evaluation probability is computed from the hazard model. Nuisance escalation rate is defined as the fraction of escalations that occur in low-risk patients due to noise-triggered reassessment findings, representing operational burden and potential alarm fatigue. A harm and cost index integrates adverse event risk, false priorities, and operational burden into a normalized metric to compare architectures.

4. RESULT AND DISCUSSION

Summary of Findings

The scenario-based evaluation shows that triage reliability is dominated by the tails of waiting time distributions under crowding and by the interaction between evidence uncertainty and reassessment governance. Architectures that rely on static thresholds and discretionary reassessment perform reasonably in average conditions but become fragile under surges, where heavy-tail delays increase under-triage harm even if initial categorization accuracy changes little. Adding more data without governance can worsen reliability by increasing process friction and nuisance escalation, thereby consuming staff attention and amplifying crowding feedback loops. Calibrated risk scoring improves stability and reduces under-triage in normal regimes, but it remains vulnerable when capacity shifts change queue dynamics and when documentation noise increases. A two-tier governed architecture provides the best reliability outcomes because it manages uncertainty explicitly, allocates reassessment resources strategically, and adapts to operational regime while preserving safety bounds.

The first table compares architectures under normal operating conditions with moderate crowding and stable staffing, where the primary reliability challenge is uncertainty in early evidence rather than extreme queue saturation.

Table 1. Core triage reliability outcomes (normal regime, per 10,000 ED arrivals)

Metric	A Baseline static triage	B More data ungoverned	C Calibrated risk score	D Two-tier governed
High-risk arrivals (latent)	420	420	420	420
Under-triaged high-risk patients	54	62	41	28
Under-triage rate (high-risk)	0.129	0.148	0.098	0.067
Over-triaged low-risk patients	620	780	540	510
Over-triage rate (low-risk)	0.071	0.089	0.062	0.058
P(time-to-provider > 30 min for high priority)	0.16	0.18	0.13	0.09
P(time-to-provider > 120 min for intermediate priority)	0.22	0.25	0.19	0.15
Adverse events before evaluation (estimated)	21	24	16	11
Nuisance escalations per 10,000 arrivals	140	260	170	155
Harm-cost index (normalized)	1.00	1.14	0.86	0.71

Source: data proceed

Table 1 indicates that baseline triage produces meaningful under-triage even in normal conditions because early evidence overlap is substantial, and because discretionary reassessment is inconsistent, allowing a subset of high-risk patients to wait beyond safe thresholds without timely escalation. The “more data” approach appears attractive because it increases information, yet the results show worse outcomes because additional data collection without governance increases operational friction and increases nuisance escalations, which in turn consumes staff attention and increases effective waiting times, particularly when escalation protocols are not staged and therefore treat noisy signals as urgent. This finding reflects a reliability principle: more measurement channels can reduce uncertainty only if the system also includes validation, staged interpretation, and disciplined response; otherwise, measurement abundance becomes operational noise.

Calibrated risk scoring improves under-triage and reduces adverse events before evaluation by mapping early evidence to a probabilistic risk and by enabling thresholds that are more consistent than ad hoc clinician judgement, yet it still produces nontrivial under-triage because static thresholds cannot fully handle uncertainty when evidence is missing or ambiguous. Moreover, a single-step policy without explicit uncertainty routing tends to push ambiguous cases either toward high priority, increasing over-triage, or toward lower priority, increasing under-triage, depending on how thresholds are tuned. The two-tier governed architecture improves reliability because it avoids this forced choice: cases with moderate risk or high uncertainty are routed to structured reassessment with explicit time targets, so the system does not rely solely on initial triage assignment to protect high-risk patients. Additionally, trigger-based escalation uses deterioration signals and persistence checks, which reduces nuisance escalation relative to the ungoverned data expansion approach while still capturing true deterioration earlier. The improvement in harm-cost index reflects the system-level nature of the benefit: lower under-triage and fewer adverse events are achieved without excessive over-triage, which means safety improvements do not require unacceptable throughput degradation.

Because emergency departments often experience surges where queueing dynamics shift dramatically, Table 2 evaluates architecture performance under a surge regime characterized by higher arrival rates, reduced staffing elasticity, and increased documentation noise, which together increase waiting time tails and reduce reassessment frequency unless explicitly governed.

Table 2. Stress test outcomes (surge regime, per 10,000 ED arrivals)

Metric	A Baseline static triage	B More data ungoverned	C Calibrated risk score	D Two-tier governed
High-risk arrivals (latent)	520	520	520	520
Under-triaged high-risk patients	98	122	76	49
Under-triage rate (high-risk)	0.188	0.235	0.146	0.094
Over-triaged low-risk patients	870	1180	760	720
P(time-to-provider > 30 min for high priority)	0.29	0.33	0.25	0.17
P(time-to-provider > 120 min for intermediate priority)	0.41	0.46	0.37	0.28
Adverse events before evaluation (estimated)	46	58	34	22
LWBS events (estimated, mostly low-risk)	310	420	290	260
Nuisance escalations per 10,000 arrivals	210	410	260	230
Harm-cost index (normalized)	1.00	1.23	0.88	0.63

Source: data proceed

The surge regime reveals architecture fragility more clearly than the normal regime, because heavy-tail waiting times become the dominant driver of harm and because discretionary reassessment breaks down when workload increases. Baseline triage shows a substantial increase in under-triage rate and adverse events before evaluation, which is consistent with a system that cannot reliably reassess patients and therefore cannot compensate for initial uncertainty. The more-data ungoverned approach performs worst because it increases intake friction and operational noise at precisely the moment when staff time is most constrained, thereby increasing waiting times and accelerating the feedback loop where crowding degrades documentation and reassessment, which further increases missed deterioration.

Calibrated risk scoring remains beneficial under surge, reducing under-triage and adverse events relative to baseline, but it becomes less effective than in the normal regime because static thresholds cannot respond to regime change: when the queue is saturated, even patients categorized as high priority face increased delay exceedance probability, and the risk score does not automatically translate into timely action unless there are explicit capacity-aware controls. Additionally, documentation noise and missingness increase under surge, which increases uncertainty and can reduce score reliability; if the scoring system does not route uncertainty to additional evidence or structured reassessment, the benefits erode. The two-tier governed architecture performs best because it is designed explicitly for such regimes: capacity-aware dynamic thresholds and reassessment allocation reduce delay exceedance probability for the most time-sensitive patients, while structured reassessment and trigger-based escalation provide a mechanism to recover from initial uncertainty and from documentation noise. The reduction in LWBS events indicates a secondary operational benefit: when the system is better governed, waiting times for low-risk patients become less extreme because over-triage and nuisance escalations are controlled, which reduces abandonment and the hidden risk of missed diagnoses in patients who leave.

The reduction in harm-cost index under Architecture D is substantial because the system avoids the worst tail events that dominate harm in surge regimes, namely long delays for high-risk patients and delayed recognition of deterioration. This finding supports an engineering conclusion that triage reliability improvements must target tail risk and must include governance mechanisms that function under crowding, rather than relying exclusively on improved scoring accuracy in calm conditions.

Documentation quality is an operational variable that often receives less attention than model accuracy, yet it can dominate reliability because missingness and noise increase uncertainty and can produce both missed risk and nuisance triggers. Table 3 compares how architectures degrade as documentation noise increases, using a simplified sensitivity test that increases vital sign noise and missingness to reflect high workload, communication barriers, and inconsistent measurement.

Table 3. Sensitivity of under-triage and nuisance escalation to evidence quality degradation

Evidence quality regime	Description	Under-triage rate with A	Under-triage rate with C	Under-triage rate with D	Nuisance escalations with C	Nuisance escalations with D
High quality	Low missingness, stable vitals	0.12	0.09	0.06	160	150
Medium quality	Moderate missingness and noise	0.15	0.12	0.08	220	190
Low quality	High missingness, noisy vitals	0.19	0.17	0.11	330	240

Source: data proceed

Table 3 clarifies why uncertainty governance is central to reliability and why a risk score alone is not sufficient when evidence quality degrades. Baseline triage degrades sharply with lower evidence quality because clinician judgement becomes less anchored, and because discretionary reassessment is not systematically triggered by uncertainty; therefore, more high-risk patients are assigned lower priority and remain in the queue longer. Calibrated risk scoring degrades as well, and the degradation is accompanied by a sharp increase in nuisance escalations because noisy measurements and missingness can produce unstable risk estimates and spurious triggers, particularly if the system is tuned to be sensitive to avoid under-triage. The two-tier governed architecture degrades less because it converts evidence quality degradation into explicit uncertainty, and uncertainty triggers structured reassessment and targeted evidence collection rather than immediate escalation or immediate deprioritization. As a result, under-triage remains lower even under low evidence quality, and nuisance escalations increase less because escalation requires corroboration through persistence checks or multi-signal confirmation, reducing the operational penalty of sensitivity. This sensitivity analysis is reliability-relevant because the most challenging operational periods are exactly those where evidence quality is poorest, and therefore architectures should be evaluated by their robustness to evidence degradation rather than only by their peak performance under ideal documentation.

Discussion

The comparative results can be interpreted through the lens of uncertainty propagation and queueing regime shifts. Under-triage harm is not solely a function of initial misclassification; it is a function of misclassification multiplied by delay, and delay distributions become heavy-tailed under crowding, meaning that a small increase in the probability that a high-risk patient is placed in a lower-priority queue can yield a disproportionate increase in the probability of a dangerous delay. Therefore, reliability improvements that modestly reduce under-triage can yield large safety benefits under surge, and this is visible in the reduction of adverse event proxy counts under Architecture D.

Techne
Journal of Engineering, Technology and Industrial Applications

Over-triage, by contrast, is harmful primarily through capacity coupling. When low-risk patients are assigned high priority, they increase the queue weight for high-priority channels, increasing waiting time for truly high-risk patients and increasing the probability of delay exceedance; this effect is nonlinear because it pushes the system toward saturation. A reliable system therefore must not simply "be conservative" by over-triaging broadly; it must be selectively conservative, increasing priority when evidence is strong or when uncertainty is high and delay sensitivity is high, while using structured reassessment to resolve uncertainty rather than defaulting to high-priority assignment. Architecture D embodies this logic, and the results indicate that it reduces under-triage without explosive over-triage.

Nuisance escalations matter because they consume the scarce resource that is central to triage reliability, namely attention and reassessment capacity. In crowded environments, additional alarms and escalations can increase workload and degrade overall observation quality, creating a negative spiral. A system that triggers frequent nuisance escalations is therefore unreliable even if it detects true deterioration, because it reduces the probability that staff can respond effectively when genuine high-risk events occur. This is why governance mechanisms such as persistence checks, corroboration across signals, and risk-aligned staging are critical: they reduce nuisance while preserving sensitivity.

Dynamic thresholding is sometimes misinterpreted as "tightening thresholds during crowding," but the results support a more nuanced view: reliable dynamic thresholding must preserve safety bounds and should adapt by reallocating reassessment intensity and by focusing verification on uncertainty rather than by globally increasing declines or deprioritizations. In emergency triage, this means that during crowding, the system should not simply reduce the number of high-priority assignments to improve throughput, because this increases under-triage; instead, it should focus on rapid identification and protection of the time-sensitive subset through expedited pathways and structured reassessment, while managing low-risk flows through mechanisms such as fast-track channels and clear communication that reduces LWBS. Architecture D, by design, uses capacity-aware adjustments that target these mechanisms.

Practical Implementation Implications

Although the study is scenario-based, the architecture principles are implementable in real emergency departments with existing data streams, because many of the required elements are operational rather than computational. Calibration of risk scores can be achieved using historical outcomes, but it must be accompanied by monitoring and governance that detects drift, such as shifts in score distribution and changes in escalation rates. Uncertainty estimation can be implemented using simple proxies such as missing vital signs, near-threshold values, and inconsistent readings across repeated measures, and uncertainty can trigger structured reassessment rather than immediate escalation. Structured reassessment can be operationalized as timed checks for selected patients, and trigger-based escalation can use simple criteria such as persistent abnormal vital signs or worsening patterns rather than single measurements, reducing nuisance triggers.

Capacity-aware controls require operational visibility into queue lengths and staffing, and while sophisticated optimization is possible, many reliability gains can be achieved through policy constraints such as ensuring that a minimum fraction of high-risk predicted patients are evaluated within a time window, and adjusting reassessment intensity to maintain that constraint. The most important requirement is governance discipline, because without disciplined response protocols, additional data and scoring outputs can create noise that undermines reliability.

5. CONCLUSION

Emergency department triage should be designed and evaluated as a reliability decision system because patient harm and operational failure are dominated by tail events that arise from uncertainty, crowding-driven heavy-tail delays, and delayed or inconsistent escalation. The scenario-based quantitative evaluation demonstrates that baseline static triage with discretionary reassessment becomes fragile under surge regimes, that adding data without governance can worsen reliability by increasing friction and nuisance escalations, and that calibrated risk scoring improves stability but remains vulnerable when operational regimes shift and when documentation quality degrades. A two-tier governed architecture that integrates calibrated risk and uncertainty estimation, structured reassessment with trigger-based escalation, and capacity-aware dynamic controls that preserve safety bounds reduces under-triage and time-to-provider exceedance probabilities while stabilizing workload and reducing adverse-event exposure, with the strongest advantages observed during surges where heavy-tail delays dominate. The central engineering implication is that triage reliability depends as much on governance and staged decision pathways as on the scoring method used to estimate risk; therefore, emergency department modernization efforts should prioritize end-to-end decision pipeline design, including uncertainty routing, reassessment discipline, drift monitoring, and operational constraints that prevent runaway crowding effects.

REFERENCES

1. Adari, V. K. (2020). Intelligent Care at Scale AI-Powered Operations Transforming Hospital Efficiency. *International Journal of Engineering & Extended Technologies Research (IJEETR)*, *2*(3), 1240–1249.
2. Akter, M. S., Sultana, N., Khan, M. A. R., & Mohiuddin, M. (2023). Business Intelligence-Driven Healthcare: Integrating Big Data and Machine Learning For Strategic Cost Reduction And Quality Care Delivery. *American Journal of Interdisciplinary Studies*, *4*(02), 1–28.
3. Bhuiyan, M. (2025). AI-Driven Crisis Communication And Emergency Response: Optimizing Nonprofit Digital Outreach During National Disasters. *International Journal of Scientific Interdisciplinary Research*, *6*(1), 293–326.
4. DeHollander, A., Karwan, M., & Casucci, S. (2025). A Comprehensive Review of Patient Prioritization Strategies for Mitigating Emergency Department Crowding and Enhancing Efficiency. *Available at SSRN 5099607*.
5. Druetto, A., Aringhieri, R., Duma, D., & Guastalla, A. (2024). Dispatching strategy with lookahead for the real-time management of ambulances. *50th Annual Meeting of the EURO Working Group on Operational Research Applied to Health Services*, 20.
6. Fan, L., Zuo, X., Tang, L., Xiong, H., You, Y., & Fan, C. (2025). Performance comparison of artificial intelligence models in predicting 72-h emergency department unscheduled return visits. *Frontiers in Public Health*, *13*, 1609206.
7. Griffin, P. M., Nembhard, H. B., DeFlitch, C. J., Bastian, N. D., Kang, H., & Muñoz, D. A. (2016). *Healthcare systems engineering*. John Wiley & Sons.
8. Hao, S., Sun, Z., & Xu, Y. (2020). *Emergency care access vs. quality: Uncovering hidden consequences of fast-track routing decisions*. Working paper.
9. Jaghni, K. K. (2025). HealthNavAI: An AI-Powered National Platform for Real-Time Healthcare Service Availability and Patient Routing. *Journal of Computer Science and Technology Studies*, *7*(10), 426–443.
10. Joseph, M. M., Mahajan, P., Snow, S. K., Ku, B. C., Saidinejad, M., & MEDICINE, C. O. N. P. E. (2022). Optimizing pediatric patient safety in the emergency care setting. *Pediatrics*, *150*(5).
11. Karkkainen, B. C. (2019). Information as environmental regulation: TRI and performance benchmarking, precursor to a new paradigm? In *Environmental law* (pp. 191–304). Routledge.
12. Lameesa, A., Hoque, M., Alam, M. S. Bin, Ahmed, S. F., & Gandomi, A. H. (2024). Role of metaheuristic algorithms in healthcare: a comprehensive investigation across clinical diagnosis, medical imaging, operations management, and public health. *Journal of Computational Design and Engineering*, *11*(3), 223–247.
13. Machireddy, J. R. (2023). Optimizing healthcare resource allocation for operational efficiency and cost reduction using real-time analytics. *Nuvern Applied Science Reviews*, *7*(3), 12–33.
14. O'Hara, R., Johnson, M., Hirst, E., Weyman, A., Shaw, D., Mortimer, P., Newman, C., Storey, M., Turner, J., & Mason, S. (2014). A qualitative study of decision making and safety in ambulance service transitions. *Health and Social Care Delivery Research*, *2*(56), 1–138.
15. Ortíz-Barrios, M. A., & Alfaro-Saíz, J.-J. (2020). Methodological approaches to support process improvement in emergency departments: a systematic review. *International Journal of Environmental Research and Public Health*, *17*(8), 2664.
16. Sahlaoui, F.-Z., Aboueljinane, L., & Lebbar, M. (2023). A review on simulation-based metamodeling in emergency healthcare: methodology, applications, and future challenges. *Simulation*, *99*(10), 989–1009.

17. Sultana, R. (2022). A Review Of Data-Driven Communication In Economic Recovery: Implications Of ICT-Enabled Strategies For Human Resource Engagement. *International Journal of Business and Economics Insights*, *2*(1), 1–34.
18. Tsao, J. Y., Trucano, T. G., Kleban, S. D., Naugle, A. B., Verzi, S. J., Swiler, L. P., Johnson, C. M., Smith, M. A., Flanagan, T. P., & Vugrin, E. D. (2016). *Complex Systems Models and Their Applications: Towards a New Science of Verification, Validation & Uncertainty Quantification*. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).
19. Vasilakis, C., & Wood, R. (2020). Analyst-driven development of an open-source simulation tool to address poor uptake of OR in healthcare. *46th Meeting of the EURO Working Group on Operational Research Applied to Health Services*, 4.
20. Youn, S., Geismar, H. N., & Pinedo, M. (2022). Planning and scheduling in healthcare for better care coordination: Current understanding, trending topics, and future opportunities. *Production and Operations Management*, *31*(12), 4407–4423.
21. Youssef, E., Benabbas, R., Choe, B., Doukas, D., Taitt, H. A., Verma, R., & Zehtabchi, S. (2024). Interventions to improve emergency department throughput and care delivery indicators: A systematic review and meta-analysis. *Academic Emergency Medicine*, *31*(8), 789–804.