An investigation into hospital triage

ICA 2 Stat0009

2024-02-20

Contents

1 Introduction to Hospital Triage

Hospital triage is a critical process used in emergency medicine to manage patient flow efficiently and prioritize care based on the urgency of medical conditions. This system originated from military practices, where it was essential to classify soldiers' injuries to decide who needed immediate attention (Mitchell, 2008). The principle behind triage is to do the greatest good for the greatest number of people, which is especially vital during mass casualty incidents or when healthcare resources are limited.

The current healthcare scenario emphasizes the critical need for efficient triage due to escalating pressures on emergency departments. Recent data from (Nuffield Trust, 2024) reveals that the median waiting time for A&E patients in England reached 1 hour and 4 minutes in November 2023, indicating significant challenges in delivering timely care. Furthermore, projections from (The Health Foundation, 2023) suggest that the NHS waiting list could surpass 8 million by summer 2024, pointing to increasing healthcare demands.

These statistics highlight the importance of triage in managing patient flow and ensuring that urgent cases receive prompt attention. Triage ensures that patients with life-threatening conditions receive immediate care, thus maximizing the chances of survival and recovery. By assessing and categorizing patients based on the severity of their conditions, triage helps in reducing waiting times for critical cases and makes the best use of available medical resources. This system also helps in preventing the overcrowding of emergency departments, thereby maintaining a manageable workflow and reducing stress on healthcare professionals (Love et al., 2012).

1.1 Overview of existing triage systems

There are several triage systems used worldwide, each with its methodologies and criteria for prioritizing patients. Some of the most common systems include (Hinson et al., 2019):

Manchester Triage System (MTS): Originating in the UK, the MTS uses a flowchart system to categorize patients into five urgency levels, ranging from immediate attention to non-urgent.

Emergency Severity Index (ESI): Widely used in the United States, the ESI is a five-level triage algorithm that categorizes patients based on the severity of their conditions and the number of resources their care is anticipated to require.

Australian Triage Scale (ATS): Similar to other systems, the ATS classifies patients into five categories, from those needing immediate attention to those who can afford to wait longer.

Canadian Triage and Acuity Scale (CTAS): In Canada, the CTAS is utilized, which similarly assigns patients to one of five categories, ranging from resuscitation to non-urgent, based on their health condition's severity and complexity.

South African Triage Scale (SATS): Used in South Africa, SATS prioritizes emergency department patients with a scoring system based on vital signs and presenting complaints to determine urgency levels for treatment.

These systems differ mainly in their assessment criteria, algorithms, and the specific terms used to categorize the levels of urgency. However, all aim to quickly identify patients who need immediate life-saving interventions and manage resources effectively. See Figure 1 for a detailed overview of the triage systems.

The evolution of hospital triage has been marked by a continuous effort to improve accuracy, efficiency, and fairness in patient assessment. From simple, subjective decision-making based on visible injuries or symptoms, triage has evolved into a more systematic and objective process. Innovations in triage include the incorporation of technology, such as computer-based algorithms and digital health monitoring systems, which help in standardizing the process, reducing human error and increasing real-time communication (Gao et al., 2007).

Triage System	CTAS	ESI	MTS	ATS	SATS
Stated objective	Provide patients with timely care	Prioritize patients by immediacy of care needs and resource	Rapidly assess a patient and assign a priority based on clinical need	Ensure patients are treated in order of clinical urgency and allocate patients to the most appropriate treatment area	Prioritize patients based on medical urgency in contexts where there is a mismatch between demand and capacity
Recommended time to physician contact, min	1: immediate 2: ≤15 3: ≤30 4: ≤60 5: ≤120	1: immediate 2: ≤15 3: none 4: none 5: none	Red: immediate Orange: ≤10 Yellow: ≤60 Green: ≤120 Blue: ≤240	1: immediate 2: ≤15 3: ≤30 4: ≤60 5: ≤120	Red: immediate Orange: ≤10 Yellow: ≤60 Green: ≤240 Blue: ≤120
Discriminators					
Clinical	Yes	No	Yes	Yes	Yes
Vital signs	Yes	Yes	Yes	Yes	Yes
Pain score	Yes (10-point)	Yes (visual analog scale)	Yes (3-point)	No	Yes (4-point)
Resource use	No	Yes	No	No	No
Pediatrics	Separate version	Separate vital sign differentiators	Considered within algorithm	Considered within algorithm	Separate flowchart

Figure 1: Triage Overview. Source: Hinson et al. (2019)

1.2 Canadian Triage and Acuity Scale (CTAS)

In this project, we will use the CTAS framework due to its comprehensive and clear directives for assigning treatment times based on patient severity, along with the availability of detailed data that supports modeling decisions. Below is a breakdown of the different triage levels (Emergency Physicians, 2012):

- **Level 1: Resuscitation**: E.g. Cardiac arrest or severe respiratory distress.
- **Level 2: Emergent**: E.g. Chest pain with sweating suggestive of a heart attack.
- **Level 3: Urgent**: E.g. Severe abdominal pain that could indicate a serious condition.
- **Level 4: Less Urgent**: E.g. A sprained ankle with swelling and moderate pain.
- **Level 5: Non-Urgent**: E.g. A patient with a minor chronic issue seeking a prescription refill.

Yoon et al. (2003) provides detailed data on the distribution of these CTAS levels by through studying $n = 894$ patients in a time frame of 7 days at the University of Alberta Hospital:

Triage level	*n* (%)	ED registration to triage assessment	Triage assessment to nursing assessment	Nursing assessment to physician assessment	Physician assessment to disposition decision	Disposition decision to actual departure	Total ED LOS (SD)
I	9 (1.0)	2.8	0.4	1.6	67.0	79.6	151.3 (99.3)
II	55 (6.2)	2.6	4.5	7.5	190.8	95.4	300.8 (251.4)
III	297 (33.2)	7.7	12.7	32.8	245.9	67.4	366.4 (266.5)
IV	327 (36.6)	13.9	25.8	35.5	155.3	20.7	251.2 (199.0)
V	206 (23.0)	13.8	18.3	34.8	83.8	11.3	162.1 (173.0)
All	894 (100)	11.0	18.2	32.4	170.2	39.2	271.0 (173.0)

*Triage levels determined by the *Canadian Emergency Department Triage and Acuity Scale* (CTAS).
ED = emergency department; LOS = length of stay; SD = standard deviation

Figure 2: Emergency Room times conditional on triage level. Source: Yoon et al. (2003).

The distribution of patients across the Canadian Triage and Acuity Scale (CTAS) levels in the study, where 93% of patients fell into levels 3, 4, and 5, reflects a common trend in emergency departments (EDs) where the majority of cases are of less severe nature. Counter-intuitively, patients of CTAS Level 3 spent the longest time in the emergency department. Yoon et al. (2003) state this is because patients often arrive with unclear symptoms, making it difficult to immediately decide on admitting or discharging them, leading to longer stays in the ED for further observation and testing.

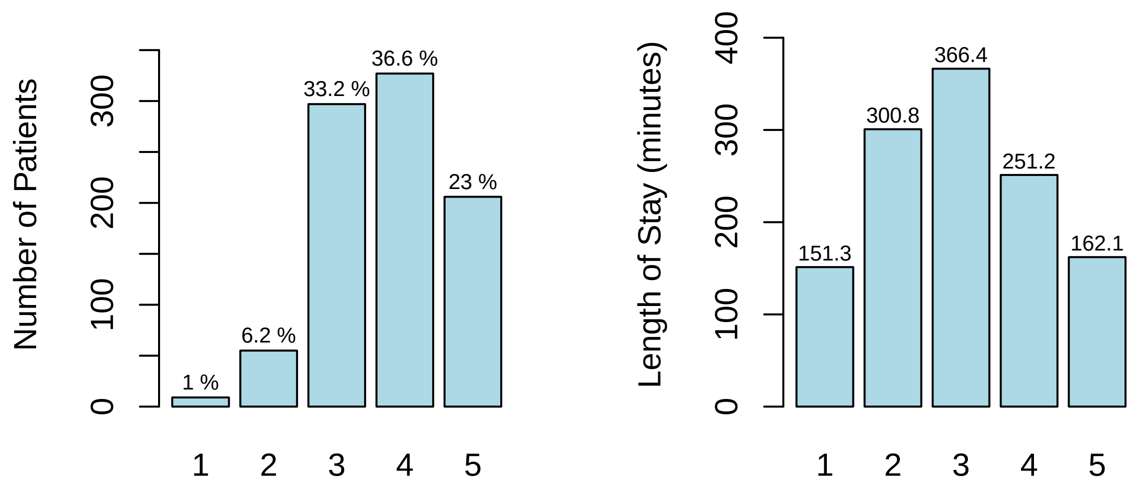

Figure 3: Number of Patients per Triage Level (left). Total Length of Stay per Triage Level (right).

We will make informed decisions based on the data from Yoon et al. (2003) within our paper, as we will have reference values to benchmark and evaluate our models.

2 Aim of project

The aim of this project is to analyse hospital triage systems through the lens of queueing theory and simulations, thereby uncovering factors that influence their efficiency and effectiveness. By simulating various queueing scenarios, this project seeks to bridge the gap between theoretical models and the complex realities of emergency department operations. Our objective is to assess the performance of current triage practices, particularly those akin to the Canadian Triage and Acuity Scale (CTAS), under different conditions and constraints. This investigation will enable us to evaluate the robustness of existing triage systems based on empirical evidence and theoretical insights.

2.1 Overview of simulations

The project is structured around five key simulation scenarios, each designed to iteratively simulate hospital triage dynamics with increasing accuracy:

1. **MG1 Queue Simulation**: This foundational model will simulate a basic queueing system with a single server, representing a simplified version of patient flow in an emergency department. This scenario will serve as a benchmark, helping us understand the baseline performance of a triage system under the simplest conditions.

2. **Priority Queue with Severity (1-5 mimicking the Canadian scale)**: Building on the MM1 model, this simulation will introduce a priority queueing mechanism based on a 5-level severity scale similar to the CTAS. This scenario aims to explore how prioritization based on severity affects patient wait times and overall system efficiency.

3. **MMC Queue with Priority**: This model will extend the priority queue concept to a more complex MMC system, where multiple servers are available, representing multiple healthcare providers in an emergency department. This scenario will help us understand the impact of resource availability on the performance of a priority-based triage system.

4. **MMC with Conditional Patient Departure**: This simulation will investigate scenarios where patients decide to leave the queue based on the severity of their condition. The likelihood of departure will vary, with those having less severe conditions potentially leaving sooner due to prolonged wait times. This approach will allow us to analyze the impact of patient abandonment on system efficiency and explore strategies to retain patients until they receive care.

5. **MMC with Markov Chain for Condition Deterioration**: This simulation will use a Markov chain to model the progression of patients' conditions over time, including the possibilities of deterioration, departure due to frustration or perceived neglect, and death. This model will provide insights into how the triage system adapts to patients' evolving needs and the effectiveness of the system in preventing adverse outcomes through timely intervention.

6. **Incorporation of Real-World Data**: The final simulation will integrate real-world data into a sophisticated model to create a more accurate representation of an emergency department's triage system. This scenario aims to validate the theoretical models and simulations against actual patient flow data, offering a grounded assessment of triage system performance.

By examining these diverse scenarios, the project aims to identify key factors that influence the efficacy of triage systems, such as resource allocation, patient prioritization strategies, and the handling of dynamic patient states.

2.2 Measures of performance

In our project, we will employ key metrics to evaluate the effectiveness and shortcomings of hospital triage systems, focusing on quantifying the system's "loss" or "badness." These metrics will help in understanding the areas where triage systems may not perform optimally under various simulated conditions.

Total Waiting Time: This metric will assess the cumulative waiting period for patients before receiving care. Extended waiting times can indicate inefficiencies and can significantly impact patient outcomes, especially for those with urgent needs.

Violations of Triage System Guidelines: We will measure instances where the simulated triage processes fail to meet the established guidelines for maximum waiting times, based on the severity of patients' conditions. This will help identify how often and by how much current triage practices deviate from their intended standards.

Number of Patients Left Unseen (Indicator of Satisfaction): This metric counts patients who leave without being seen, serving as an indirect measure of patient satisfaction. High numbers may indicate dissatisfaction due to long waits, highlighting areas for improvement in the triage system to enhance patient experience and reduce departures. We aim to minimize the number of patients left unseen, ensuring everyone receives timely care, especially those in urgent need.

By integrating these metrics into our simulation analysis, we aim to create a holistic assessment of triage systems, identifying key areas for improvement. This approach will enable us to propose targeted recommendations for enhancing triage efficiency, reducing waiting times, adhering more closely to triage guidelines, and ultimately, minimizing preventable patient deaths.

3 Simulations

3.1 M/G/1 FIFO Queue

For modeling an emergency department as a simple starter case with one doctor, we employ the MG1 queueing model. In this scenario, "M" indicates that patient arrivals follow a Poisson process, reflecting the randomness of emergency visits. "G" signifies that the service times, or the time each patient spends with the doctor, follow a general distribution—specifically, a Gamma distribution with shape 6 and rate 0.1, as recommended by Oladimeji & Ibidoja (2020) for healthcare settings. The "1" in MG1 highlights that there is a single server, in this case, one doctor, available to attend to patients. Additionally, FIFO means the queue uses a First In First Out System, where patients are prioritized by their arrival time.

We know that the median time spent waiting in the emergency department was 64 minutes in November 2023 (Nuffield Trust, 2024). However, we could not find a reasonable parameter value for the exponential arrivals (λ) and will thus calculate it theoretically.

To calculate the value of λ, we will use the Pollaczek–Khinchine formula (Pollaczek, 1930) (Khinchin, 1932), which gives the relationship between the Laplace transform of the service time distribution and the queue length. However, it can be recast to show the relationship between the mean length of the queue, and the mean waiting time (Asmussen, 2003, p. 192):

$$L = \rho + \frac{\rho^2 + \lambda^2 \operatorname{Var}(S)}{2(1-\rho)} \tag{1}$$

In the above, L denotes the mean length of the queue, ρ renotes the capacity utilization of the queue, and S is the distribution of the service time. Furthermore, ρ is defined as:

$$\rho = \frac{\lambda}{\mu}$$

where:

$$\mu = \frac{1}{\mathbb{E}(S)}$$

thus:

$$\rho = \lambda \mathbb{E}(S)$$

For a stable system, we desire $\rho \leq 1$ (Virtamo, 2008). If $\rho \geq 1$, the queue length would tend to infinity, as more people arrive than can be served. In our case, we want to find a value of λ such that $S \sim \text{Gamma}(6, 0, 1)$, and ρ. In order to find a value of λ such that the expected waiting time is 64 minutes, we turn to Little's Law, which gives:

$$L = \lambda W \tag{2}$$

where W is the mean total time spent in a system. We will define W_1 as the time spent in the waiting room, and thus:

$$W = W_1 + \mathbb{E}(S) \tag{3}$$

Substituting Eq. (1) and Eq. (3) into Eq. (2) and re-arranging:

$$\frac{L}{\lambda} = W_1 + \mathbb{E}(S)$$

$$\frac{\rho}{\lambda} + \frac{\rho^2 + \lambda^2 \operatorname{Var}(S)}{2\lambda(1-\rho)} = W_1 + \mathbb{E}(S)$$

As $\frac{\rho}{\lambda} = \mathbb{E}(S)$:

$$\frac{\rho^2 + \lambda^2 \operatorname{Var}(S)}{2\lambda(1-\rho)} = W_1$$

Using $\rho = \lambda \mathbb{E}(S)$:

$$\frac{(\lambda \mathbb{E}(S))^2 + \lambda^2 \operatorname{Var}(S)}{2\lambda(1-\lambda \mathbb{E}(S))} = W_1$$

Simplifying:

$$\lambda \frac{\mathbb{E}(S)^2 + \operatorname{Var}(S)}{2(1-\lambda \mathbb{E}(S))} = W_1$$

The above could be further simplified by noticing $\mathbb{E}(S)^2 + \operatorname{Var}(S) = \mathbb{E}(S^2)$, but we choose to leave the equation in the above form for ease of calculation.

Recall the objective is to calculate λ such that $w_1 = 64$, as motivated by (Nuffield Trust, 2024). We therefore substitute parameters α and β into our above equation:

$$\lambda \frac{\left(\frac{\alpha}{\beta}\right)^2 + \frac{\alpha}{\beta^2}}{2\left(1 - \lambda \frac{\alpha}{\beta}\right)} = W_1$$

Substituting our desired values:

$$\lambda \frac{\left(\frac{6}{0.1}\right)^2 + \frac{6}{0.1^2}}{2\left(1 - \lambda \frac{6}{0.1}\right)} = 64$$

$$\lambda \frac{60^2 + 600}{2 - 120\lambda} = 64$$

This can be solved numerically (e.g. via. Newton-Raphson) to find:

$$\lambda \approx 0.0108$$

We can verify this using simulations to find that the mean waiting time in our M/G/1 queue is indeed 64 minutes, using the law of large numbers.

```r
# Define a function to create severities with probabilities from Yoon et al. (2003)
severity_gen <- function(n) {
    # Generates a sample of severity levels (1 to 5) for 'n' patients
    sample(1:5, n, replace = TRUE, prob = c(0.01, 0.062, 0.332, 0.366, 0.23))
}

# Simulates a triage system using a gamma distribution for service times
# from Oladimeji & Ibidoja (2020)
triage_sim_gamma <- function(n, arrival_rate) {
    # Generate cumulative arrival times based on an exponential distribution
    arrivals <- cumsum(rexp(n, arrival_rate))
    # Generate a list of severities for 'n' patients
    severity_list <- severity_gen(n)
    # Combine arrivals and severities to represent patients in the waiting room
    waiting_room <- cbind(arrivals, severity_list)

    # Initialize a matrix to store wait times and priorities
    wait_times <- matrix(numeric(0), nrow = 0, ncol = 2)
    # Initialize service time
    services <- 0

    # Process patients until all have been served
    while(services < max(arrivals) && nrow(waiting_room) > 0) {

        # Identify patients available for service
        available_patients <- waiting_room[waiting_room[, 1]<=services,,drop = FALSE]
        # Assume the next patient is the first in line (FIFO)
        next_patient_index <- 1
        next_patient <- waiting_room[next_patient_index, , drop = FALSE]
        # Calculate wait time for the patient
        wait <- max(0, services - next_patient[1])
        priority <- next_patient[2]

        # Add patient's wait time and priority to the record
        wait_times <- rbind(wait_times, c(wait, priority))
        # Remove the served patient from the waiting room
        waiting_room <- waiting_room[-next_patient_index, , drop = FALSE]
        # Update the chosen server's next available service time
        services <- max(services, next_patient[1]) + rgamma(1, 6, 0.1)
    }

    # Assign column names to the wait times matrix
    colnames(wait_times) <- c("Wait", "Priority")
    return(wait_times)
}

# Using arrival rate = 0.0108 as calculated
queue = triage_sim_gamma(n=20000, arrival_rate = 0.0108)
# Calculate and print the mean wait time
mean(queue[,"Wait"])
```

```
## [1] 64.15783
```

In comparing the average waiting times per severity level within this service system, it becomes evident that

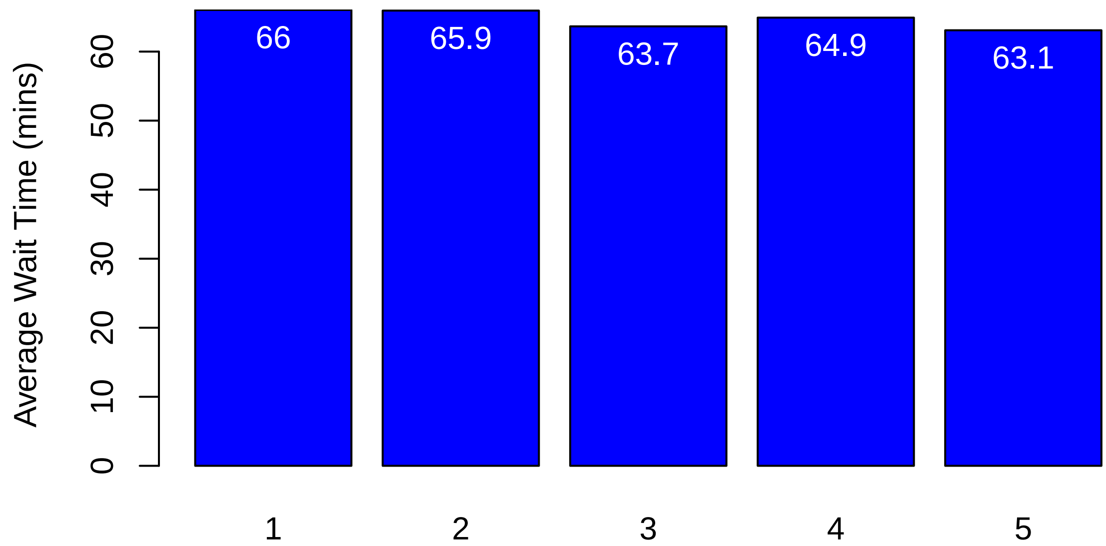

Figure 4: Average Wait Times For M/G/1 Queue by Severity

despite variations in patient severity, the average waiting times remain remarkably uniform. This uniformity raises concerns regarding the efficiency of the FIFO service process in adequately addressing the needs of patients with varying degrees of severity. Particularly troubling is the potential impact on patients with lower severity levels, who may face disproportionately longer waits in spite of higher urgency. Such delays in accessing medical attention pose significant risks, as prolonged wait times could exacerbate their conditions and lead to detrimental outcomes. This highlights the critical importance of reevaluating the FIFO service approach within the broader context of patient care and service system. We know that hospitals typically employ more nuanced triage systems that take other factors such as severity and total time waited, into account. Moreover, the inclusion of the FIFO system in our analysis serves as a foundation benchmark for our coming exploration from which we will evaluate and refine more sophisticated triage models.

3.2 M/G/K Queue

Expanding our analysis to a M/G/K queue, which involves multiple servers, brings us closer to real-world situations by adding complexity to the model. One significant factor that has not yet been incorporated is the consideration that patients with higher severity levels require longer treatment times. To simulate this, we have chosen to employ a service time distribution modeled by a $Gamma(6, \frac{severity}{35})$ distribution, with the parameters being conditional on the patient's severity level. This decision is grounded in the work of Oladimeji & Ibidoja (2020), which indicates that the mean service time in a hospital setting is approximately 60 minutes. Consequently, the service time distribution for a patient X_i, with a severity level $i \in 1, 2, 3, 4, 5$, is given by $X_i \sim Gamma\left(6, \frac{i}{35}\right)$. The weighted average of the service time across our severity-based system can be represented as:

$$\sum_{i=1}^{5} \mathbb{E}[X_i] P(X = i).$$

In our specific case, this equation simplifies to:

$$210 \left(0.01 + \frac{1}{2} \cdot 0.062 + \frac{1}{3} \cdot 0.332 + \frac{1}{4} \cdot 0.366 + \frac{1}{5} \cdot 0.23\right) = 60.725.$$

The proportions of severities follows from Yoon et al. (2003). This approach allows us to more accurately model the service dynamics within a hospital environment, accounting for the variability in treatment times as a function of patient severity. Additionally, our figure of 60.7 is close to our value of a 60 minute service time. According to (Millington, 2018), there are typically 12 specialist ER doctors available, which is the figure we will use henceforth.

The above plot indicates that waiting times remain indistinguishable across different severities, mirroring the findings observed in the M/G/1 FIFO system. This uniformity in waiting times raises the same concerns regarding its implications for patient outcomes and satisfaction as before.

This suggests that, irrespective of the complexity of the service distribution, a FIFO service approach fails to adequately address the varying urgency of patient needs, thus motivating the development of a dynamic service process that incorporates several factors other than the arrival time of patients when prioritising the queue.

Our choice of including multiple servers in the simulation aligns with the resource allocation typically observed in triage settings found in average hospitals in Europe and North America. However, we anticipate that the results from our recommendations will be effective in a variety of triage scenarios regardless of their respective resources.

3.3 M/G/K Priority Queue

In our revised queuing system, we have implemented a prioritization process aimed at optimizing the time until a patient is attend to. Whereas the previous approach where patients were ordered by the time of their arrival, our system now prioritises patients according to their severity scores, with higher urgency (lower severity score) patients receiving immediate attention. In cases where multiple patients share the same severity score, priority is given to the patient who arrived first at the hospital.

Our analysis of the resulting waiting times, as depicted in the plot above, demonstrates the efficacy of this new process in minimizing wait times for patients requiring urgent care. The waiting times across different severity levels exhibit an observable pattern, with patients of higher urgency experiencing shorter wait times compared to those with lower severity scores. This reduction in wait times for high-urgency patients underscores the success of our system in aligning service delivery with patient needs and priorities.

While it is evident that the total time spent in the system remains relatively high for patients with the highest urgency (severity of 1), it is important to emphasize that our primary objective is to minimize the average time until a patient is seen, particularly for those in urgent need of care. By focusing on this key metric, this system could be considered as significantly more effective than the FIFO system.

However, although the urgency of the patients with a CTAS score of 5 is low, in accordance with patient care thresholds regarding the maximum amount of time patients should have to wait to be seen, the waiting time for these patients could be seen as too high. Therefore we will attempt to lower this time, on average, by introducing other factors into our service process prioritisation.

3.4 Patient depature

To model patient departure due to exceeded wait times in our simulation, we employ the Weibull distribution, as proposed by Wiler et al. (2013), to determine their tolerance levels. If the waiting period surpasses a patient's calculated tolerance threshold, the patient will opt to leave. Specifically, we utilize the Weibull distribution parameters with a shape factor of 1.3 and a scale factor of 1000, denoted as $\text{Weibull}(1.3, 1000)$, to compute these tolerance times.

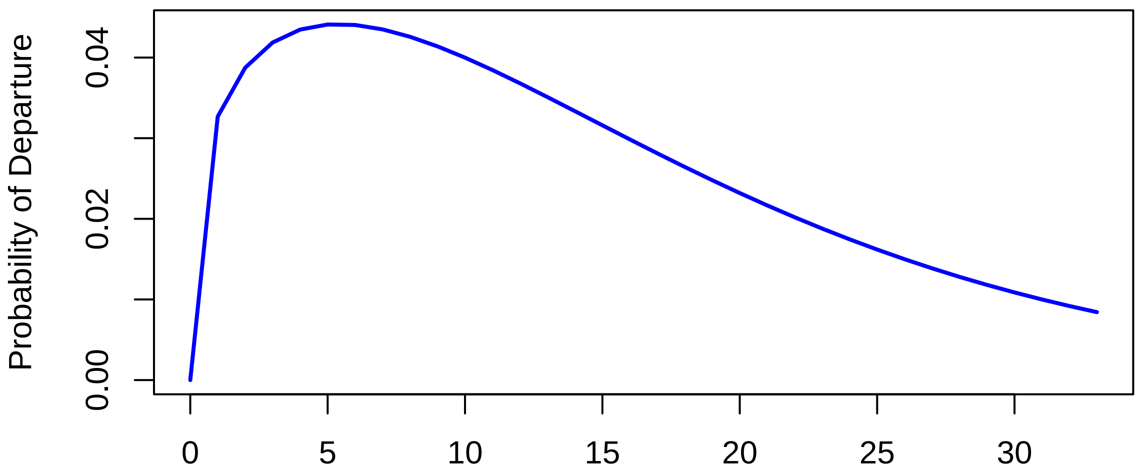

Figure 5: Weibull distribution of patient waiting time tolerance.

3.5 Condition Deterioration

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \frac{1}{20} & -\frac{1}{20} & 0 & 0 & 0 \\ 0 & \frac{1}{40} & -\frac{1}{40} & 0 & 0 \\ 0 & 0 & \frac{1}{60} & -\frac{1}{60} & 0 \\ 0 & 0 & 0 & \frac{1}{80} & -\frac{1}{80} \end{bmatrix}$$

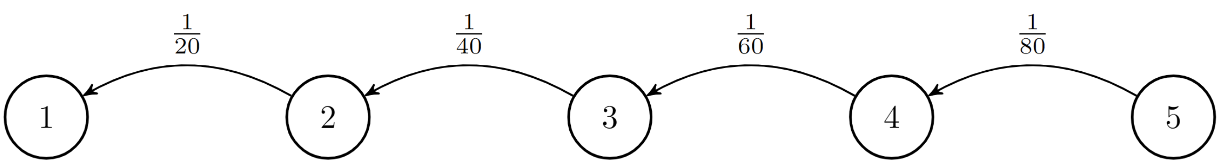

Figure 6: Diagram of Markov Chain Representing Severity Deterioration.

4 Extreme Scenario Simulation

5 Limitations

In our study, we simulated the triage process within a hospital setting under the assumption that all individuals undergo correct classification according to their immediate medical needs. However, this assumption presents a significant limitation to the realism and applicability of our findings to actual healthcare environments. In real-life scenarios, the triage process is subject to human error and variability in judgment, leading to potential misclassification of patients (Zachariasse et al., 2019). This discrepancy can affect patient outcomes, resource allocation, and overall efficiency of emergency care services. Consequently, while our simulation provides valuable insights into the potential functioning of an ideal triage system, it may not fully capture the complexities and challenges inherent in real-world triage practices, including the implications of misclassification on patient care and system performance.

Furthermore, we assume a constant rate of patient arrivals throughout each day and year, which simplifies real-world dynamics where patient flow varies by time of day and season. This approach overlooks important fluctuations, such as increased visits during holidays or flu season, noted by Zhang et al. (2022). Ignoring these variations can limit the accuracy of our findings in predicting hospital operations and resource needs, as actual patient arrivals often dictate staffing and resource allocation.

References

Asmussen, S. (2003). *Applied probability and queues* (Vol. 51). Springer New York. https://doi.org/10.1007/b97236

Emergency Physicians, C. A. of. (2012). *The canadian triage and acuity scale combined adult/paediatric educational program*. https://ctas-phctas.ca/wp-content/uploads/2018/05/participant_manual_v2.5b_november_2013_0.pdf

Gao, T., Massey, T., Selavo, L., Crawford, D., Chen, B., Lorincz, K., Shnayder, V., Hauenstein, L., Dabiri, F., Jeng, J., Chanmugam, A., White, D., Sarrafzadeh, M., & Welsh, M. (2007). The advanced health and disaster aid network: A light-weight wireless medical system for triage. *IEEE Transactions on Biomedical Circuits and Systems*, *1*, 203–216. https://doi.org/10.1109/TBCAS.2007.910901

Hinson, J. S., Martinez, D. A., Cabral, S., George, K., Whalen, M., Hansoti, B., & Levin, S. (2019). Triage performance in emergency medicine: A systematic review. *Annals of Emergency Medicine*, *74*, 140–152. https://doi.org/10.1016/j.annemergmed.2018.09.022

Khinchin, A. Ya. (1932). Mathematical theory of stationary queues. *Matematicheskii Sbornik*, *39*, 73–84.

Love, R. A., Murphy, J. A., Lietz, T. E., & Jordan, K. S. (2012). The effectiveness of a provider in triage in the emergency department: A quality improvement initiative to improve patient flow. *Advanced Emergency Nursing Journal*, *34*, 65–74. https://doi.org/10.1097/TME.0B013E3182435543

Millington, H. (2018). *Fifty years in the NHS doesn't feel like very long...*. https://www.imperial.nhs.uk/about-us/blog/fifty-years-in-the-nhs-does-not-feel-like-very-long.

Mitchell, G. W. (2008). A brief history of triage. *Disaster Medicine and Public Health Preparedness*, *2*(S1), S4–S7. https://doi.org/10.1097/DMP.0b013e3181844d43

Nuffield Trust. (2024). *A&E waiting times*. https://www.nuffieldtrust.org.uk/resource/a-e-waiting-times

Oladimeji, A. S., & Ibidoja, O. J. (2020). On the distribution of service time of patients. *Journal of Reliability and Statistical Studies*. https://doi.org/10.13052/jrss0974-8024.1313

Pollaczek, F. (1930). Über eine aufgabe der wahrscheinlichkeitstheorie. i. *Mathematische Zeitschrift*, *32*, 64–100. https://doi.org/10.1007/BF01194620

The Health Foundation. (2023). *NHS waiting list to peak at more than 8 million by summer 2024*. https://www.health.org.uk/news-and-comment/news/nhs-waiting-list-to-peak-at-more-than-8-million-by-summer-2024

Virtamo, J. (2008). *38.3143 queueing theory / priority queues 1: Priority queues*. Online slides. https://www.academia.edu/4342667/J_Virtamo_38_3143_Queueing_Theory_Priority_queues_1_Priority_queues

Wiler, J. L., Bolandifar, E., Griffey, R. T., Poirier, R. F., & Olsen, T. (2013). An emergency department patient flow model based on queueing theory principles. *Academic Emergency Medicine*, *20*, 939–946. https://doi.org/10.1111/acem.12215

Yoon, P., Steiner, I., & Reinhardt, G. (2003). Analysis of factors influencing length of stay in the emergency department. *Canadian Journal of Emergency Medicine*, *5*(3), 155–161. https://doi.org/10.1017/S1481803500006539

Zachariasse, J. M., Hagen, V. van der, Seiger, N., Mackway-Jones, K., Veen, M. van, & Moll, H. A. (2019). Performance of triage systems in emergency care: A systematic review and meta-analysis. *BMJ Open*, *9*(5). https://doi.org/10.1136/bmjopen-2018-026471

Zhang, Y., Zhang, J., Tao, M., Shu, J., & Zhu, D. (2022). Forecasting patient arrivals at emergency department using calendar and meteorological information. *Applied Intelligence*, *52*, 11232–11243. https://doi.org/10.1007/s10489-021-03085-9