Fairness of Medical Triage Systems

Yunfei Kevin Wang
ECO317, Professor Gabriel Carroll
University of Toronto

April 19, 2024

Contents

1 Introduction

One of the highest-stakes situations of scarcity a society can face is limited healthcare resources in a medical emergency. During COVID-19, emergency rooms were flooded, so difficult decisions had to be made on which critical patients got ventilators and hospital beds (Jabakhanji, 2021). Public health agencies also faced a difficult question: Who should get the vaccine first? In war zones and disaster relief, the scarcity is even more pressing — often there are simply not enough resources to go around to treat everyone. Faced with such a pressing situation, it is important to act not only swiftly, but fairly. Since we cannot help everyone, on what ethical basis do some people get prioritized over others? We will scrutinize this using the economics approach to distributive justice.

2 Elements of a Triage Problem

2.1 Triage in Practice

Before we try to model this problem, it is important to understand how society already deals with distributing limited healthcare resources in an emergency. This problem has been so prevalent throughout history that we have a word for it: triage. At a high level, triage can be defined as a systematic prioritization of patients based on relevant information. The most common "relevant information" used to decide on prioritization are two factors: the severity of the condition and the resources needed to treat the condition, but one can think of others such as one's willingness and ability to pay for the resources.

2.1.1 Edwin Smith Papyrus

The earliest evidence of triage was found in the Edwin Smith Papyrus over 3000 years ago in Ancient Egypt (van Middendorp et al., 2010). In it, he divides patients into three categories:

1. "A medical condition I can heal"
2. "A medical condition I intend to fight with."
3. "A medical condition that cannot be healed."

In this triage, category 1 and 2 patients are prioritized over the 3rd category. While the relevant information used to define the 3rd category is unclear, presumably it was based on the difficulty of treating the condition which heavily depends on the available resources and severity of the condition. So, this historical triage system aligns with our definition.

2.1.2 Modern Emergency Department Triage

Worldwide, there are various triage systems used by hospitals, all aiming to have "patients receive care in an appropriate and timely manner... limit[ing] their injuries and their complications." (Yancey and O'Rourke, 2024) One common system used by emergency departments is the Emergency Severity Index (ESI) triage algorithm. It divides patients into the following:

- Level 1: Requires immediate life-saving interventions
- Level 2: High risk of deterioration into a life-threatening condition
- Level 3: Stable. Requires 2 or more hospital resources (i.e. lab test, X-rays)
- Level 4: Stable. Requires 1 hospital resource
- Level 5: Stable. No hospital resources needed besides basic medications

In this triage, patients seen are in order of their level, with level 1 being first, and level 5 being last. To categorize patients into levels 1 and 2, nurses check things like vital signs (pulse, breathing), responsiveness, etc. which fall under the severity of the condition in our definition. To categorize patients into levels 3, 4, and 5, nurses check the resources needed. So, this modern-day triage system also matches our definition. Interestingly, the severity of the condition is considered before the resources needed when deciding prioritization.

2.1.3 Modern Field and Disaster Triage

In a field and disaster triage (i.e. in the setting of a natural disaster or war), there are two widely popular triage algorithms, START, and SALT (Yancey and O'Rourke, 2024). The START triage divides patients into the well-known four armband color categories:

- Red (Immediate): Life-threatening injury
- Yellow (Delayed): Major injury, but can be delayed
- Green (Minimal): Minor injury
- Black (Expectant): Presumed dead

In the START triage, patients are treated in order of Red, Yellow, Green, and finally Black. Categorization is done quickly: Minimal (green) patients are determined first as the patients who can walk to an injury care site after everyone is instructed to. The remaining patients are determined using basic vital signs such as respiration and pulse. While this aligns with our definition of triage, notice only the severity of the condition is considered in this system.

The SALT triage is quite similar, with a key difference in the categorization of Expectant (black armband) patients. It asks the question: "Is the patient likely to survive the current circumstance given the resources available?" (Yancey and O'Rourke, 2024). If no, then the patient is categorized as Expectant. This means that patients who may not be dead can be denied treatment. Notice the SALT system considers both severity and resources needed.

2.2 Triage in a Theoretical Framework

Faced with so many different situations requiring triage and different approaches to triage, we attempt to boil down the problem to some high-level variables, structure, and fairness principles using some simplifying assumptions. As a starting point, we will use our definition of triage that we motivated using real-world examples. It can reformulated and simplified as: *A triage is a systematic prioritization of patients based on their condition's severity, and the resources needed to treat their condition.*

Admittedly, the definitions of severity and resources needed can refined further. For example, severity can be broken down into two dimensions: time-sensitive conditions that require immediate attention (i.e. choking), and life-threatening conditions that will likely kill you (i.e. cancer). Resources needed can be broken down into two dimensions: quantity needed (i.e. number of nurse hours), and type needed (i.e. common medications v.s. requiring specialized equipment). However, modeling this will be unwieldy, and we will argue that ignoring these distinctions still aligns with the Field and Disaster triage setting.

2.2.1 Emergency Department v.s. Field and Disaster Triage

While the definition of triage is widely applicable to both emergency departments and field and disaster triage, the underlying situation we need to model differs greatly.

For the emergency department setting, we have a somewhat predictable inflow of new patients, and the time-sensitivity of the condition is important due to the various kinds of patients visiting. Also, hospitals usually have enough resources to treat everyone eventually, so reducing wait times and having surge capacity are priorities. This strongly suggests that a multiple time-period model is needed to properly capture the inflow of new patients, how the severity of the condition evolves over time, the cost of wait times, and unexpected surges. Also, the types of medical resources are much more varied (medical tests, X-rays, surgeons available, etc.) so simply considering the quantity of resources may not be sufficient.

However, the Field and Disaster setting is simpler. Typically, there is a one-time appearance of new patients (i.e. immediately after an earthquake), and injuries are similar since they have the same cause, so time-sensitive and life-threatening conditions often go hand in hand. Thus, the time dimension is not as important, and the setting can be modeled as a one-off event. Also, the resources available tend to be very limited and basic compared to a hospital, so assuming one homogeneous medical resource is more justifiable. Thus, it seems like modeling the Field and Disaster setting is far more tractable.

2.2.2 Key Variables

Given our careful consideration and simplifying assumptions, these key variables capture the definition of triage in the Field and Disaster setting:

- N patients in need of treatment
- t the total number of medical resources available
- s_i the severity of the medical condition of patient i
- x_i the resources needed to cure patient i

We will assume a single period, so a triage algorithm will not necessarily prioritize which patient receives treatment first, but which patients receive treatment and which do not. Notice how this sketch of the key variables captures all aspects of our definition of triage.

One thing we will also abstract away from is the cost of obtaining the value of s_i, and the accuracy of the value. In reality, these are important considerations, as the START triage algorithm emphasizes speed over thoroughness, while the ESI is more thorough as it is in a hospital setting. Modelling the trade-off between obtaining better information (i.e. more testing and observation) and just going ahead with treating patients could be an interesting extension of the model. For our discussion, we assume that s_i is directly observed without any costs, and is correct (or at least an unbiased estimator of the true severity).

2.2.3 Fairness Principles

This setup has similarities with some well-known problems in economics. Each patient i needing a certain number of resources x_i is like the claims problem. Allocating medical resources to N patients is like the fair division problem. So we will borrow some ideas from claims problems and fair division problems for the following principles of a triage algorithm:

- Symmetry: The order of the patients listed in the problem should not matter.
- Pareto Efficient: No one can be treated without foregoing someone else's treatment.
- Envy-Free: After resources are distributed, someone who started "better off" should not wish they were someone who started "worse off".

Finally, to make a judgment on which outcome for the patients is best over all possible distributions of resources, we will borrow ideas from cardinal welfare, namely the collective utility function. We now formalize these key variables and fairness principles.

3 The Simple Triage Problem

3.1 Basic Definitions

Definition 3.1. *A* ***Simple Triage Problem*** *has* $N = \{1, 2, ..., n\}$ *patients, total number of resources* $t > 0$*, each patient has severity* $s_i \in (0, 1]$*, and resources needed* $x_i > 0$*, where* $\sum_{i=1}^{n} x_i > t$*. Such a problem can be represented by* $(t; (s_1, s_2, ..., s_n), (x_1, x_2, ..., x_n))$

This definition summarizes our discussion in 2.2.2 Key Variables. The severity s_i has the natural interpretation as the probability of death of patient i without treatment. For example, $s_i = 0.6$ means that the patient has a 60% chance of dying. The condition that $\sum_{i=1}^{n} x_i > t$ means more resources are needed than available, like a claims problem.

Definition 3.2. *A* ***Solution to a Simple Triage Problem*** *is a rule that assigns for any Simple Triage Problem* $(t; (s_1, s_2, ..., s_n), (x_1, x_2, ..., x_n))$*, a set of cured patients* $C \subset \{1, 2, ..., n\}$ *such that* $\sum_{i \in C} x_i \leq t$*.*

Definition 3.3. *The* ***Outcome*** *of a Solution to a Simple Triage Problem, is the list* $O = (s'_1, s'_2, ..., s'_n)$ *where* $s'_i = s_i$ *for all* $i \notin C$*, and* $s'_i = 0$ *for all* $i \in C$*.*

Note that a Solution to a Simple Triage Problem does not only require choosing a set of cured patients C for one problem but a rule for any Simple Triage Problem. However, the Outcome of a Solution to a Simple Triage Problem corresponds to applying that solution to a *specific* Simple Triage Problem. This outcome can be thought of as the actual mortality probabilities of the patients after curing the patients prescribed by the solution.

One important implication of this definition is that one cannot partially cure a patient, one must either completely cure the patient by fulfilling their x_i, or leave them alone. While this may seem like a strong assumption, in some medical situations, the full course of treatment is required for positive results. One may also conceivably extend this model to include a treatment function $s'_i = f(s_i, x_i, y_i)$ such as $f(s_i, x_i, y_i) = s_i - s_i \frac{y_i}{x_i}$, where y_i is the resources given to patient i. However, we will focus our attention on this simpler model.

3.2 Axioms of Solutions

With these definitions, we formally define the Fairness Principles we introduced in 2.2.3.

For the following, let $(t; (s_1, s_2, ..., s_n), (x_1, x_2, ..., x_n))$ be any Simple Triage Problem. We have a Solution to a Simple Triage Problem that assigns C to this problem, resulting in Outcome O. A Solution to a Simple Triage Problem can satisfy the following axioms:

Definition 3.4 (Pareto Efficiency Axiom)**.** *For all $j \notin C$, $\sum_{i \in C} x_i + x_j > t$.*

This axiom means that under the solution, one cannot treat another person without exceeding the available resources. Thus, no patient can be made better off without making another patient worse off, which is Pareto Efficiency in the usual sense.

Definition 3.5 (Symmetry Axiom)**.** *For any new problem defined by reordering $(s_1, s_2, ..., s_n)$ and $(x_1, x_2, ..., x_n)$ in the same way, the solution C applied to this new problem yields outcome $O = (s'_1, s'_2, ..., s'_n)$ reordered in the same way.*

This axiom means that under the solution, the order in which the patients are listed does not affect the patients' outcomes in reality. Like claims problems, this implies an axiom like Equal Treatment of Equals. The idea of an Envy-Free Test generalizes in two ways:

Definition 3.6 (Equal Severity Envy-Free Test)**.** *Assume $s_i = s_j$. If $x_i < x_j$, then $s'_i \leq s'_j$*

This axiom means that under the solution, if two patients have the same severity, but patient i requires fewer resources than patient j, then patient i's severity outcome should be at least as good as patient j's. One reason why we might want to satisfy this axiom is in situations where patients can demand more resources than needed. We do not want to incentivize patients to demand more medical resources in hopes of a better outcome.

Definition 3.7 (Equal Resource Envy-Free Test)**.** *Assume $x_i = x_j$. If $s_i < s_j$, then $s'_i \leq s'_j$*

This axiom means that under this solution, if two patients require the same resources, but patient i's severity is lower than patient j, then patient i's severity outcome should be at least as good as patient j's. One reason why we might want to satisfy this axiom is in situations where patients can feign severity. We do not want to incentivize patients to pretend to be more sick than they actually are in hopes of a better outcome.

3.3 Solutions to A Simple Triage Problem

For the following section, let $(t; (s_1, s_2, ..., s_n), (x_1, x_2, ..., x_n))$ be any Simple Triage Problem. To find the best Solution to a Simple Triage Problem, one must naturally consider the Outcome of applying the Solution. But how does society decide which outcome is best? One approach is to use Collective Utility Functions over the possible Outcomes.

3.3.1 Utilitarian Solution

If our goal is to maximize the expected number of lives saved, then we should evaluate the outcome $O = (s'_1, s'_2, ..., s'_n)$ using the Utilitarian Collective Utility Function:

$$W_{util}(s'_1, s'_2, ..., s'_n) = -(s'_1 + s'_2 + ... + s'_n)$$

The negative is since higher severity s'_i is bad. The Utilitarian Solution prescribes:

$$C_{util} := \underset{C \subset \{1,...,n\}}{\arg\max}\ W_{util}(s'_1, s'_2, ..., s'_n)$$

However, this definition is quite cumbersome, as it requires one to check all possible subsets of patients to find the best one. Instead, we propose the following algorithm:

Definition 3.8 (Utilitarian Algorithm Solution)**.** *For each patient i, compute the cost-effectiveness ratio $r_i := \frac{x_i}{s_i}$. Order patients in a list from lowest to highest r_i (most cost-effective to least cost-effective), and add them to the cured patient set C in that order. If there are not enough resources to treat patient j, treat patient $j+1$, until exhausting the list.*

Proposition 3.1. *The Utilitarian Algorithm Solution is equivalent to the Utilitarian Solution under certain assumptions.*

We will not prove our propositions here, but a heuristic argument is that to maximize the expected number of lives saved, we need to put our resources into the most cost-effective cases. The assumptions we need are to handle edge cases where we do not have resources to cure patient j but do for some later patient $k > j$ in the list. For example, if one can partly cure patient j for a proportional reduction in their severity outcome s'_j then it is enough.

Proposition 3.2. *The Utilitarian Algorithm Solution satisfies the Pareto Efficiency Axiom, Symmetry Axiom (under conditions), and Equal Severity Envy-Free Test. It violates the Equal Resource Envy-Free Test.*

The Pareto Efficiency follows since the Utilitarian Algorithm Solution exhausts the list of patients. Symmetry follows since we order the list by cost-effectiveness ratio r_i so the original order of the patients in the problem does not matter (except in the case of two patients having the same r_i's, which the conditions can handle). Equal Severity Envy-Free follows since if $x_i < x_j$, then $r_i < r_j$, which means i has priority over j to be cured, so $s'_i \leq s'_j$. There are many counter-examples to violate Equal Resource Envy-free.

3.3.2 Egalitarian Solution

If our goal is to equalize the chances of everyone's survival as much as possible, then we should evaluate the set $O = (s'_1, s'_2, ..., s'_n)$ using the Leximin Collective Utility Function:

$$W_{egal}(s'_1, s'_2, ..., s'_n) = \text{Leximin}(-s'_1, -s'_2, ..., -s'_n)$$

Again, the negative is since higher severity s'_i is bad. The Egalitarian Solution prescribes:

$$C_{egal} := \underset{C \subset \{1,...,n\}}{\arg\max}\ W_{egal}(s'_1, s'_2, ..., s'_n)$$

Again, we propose an equivalent algorithm that is easier to compute:

Definition 3.9 (Egalitarian Algorithm Solution)**.** *Order patients in a list from highest to lowest s_i (most severe condition to least severe), and add them to the cured patient set C in that order. If there are not enough resources to cure patient j, cure patient $j+1$, until exhausting the list.*

The justification for each part of the following proposition is similar to Proposition 3.1 and 3.2, or just very straightforward, so its argument is omitted for the sake of brevity.

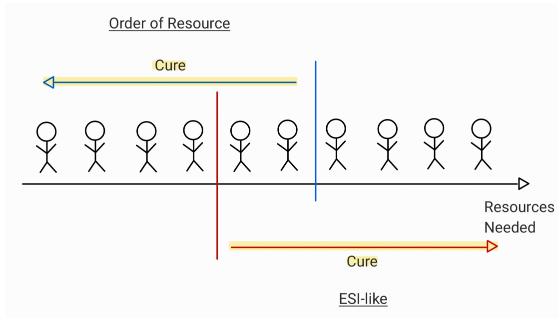
Proposition 3.3. *The Egalitarian Algorithm Solution is equivalent to the Egalitarian Solution. The algorithm satisfies the Pareto Efficiency Axiom, Symmetry Axiom, and Equal Severity Envy-Free Test (under conditions), and violates the Equal-Resource Envy-Free Test.*

3.3.3 Other Algorithms

Some other algorithmic solutions that are not derived from a Collective Utility Function are also reasonable. Similar to the Utilitarian and Egalitarian algorithms, the following algorithms will order the patients in a list and treat them in that order. Assume that once there are not enough resources to treat patient j in the list, then there are not enough resources to treat all of the remaining patients $j+1, j+2, ..., n$ (i.e. the Pareto Axiom is automatically satisfied). While this assumption can be easily forgone by using a similar method as the previous two algorithms, this assumption will simplify our definitions and yield nice graphs to analyze.

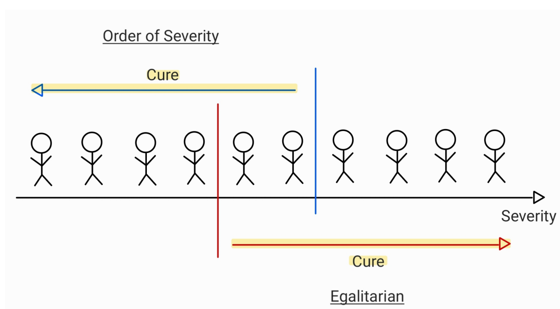
Definition 3.10 (Order of Resource Algorithm Solution)**.** *Order patients in a list from lowest to highest x_i, and add them to cured patient set C in that order until resources deplete.*

The picture below summarizes how the Order of Resource Algorithm works. The ESI-Like algorithm name is inspired by the Emergency Severity Index algorithm presented in 2.1.2, where level 3-5 patients were treated in the order of whoever needed the most resources. The ESI-Like Algorithm is simply Definition 3.10 but the list is from highest to lowest x_i.



Definition 3.11 (Order of Severity Algorithm Solution). *Order patients in a list from lowest to highest s_i, and add them to cured patient set C in that order until resources deplete.*

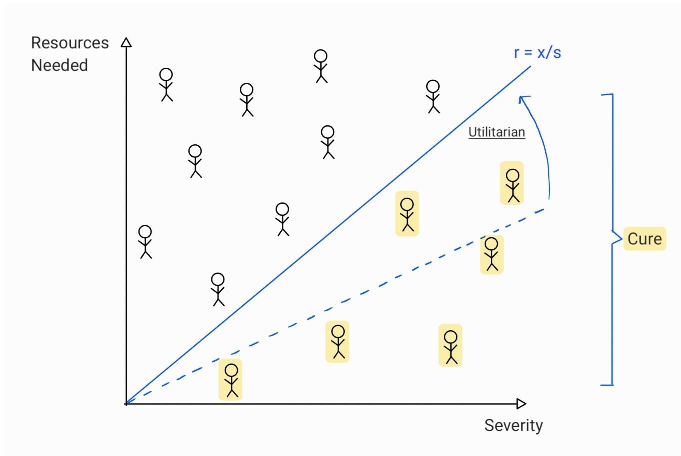
The picture below summarizes how the Order of Severity Algorithm works. Notice how the Egalitarian Algorithm is in some sense the opposite of the Order of Severity Algorithm. Also, the way the past two diagrams are presented is suggestive of a new perspective. For example, the Order of Severity can be interpreted as we first determine a threshold of severity (vertical blue line) such that we have the resources to cure everyone with severity below the threshold. This is reminiscent of the SALT Triage introduced in 2.1.3, where the Expectant (black armband) were not treated because their injury was too severe given the resources available.



We have the following proposition, and the argument is similar to previous propositions.

Proposition 3.4. *The Order of Resource and Order of Severity Algorithm Solutions satisfies the Pareto Efficiency Axiom, Symmetry Axiom, Equal Severity Envy-Free Test, and the Equal-Resource Envy-Free Test (under conditions).*

Is there a way to somehow combine the Order of Resource and Order of Severity Algorithms? Yes, there is: we can arrange the patients in 2D space and find some kind of line to split the patients into two groups. One natural line to decide on is the cost-effectiveness of treating a patient ($r = \frac{x}{s}$), and this is simply the Utilitarian Algorithm as shown below.



3.3.4 Resource-Severity Correlation Property

One implicit assumption we have been making throughout our discussion is that there is no relationship between the severity and resources needed. However, in practice, almost certainly these two variables are positively correlated (more severe conditions require more resources). So let us assume that these two variables are related by the following equation:

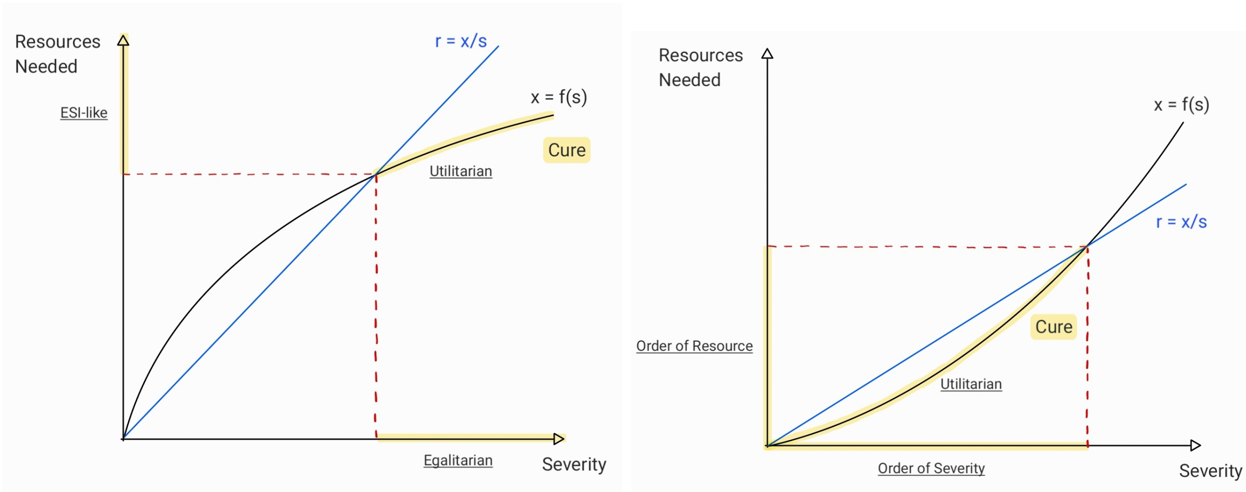
$$x_i = f(s_i) + \varepsilon_i$$

where f is an increasing continuous function, ε_i is some Gaussian noise. If $\varepsilon \equiv 0$, we have an interesting result connecting all parts of our discussion so far into the following proposition.

Proposition 3.5 (Resource-Severity Correlation Property). *Consider a Simple Triage Problem, where $x_i = f(s_i)$ for some increasing continuous function f. Under certain conditions*[1], *the following are true: If f is concave, then the Utilitarian Algorithm is equivalent to the ESI-like and Egalitarian Algorithms. If f is convex, then the Utilitarian Algorithm is equivalent to the Order of Resource and Order of Severity Algorithms.*

This proposition is demonstrated in the graph below. Because of the strong $\varepsilon \equiv 0$ assumption, it may seem like this property is of limited practical use. However, relaxing the assumption would only weaken the equivalencies to be "roughly equivalent". This result implies that as long as we can estimate the concavity of the correlation between the x_i and s_i the Utilitarian Algorithm which requires two pieces of information x_i and s_i, can be approximated by only considering only one of x_i or s_i that the Egalitarian, ESI-like, Order of Severity, and Order of Resource Algorithms require.

[1]We require conditions to address the edge case of satisfying the Pareto Efficiency Axiom as usual.



4 Conclusion

So, we will now use all our propositions to deduce what triage system is fairest. Firstly, by Propositions 3.1 and 3.3, we can reduce our list of candidates of formally defined solutions to four: the Utilitarian, Egalitarian, Order of Severity, and Order of Resource Algorithms.

If we were to require our solution to satisfy all our axioms (Definitions 3.4 - 3.7), then only the Order of Resource and Order of Severity remain by Proposition 3.4. But something is unsettling with these two approaches: we are only curing those who are the most well off to begin with, either in terms of the least resources needed or in the least severe condition.

So perhaps we should be more Egalitarian in our approach. However, the Egalitarian Algorithm falls flat in the following scenario. Suppose there was one person who certainly going to die and one million people who all had a ten percent chance of dying. We can use all our resources to save one person, or divide it up and protect the one million. The Egalitarian Algorithm will require us to save one person for certain, and let the one million people take the gamble, almost certainly resulting in more than one death.

So what about being Utilitarian? Proposition 3.2. tells us that it violates the Equal Resource Envy-Free Test, but recall one reason why we might care about this axiom is to prevent patients from pretending to be sicker to receive a better outcome. If this is not a concern, then saving the most number of lives possible is an easy idea to get behind, but must be applied carefully since a common criticism of utilitarianism in the medical context is that can be discriminatory (Okorie, 2019). In emergencies with limited time, where it is costly to assess both the patients' severity and the resources needed, we then have Proposition 3.5. If the correlation between severity and resources needed is convex, then indeed we can just use the Order of Severity or Order of Resource methods to achieve the Utilitarian outcome.

References

Jabakhanji, S. (2021, January). ICU doctor calls triage protocol 'morally distressing' as province sees continued stress on care units. *CBC News*.

Okorie, N. (2019). Partiality, impartiality and the ethics of triage. *Developing World Bioethics 19*(2), 76–85. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/dewb.12201.

van Middendorp, J. J., G. M. Sanchez, and A. L. Burridge (2010, November). The Edwin Smith papyrus: a clinical reappraisal of the oldest known document on spinal injuries. *European Spine Journal 19*(11), 1815–1823.

Yancey, C. C. and M. C. O'Rourke (2024). Emergency Department Triage. In *StatPearls*. Treasure Island (FL): StatPearls Publishing.