Polarized Citizen Preferences for the Ethical Allocation of Scarce Medical Resources in 20 Countries

Edmond Awad, Bence Bago, Jean-François Bonnefon, Nicholas A. Christakis, Iyad Rahwan, and Azim Shariff

Abstract
Objective. When medical resources are scarce, clinicians must make difficult triage decisions. When these decisions affect public trust and morale, as was the case during the COVID-19 pandemic, experts will benefit from knowing which triage metrics have citizen support. Design. We conducted an online survey in 20 countries, comparing support for 5 common metrics (prognosis, age, quality of life, past and future contribution as a health care worker) to a benchmark consisting of support for 2 no-triage mechanisms (first-come-first-served and random allocation). Results. We surveyed nationally representative samples of 1000 citizens in each of Brazil, France, Japan, and the United States and also self-selected samples from 20 countries (total N = 7599) obtained through a citizen science website (the Moral Machine). We computed the support for each metric by comparing its usability to the usability of the 2 no-triage mechanisms. We further analyzed the polarizing nature of each metric by considering its usability among participants who had a preference for no triage. In all countries, preferences were polarized, with the 2 largest groups preferring either no triage or extensive triage using all metrics. Prognosis was the least controversial metric. There was little support for giving priority to healthcare workers. Conclusions. It will be difficult to define triage guidelines that elicit public trust and approval. Given the importance of prognosis in triage protocols, it is reassuring that it is the least controversial metric. Experts will need to prepare strong arguments for other metrics if they wish to preserve public trust and morale during health crises.

Highlights
- We collected citizen preferences regarding triage decisions about scarce medical resources from 20 countries.
- We find that citizen preferences are universally polarized.
- Citizens either prefer no triage (random allocation or first-come-first served) or extensive triage using all common triage metrics, with “prognosis” being the least controversial.
- Experts will need to prepare strong arguments to preserve or elicit public trust in triage decisions.

Keywords
cross-cultural study, medical ethics, triage preferences

Date received: May 31, 2022; accepted: June 14, 2022

Corresponding Author
Azim Shariff, Department of Psychology, University of British Columbia, 2136 West Mall, University of British Columbia, 2136 West Mall, Vancouver, BC V7T1Z4, Canada (shariff@psych.ubc.ca).
During the COVID-19 pandemic, many countries anticipated a scarcity of essential medical resources, such as the ventilators required for patients with the most severe respiratory conditions. In a scarcity context, clinicians and other decision makers need to make difficult triage decisions about how to prioritize treatment. Such decisions impose a heavy practical and emotional burden on clinicians and may expose them to liability risks. To alleviate that burden and those risks, clinicians need guidance, ideally that has wide public support. Defining such guidelines has proven to be a formidable challenge. There is a substantial heterogeneity between (and sometimes within) countries with respect to official governmental guidelines about which metrics are deemed acceptable as bases for triage and how these various metrics should be prioritized. This heterogeneity is perhaps unavoidable given the difficulty of triage decisions and the ethical controversies raised by any single metric proposed as a guide (Table 1). Moreover, with these controversies comes a risk of social discord when citizens can imagine themselves and their relatives as being potentially affected by a life-and-death triage decision. The mere existence of a triage protocol is enough to provoke people’s discomfort, triggering discussions of “death panels” or physicians inappropriately “playing God.” If the metrics are perceived as biased or unfair, then trust and morale (among both providers and citizens) are likely to suffer even more.

In sum, to relieve clinicians of the practical and emotional burden of triage decisions, and to mitigate their liability risk, policy makers and their advisors must provide clear guidelines about the allocation of scarce resources, especially when this allocation is the focus of intense public attention. In turn, policy makers and their

<table>
<thead>
<tr>
<th>Triage Metric</th>
<th>Rationale</th>
<th>Controversies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prognosis</td>
<td>Prioritize patients who have better odds to survive treatment or better odds to survive on a longer term</td>
<td>Short-term prognosis is consensual, but controversies arise when considering long-term survival, which can be affected by comorbidities unrelated to the probability of surviving treatment, especially when these comorbidities are more frequent in patients from disadvantaged backgrounds.</td>
</tr>
<tr>
<td>Quality of life</td>
<td>Prioritize patients without comorbidities that may affect quality of life after surviving the disease</td>
<td>Deprioritizing patients with impaired physical ability, dementia, cerebral damage, or yet other conditions could breach the ethics of nondiscrimination.</td>
</tr>
<tr>
<td>Age</td>
<td>Maximize saved life-years (or opportunities to experience life stages) among patients with a similar prognosis</td>
<td>Deprioritizing older patients solely because of their age may breach the ethics of nondiscrimination, especially when an age cutoff is defined as an exclusion criterion.</td>
</tr>
<tr>
<td>Social value (past)</td>
<td>Prioritize health care workers who contracted the disease in the line of duty</td>
<td>Many guidelines prohibit the use of social value, only to make an exception for health care workers, which may seem unfair to other key workers.</td>
</tr>
<tr>
<td>Social value (future)</td>
<td>Prioritize health care workers to preserve their ability to fight the disease in the future</td>
<td>It is not always clear whether health care workers can be back to work in a realistic time frame and whether the logic should be extended to other key workers.</td>
</tr>
</tbody>
</table>
advisors need to be informed about public preferences. We do not mean that policies should necessarily be aligned to public preferences or that public preferences can help improve the outcomes of the resource allocation process, as the public is much less informed than experts on the feasibility, ethics, and efficiency of this resource allocation. However, decision makers can use citizen preferences to improve the legitimacy of the resource allocation process and preserve public trust. First, they can prepare especially strong justification when they decide to apply priorities that are not well accepted by the public. Second, they can use citizen preferences as tiebreakers when experts see the merits of 2 choices but cannot decide which is best. For these purposes, they need to understand which triage metrics have citizen support. Here, we provide them with such data, collected in 20 countries, using the example of ventilator allocation, which became familiar to the general public in the context of the COVID-19 pandemic.

We now briefly introduce the 5 metrics on which we collected citizen support data. These 5 metrics were selected on the basis of their prevalence in official policy documents. We did not involve participants in the selection of these metrics, since we wanted to measure support for metrics that are the most likely to be considered by clinicians or policy makers, not for metrics that may be spontaneously imagined by citizens but impossible to implement due to medical or legal constraints.

Prognosis is often considered the highest priority and most sensible metric for triage.\textsuperscript{11} It amounts to allocating ventilators to patients who are the most likely to benefit from them. There is variation, however, in the time scale that is to be considered for this prognosis: short-term probability of surviving treatment, long-term survival, or even longer-term life span reductions due to comorbidities. Together with the fact that prognosis is often inexact, which complicates matters,\textsuperscript{13} this ambiguity can sometimes blur the line between the prognosis metric and the quality-of-life metric.\textsuperscript{14} Indeed, the quality-of-life metric considers comorbidities that can affect quality of life without being tied to the short-term probability of surviving treatment.\textsuperscript{15} For example, some guidelines lower the priority of patients suffering from impaired physical ability, dementia, or cerebral damage, which has alarmed disability advocates and fueled concerns that triage protocols may breach the ethics of nondiscrimination.\textsuperscript{16,17} For example, several disability advocacy organizations filed complaints requiring the state of Alabama to clarify its protocol and to clarify in particular whether Down syndrome patients would be given low priority for ventilation.

Many guidelines rule out the use of age to make triage decisions, although they allow it a role in determining prognosis. Other guidelines consider an age cutoff as an exclusion criterion (e.g., 85 y, or 75 y in case of increased scarcity of resources). Still, other guidelines consider age as a tiebreaker between patients with a similar prognosis, prioritizing younger patients to save more life-years or to increase opportunities to experience life stages. Because of this lack of consensus, the use of age as a triage metric is a sensitive issue from a political and psychological perspective.\textsuperscript{18} Social value is an even more contentious metric. Some guidelines explicitly forbid any consideration of a patient’s utility to society—only to make an exception for health care workers. Two justifications are advanced for giving a priority to health care workers: this priority can be a reward for their past contribution or personal risk taking in fighting a disease, or it can be justified instrumentally, as a way to preserve their future contribution in fighting the disease, thus benefiting others. Both justifications are controversial, which likely explain their disparate presence in official guidelines. Some guidelines give a high priority to health care workers, others use this metric as a tiebreaker between patients with a similar prognosis, and yet others mention this metric but do not give it a clear priority, suggesting that its use should reflect community values. All of this variation emphasizes the need to collect data on citizens’ willingness to see metrics in use.

Using a triage metric based on any such considerations amounts to giving it a higher priority than mechanisms that do not explicitly take into account the characteristics of the patients, such as first-come-first-served or random lottery. In the extreme case in which either first-come-first-served or random lottery is given top priority, all metrics are moot, and the allocation process uses no triage for the restricted set of the 5 metrics featured in our survey (henceforth abbreviated as “no triage”). In the opposite extreme case, where first-come-first-served and random lottery are given the lowest priority, any and all metrics can be included in the allocation process—a situation we will call “full triage for the restricted set of the 5 metrics featured in our survey,” henceforth abbreviated as “full triage.” In between these extremes, there are many possibilities for partial triage, using one of many possible subsets of metrics. Given the 5 metrics we consider here (prognosis, quality of life, age, past contribution, future contribution), there would be 32 such possible subsets. A first goal of our data collection is to identify, in each country, which of these subsets are preferred by which proportion of citizens. A second goal of our data collection is to identify, in each
country, the metrics that are the most controversial or consensual.

Methods

We collected data from 2 sources. First, we polled nationally representative panels of 1000 participants recruited by the YouGov company in each of Brazil, France, Japan, and the United States. The Supplementary Analysis (SA) file provides a detailed description of the polling process, and Table 2 displays the demographic characteristics of the 4 samples. These 4 countries were chosen because their citizens showed substantial differences in their responses to moral dilemmas in a previous survey. Accordingly, they offered good prospects to capture cultural differences in triage preferences, if any.

Second, we posted the same survey on the Moral Machine website (moralmachine.net). The Moral Machine is a highly popular citizen science website that was designed in 2016 to collect public preferences related to the moral dilemmas of self-driving cars. It receives a constant flow of visitors interested in contributing responses to moral dilemmas and thus offers a convenient way to collect data from participants worldwide. It does not, however, offer a representative sample of the population, and we discuss the impact of this self-selection in the “Results” section and in the SA. We retained for analysis the 20 countries with the largest samples (min N = 96 in Mexico, max N = 2153 in the United States [see the SA for a detailed description], for a total N of 7599). These 20 countries included the 4 countries in which we collected nationally representative samples, allowing for a comparison of the results obtained with the 2 data sources.

All participants received the same survey, regardless of whether they were recruited through the YouGov company or through the Moral Machine website. They were asked to rate the usability of 5 triage metrics (prognosis, quality of life, age, past contribution, future contribution) and 2 no-triage mechanisms (first-come-first-served, random lottery) on a 0 to 100 scale anchored at should not be considered and should be considered. Whenever we write about the “usability” of a triage metric or no-triage mechanism, we refer to this rating.

Here is the wording of each metric in the survey. Prognosis was described as “the chance of recovery (i.e., prioritize patients without any medical conditions that worsen their progress).” Age was described as “how many years of life they’re likely to have after the illness (i.e., prioritize younger patients).” Quality of life was described as “the likely physical quality of life after the illness (i.e., prioritize patients without any medical conditions that would reduce quality of life after COVID-19 resolves).” Past contribution was described as “whether they’ve made sacrifices helping with the virus (e.g., prioritize medical professionals and research participants who’ve put their lives at risk).” Future contribution was described as “whether they might help with the virus in the future (e.g., prioritize medical professionals and students).” The 2 no-triage mechanisms were worded as follows: first-come first-served was described as “when they arrived at the hospital (i.e., prioritize patients who were first in line),” and random lottery was described as “ventilators should be allocated by random lottery (i.e., individual characteristics not considered).”

The full text of the survey is available in the SA. For the 4 representative panels recruited through YouGov, the survey was presented in the official language of the country (i.e., Brazilian Portuguese, English, French, and Japanese). For the Moral Machine samples, the survey was available in 10 languages from which people could choose: Arabic, Chinese, English, French, German, Japanese, Korean, Portuguese, Russian, and Spanish.

Table 2 Demographic Description of the 4 National Samples

<table>
<thead>
<tr>
<th>Country</th>
<th>N</th>
<th>Male (%)</th>
<th>Age (SD)</th>
<th>Smoker100 (%)</th>
<th>College (%)</th>
<th>Conservatives (%)</th>
<th>Religious (%)</th>
<th>White (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRA</td>
<td>1000</td>
<td>49</td>
<td>36.2 (12.7)</td>
<td>54</td>
<td>20</td>
<td>41.1</td>
<td>30.5</td>
<td>64.2</td>
</tr>
<tr>
<td>FRA</td>
<td>1000</td>
<td>49</td>
<td>47.8 (17.1)</td>
<td>21</td>
<td>56</td>
<td>36.3</td>
<td>69.8</td>
<td>69.8</td>
</tr>
<tr>
<td>JPN</td>
<td>1000</td>
<td>54</td>
<td>49.5 (15.6)</td>
<td>3</td>
<td>44</td>
<td>36.3</td>
<td>69.8</td>
<td>69.8</td>
</tr>
<tr>
<td>USA</td>
<td>1000</td>
<td>46</td>
<td>48.8 (17.4)</td>
<td>25</td>
<td>46</td>
<td>36.3</td>
<td>69.8</td>
<td>69.8</td>
</tr>
</tbody>
</table>

*YouGov offers different default demographic packages in all 4 countries. Hence, the recorded demographic characteristics of the samples differ among the surveyed countries. N, number of participants; male, percentage of males; age, mean age in years, with standard deviations in parentheses; know COVID patient, percentage of participants who reported to have known a COVID patient at the time of responding; smoker100, percentage of participants who smoked at least 100 cigarettes in their entire life; college, percentage of participants graduating from college; conservatives, percentage of participants reported being conservative; religious, percentage of participants reported to be religious; White, percentage of participants who reported to be White.*
We consider that a participant “accepts” the use of a triage metric if they rate the usability of this metric higher than the usability of both random lottery and first-come-first-served. With this definition, we can look at the set of triage metrics that are most likely to be accepted by citizens within each country. Ethical approval was obtained by local Ethics Committees (approval Nos. H20-01190 and A 2020-12; specific information about the committees is hidden for the purpose of double blind peer review). Before the experiments, we initially ran a pilot on April 23, 2020, of 200 people using a nonrepresentative sample of American residents via the Prolific survey platform. This data were excluded from the analysis.

**Results**

In all 4 countries with representative samples, the 2 largest groups of participants expressed preferences for the same polarized sets of metrics (Figure 1A), namely, either no triage at all or using the full complement of metrics (full triage). A plurality of participants (23%–35% of the sample) did not rate any triage metric higher than random lottery or first-come-first-served, which would correspond to a preference for no triage. The preference for no triage was driven by high ratings of the first-come-first-served mechanism (range = 50–66) more than by high ratings of the random mechanism (range = 13–19, which is lower than the first-come-first served mechanism for 81%–87% participants across countries). The second largest group of participants (12%–23% of the sample) rated every metric higher than both random lottery and first-come-first-served, indicating a preference for full triage. The no triage and full triage groups together account for about half the sample in each country, and they are always significantly larger than the third largest group (all P values lower than 0.001; see SA).

Although the Moral Machine participants self-selected into the survey (rather than being recruited as representative samples), their responses are strikingly similar. The results displayed in Figure 1A are replicated in all 20 countries from the moral machine data set (Figure 1B). In every country, the 2 largest groups of participants expressed preferences consistent with either no triage or full triage (the third largest group is significantly smaller than both these groups in 17 countries out of 20; see SA). Once more, the preference for no triage was driven by high ratings of the first-come-first-served mechanism (range = 47–63) more than by high ratings of the random mechanism (range = 9–18, lower than the first-come-first served mechanism for 69%–84% participants across countries).

While the no triage and full triage groups are always the largest, their respective sizes vary across countries. We could consider country-level correlates of this variation, but we need to exercise caution when interpreting such correlations. Consider, for example, Figure 1C, which shows a significant negative correlation between the size of the no triage group and the COVID-19 death rate per million across countries at the time of data collection ($r = -0.67, P = 0.002$). It would be tempting to think that citizens of countries that are hit the hardest are more likely to realize the necessity of triage, but the correlation alone does not offer support for this causal claim. In fact, we did not find cross-sectional evidence in our data that the size of the no triage group in the United States tracked the progression of the epidemics across time nor the death rate across states. Accordingly, the correlation between low death rates and rejection of triage presumably reflects an association with a third variable, perhaps the fluidity with which people make new social connections, which is known to be negatively correlated with the spread of COVID-19 and with the acceptance of utilitarian solutions to moral dilemmas.

The demographic breakdown of the no triage and full triage groups is not consistent across countries. We discuss these matters in detail in the SA, but to give one example, a conservative ideology is significantly associated with a preference for no triage in the United States, but the effect goes in the opposite direction in France. Overall, it would seem that triage preferences do not neatly line up with demographic characteristics. They may instead reflect idiosyncratic dispositions for outcome-based versus quality- or communitarian-based ethics, which would make it harder to find a middle ground, that is, a set of triage metrics that would be reasonably acceptable for a majority of citizens. We now consider how our data could inform this reconciliation effort.

We can explore the extent to which each triage metric may reconcile citizens with polarized preferences. A perfectly consensual metric would be accepted by all participants, with an average usability rating of 100/100. Obviously, no metric passes this test in our data, but we can check how close each metric is to this ideal. A metric is closer to the ideal potential for reconciliation when it is rejected by fewer participants and when these same participants who reject it still rate its usability reasonably high. Figure 2 offers a visualization of the potential of each metric according to these criteria.

Figure 2 suggests that the metric with the best potential is prognosis (see the SA for detailed results).
Qualitatively speaking, it is accepted by most participants in 2 nationally representative samples. In parallel, prognosis is accepted by most participants in 17 of the 20 Moral Machine samples. However, this majority is statistically significant in only 5 samples, and we must be careful when interpreting these data, since Moral Machine participants rate the prognosis metric 3 to 11 points higher than participants from nationally representative samples.
The prognosis metric has the highest acceptance rates of all metrics, and its usability remains high even among respondents who reject its use. In our nationally representative samples, Japanese and French participants who reject the use of prognosis still rate its usability significantly higher than the midpoint of the scale. The same is true in 17 of 20 countries in the Moral Machine samples, although the comparison is statistically significant in only 5 countries.

In contrast, the social value of patients for fighting the pandemic (reflected in their past or future contributions) would seem to be a sensitive topic in many countries, with many participants feeling moderately to strongly against its use as a triage metric. This may lead to tensions between health care workers and the rest of the population. In any case, the country-level variations observed in Figure 2 suggest that while the polarization of citizen preferences was universal in our surveys, the process of reconciling these polarized preferences should be country specific, as different metrics have different reconciliation potentials in different countries.

**Figure 2** Potential for reconciliation, by metric and country. Proportion of participants who reject each triage metric by country, together with its usability rate among these same respondents, in (A) nationally representative samples and (B) self-selected samples from the Moral Machine website.

The prognosis metric has the highest acceptance rates of all metrics, and its usability remains high even among respondents who reject its use. In our nationally representative samples, Japanese and French participants who reject the use of prognosis still rate its usability significantly higher than the midpoint of the scale. The same is true in 17 of 20 countries in the Moral Machine samples, although the comparison is statistically significant in only 5 countries.

In contrast, the social value of patients for fighting the pandemic (reflected in their past or future contributions) would seem to be a sensitive topic in many countries, with many participants feeling moderately to strongly against its use as a triage metric. This may lead to tensions between health care workers and the rest of the population. In any case, the country-level variations observed in Figure 2 suggest that while the polarization of citizen preferences was universal in our surveys, the process of reconciling these polarized preferences should be country specific, as different metrics have different reconciliation potentials in different countries.

**Discussion**

When health care infrastructures are strained to the point of scarcity, clinicians or other allocators need clear official guidelines to make difficult triage decisions. If these triage decisions are of wide relevance, as they were during the COVID-19 pandemic, it can be important to know which triage metrics citizens consider as acceptable. Our data show that it will be hard to find a consensual set of metrics. In all 20 countries that we polled, public opinion was strikingly polarized between citizens who would prefer no triage, on one hand, and citizens who would accept an extensive triage based on prognosis, age, expected quality of life, and prioritization of health care workers, on the other hand.

This polarization is ubiquitous despite the cultural differences between the countries we polled and the different stages of their epidemics when the data were collected. It emphasizes the challenge that experts will face if they seek to establish public trust for triage protocols.

But this challenge is not insurmountable. For example, it would be very unlikely for experts to renounce the use of prognosis as a triage metric. But, fortunately, our data show that there would be reasonable support for the use of this metric, even among citizens who would prefer no triage. We do not mean that prognosis should take precedence in the absence of a consensus view—we mean that even when there is no consensus on an acceptable set of metrics, some individual metrics are more polarizing than others. Experts who have to set priorities can therefore identify the metrics for which they need to prepare an especially strong justification.
Prognosis, the least polarizing metric, may not need an especially strong justification. In contrast, if experts decide to use more polarizing metrics such as age (e.g., as an exclusion criterion) or to give priority to health care workers, they should prepare a careful argument for this decision. Psychological research may be useful in this respect. For example, it appears that using a veil-of-ignorance argument (e.g., asking citizens to judge whether a metric is acceptable while ignoring their personal characteristics, such as their own age) can increase their approval of decisions that favor the greater good.28

While we focused on most commonly discussed triage metrics, future research may consider other metrics, such as pregnancy or the existence of dependents or social connections.27,29 For exploratory purposes, we collected data on the ability to pay for treatment. Perhaps unsurprisingly, this metric was very unpopular (detailed results are presented in the SA). Future research may also consider other triage decisions. While we focused on ventilator allocation, an even more difficult decision is to reallocate a ventilator from a current patient to an incoming patient. We collected data on such decisions (results are presented in the SA) and found that, although the usability rating of all metrics decreased, all our findings were by and large reproduced for reallocations.30

Finally, while we focused on ventilator allocation as an easily understood and illustrative example of a scarce medical resource, public preferences may show a different pattern for different types of resources. It will be especially useful to compare our data to data on public preferences for the allocations of COVID-19 vaccines, which continue to be a scarce resource in most of the world.31,32 The priorities for resources meant to prevent illness (e.g., vaccines) rather than to treat it (e.g., ventilation) differ in ways that may not be obvious to the general public.33,34 Thus, it will be important to assess whether citizens’ preferences are sensitive to these different priorities, as well as the extent to which they correlate across nations. Making allocation decisions and public preferences transparent can help because dividing resources in the open is wiser, more just, and more acceptable than dividing them in secret.

Ethics Statement

The study obtained ethical approval from the Research Ethics Board at the University of British Columbia (approval ID: H20-01190) and the Ethics Committee at Max Planck Institute for Human Development (approval ID: A 2020-12). Participants gave informed consent either before (YouGov) or after (Moral Machine) the survey (see details in the Supplementary Appendix).

ORCID iD
Bence Bago https://orcid.org/0000-0001-6905-1832

Supplemental Material

Supplementary material for this article is available on the MDM Policy & Practice website at https://journals.sagepub.com/home/mpp.

Data Availability Statement

Data and code are open and available at https://github.com/bencebago/ventilators.

References


