



Policy Analysis

Organ Allocation Algorithms and the Problem of Algorithmic Fairness

Sofia Benitez^{1*}, Mette Novak² and Claire Muller³

¹Division of Clinical Ethics, Alpine Institute of Health Policy, Austria

²Department of Bioethics, Alpine Institute of Health Policy, Austria

³Department of Public Health Ethics, Baltic Institute of Medical Humanities, Sweden

ABSTRACT

Organ allocation systems have always required a formula for deciding who receives a scarce organ, but the introduction of machine-learning tools alongside established scoring systems such as the Kidney Allocation System and the Model for End-Stage Liver Disease score has made the ethical basis of that formula considerably harder to inspect.

This policy analysis uses the 2021 removal of a race-based coefficient from the estimated glomerular filtration rate equation as a case study in how a supposedly neutral clinical variable can embed unexamined assumptions that delay transplant access for specific patient groups.

It then examines the growing use of machine-learning models in liver and kidney allocation, the arguments for and against requiring explainability in these tools, and the accountability problems that arise when a continuously retrained model replaces a publicly reviewable scoring rule.

We argue that formal accuracy and fairness are distinct targets, that algorithmic reform addresses only one stage of a longer pathway of unequal access to transplantation, and that meaningful progress requires publicly auditable model documentation alongside, not instead of, attention to upstream disparities in referral and care.

The paper concludes that no allocation algorithm, however well designed, can substitute for addressing the structural inequities that determine who reaches the transplant waiting list in the first place.

Keywords: algorithmic fairness; organ allocation; transplant ethics; machine learning; distributive justice

Citation: Benitez S, Novak M and Muller C (2023) Organ Allocation Algorithms and the Problem of Algorithmic Fairness. *Bioethics*. 1(2):01-05.

Corresponding Author: Sofia Benitez

E-mail Address: s.benitez@aihp.at

Article History: Received 09 June 2023 | Accepted 10 September 2023 | Published 10 December 2023

Copyright: © 2023 Sofia Benitez, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (<https://creativecommons.org/licenses/by/4.0/>).

INTRODUCTION

Every organ allocation system is, at bottom, a formula

for deciding who lives and who keeps waiting. For decades that formula was relatively legible: a handful of clinical variables, a waiting-list clock, and a scoring

rule that a transplant coordinator could explain to a patient in a few sentences. Kidney allocation weighed waiting time and HLA matching; liver allocation moved to the Model for End-Stage Liver Disease (MELD) score, a mortality-risk equation built from three or four lab values. These systems were never free of controversy, but their logic was inspectable [1]. What has changed over the past decade is the sophistication of the tools doing the ranking, and with it, the difficulty of answering a question that used to have a fairly direct answer: on what basis, exactly, was this patient placed above that one?

The current Kidney Allocation System (KAS) already illustrates how contested these formulas can be even without machine learning in the mix. It layers longevity matching, prior living-donor status, pediatric priority, and sensitization scoring into a single point total, and commentary on how to make the KAS fair has repeatedly noted that no single point-weighting scheme can simultaneously satisfy utility, equity, and urgency in full [2]. A recent review of the values embedded in U.S. organ allocation systems found that the priority rules for different organs do not even agree with each other about which ethical principles -- utility, equity, or urgency -- should dominate [3]. That inconsistency is worth pausing on, because it means "fairness" was never a single, settled target that allocation algorithms could simply be built to hit; it is a contested variable, and different organ systems have quietly picked different answers.

Two developments have pushed this from a background concern to an urgent one. First, the 2021 decision to remove race from the estimated glomerular filtration rate (eGFR) equation used to gauge kidney function -- and therefore transplant eligibility -- exposed how a supposedly neutral clinical variable had been embedding a racial correction factor that systematically delayed Black patients' access to listing [4,5]. Second, transplant programs are increasingly experimenting with machine-learning models for organ-quality assessment, donor-recipient matching, and post-transplant survival prediction, tools whose internal logic is often not fully inspectable even by the clinicians who deploy them [6]. Table 1 summarizes how the ethical stakes shift as allocation tools move from simple scoring rules toward opaque predictive models, and Figure 1 sketches the same progression schematically. This review examines what "algorithmic fairness" has come to mean in organ allocation, why the race-based eGFR correction became the field's clearest case study in unintended algorithmic harm, and what the push toward machine-learning-based allocation is likely to change -- and what it will not.

THE RACE-NEUTRAL EGFR CORRECTION AS A CASE STUDY

For years, the standard eGFR equations applied a coefficient that produced a higher (and therefore more reassuring) estimate of kidney function for patients

coded as Black, on the premise that average muscle mass and creatinine generation differed by race. Critics pointed out that this correction used race as a crude biological proxy, baked a population-level assumption into an individual patient's number, and had the practical effect of making some Black patients' true kidney function look better than it was -- delaying the point at which they crossed the threshold for transplant listing [5]. A national task force ultimately recommended removing the race coefficient altogether and adopting a race-free equation, and the Organ Procurement and Transplantation Network moved to implement it prospectively, including retroactively crediting waiting time to patients whose eligibility had been delayed by the old formula [4].

The aftermath has been genuinely instructive. Empirical work published since the change has found that removing race from the eGFR calculation measurably increased identification of Black patients as eligible for listing and modestly narrowed a longstanding access gap, though it did not close it [7,8]. That partial result matters for how the field thinks about algorithmic fairness more broadly: correcting one biased input variable improved outcomes without fully equalizing them, which suggests that the eGFR coefficient was one contributor to a disparity with several other causes -- referral patterns, insurance status, transportation access to transplant centers -- that a formula change alone cannot touch. Educational efforts built around this episode now use it explicitly to teach clinicians how ostensibly objective, quantitative tools can encode contestable social assumptions without anyone deciding, at any single point, to put them there [9]. The lesson generalizes uncomfortably well to the newer machine-learning tools now entering the same space: those, too, are trained on historical data that reflects the same referral and access patterns the eGFR correction helped perpetuate.

MACHINE LEARNING ENTERS ALLOCATION

Liver allocation has become the most active testing ground for machine-learning approaches to organ allocation, in part because post-transplant survival is a genuinely hard prediction problem that MELD-based scoring handles only crudely. A systematic review of AI and ML applications in liver allocation found a proliferating set of models aimed at donor-recipient matching, graft-survival prediction, and organ-quality triage, most still confined to retrospective validation rather than live deployment [10]. One recent proposal, FairAlloc, tries to build equity constraints directly into a reinforcement-learning allocation policy rather than treating fairness as a check applied after the fact to a model optimized purely for predicted survival [11] -- an explicit acknowledgment that an unconstrained accuracy-maximizing model will, left alone, tend to reproduce whatever biases exist in the training data on who historically received organs and did well.

The appeal of these models is straightforward: better prediction of graft and patient survival should, in principle, route organs to the transplants most likely to succeed, reducing wasted organs and improving population-level outcomes. The difficulty is that "most likely to succeed" is not an ethically neutral output. A model trained on historical outcomes will tend to rate as lower-probability-of-success exactly the patients who have historically been sicker at the time of transplant, had less access to post-transplant follow-up care, or belonged to demographic groups the training data underrepresents -- and it will do so with a statistical confidence that can look more objective than it is [6]. A recent framework paper argues that existing bioethical principles for organ allocation (utility, equity, urgency) were articulated for scoring systems simple enough that a clinician could trace a decision back to its inputs, and that applying those same principles to opaque predictive models requires new procedural safeguards that the field has not yet built [12].

EXPLAINABILITY, ACCOUNTABILITY, AND THE LIMITS OF "FAIRNESS"

A recurring argument in the broader algorithmic-fairness literature is that accuracy and fairness are not the same target, and that optimizing hard for one can come at the expense of the other. One influential formulation holds that procedural fairness in AI-based medical resource allocation requires explainability as a matter of principle, not merely as a technical nicety -- patients and clinicians are entitled to know the basis for a life-affecting allocation decision even when a black-box model outperforms an interpretable one on raw predictive accuracy [13]. Others take a more permissive view, arguing that black-box AI can be ethically justifiable in healthcare under specific conditions -- rigorous external validation, robust human oversight, and a plausible causal story for why the model works -- even without full interpretability [14].

Organ allocation is a particularly hard test case for this debate because the decision is discrete, irreversible, and zero-sum in a way that most clinical AI applications are not: routing an organ to one patient necessarily means not routing it to another. A widely cited framing of "bounded justice" in clinical algorithms more generally argues that fairness audits conducted only within the scope of a single model's outputs will systematically miss unfairness introduced upstream, in whose data was collected, whose outcomes were coded as successes, and which populations were adequately represented in training cohorts to begin with [15]. Applied to organ allocation, this suggests that even a perfectly calibrated, formally "fair" machine-learning allocation model could still reproduce the access disparities documented in the eGFR case, if the data it learned from already encodes decades of unequal referral and listing practices [16,17].

Accountability is a related but distinct problem. A

scoring formula like MELD can be defended or revised through an open rule-making process at the policy level; a proprietary or continuously-retrained machine-learning model resists that kind of static, publicly reviewable accountability, since its behavior can shift as new data accumulates without any single documented policy change [18]. Reviews of international deceased-donor kidney allocation algorithms have found substantial variation across countries in how much algorithmic detail is published for public and clinical scrutiny at all, which suggests transparency is currently a matter of jurisdictional choice rather than a settled professional norm [19]. A European human rights law analysis of algorithmic organ allocation goes further, arguing that opaque allocation criteria may run afoul of equal-access obligations under human rights instruments regardless of whether a given jurisdiction's transplant law formally requires algorithmic transparency, which suggests the accountability problem is not merely a matter of professional best practice but potentially of legal obligation [20].

DISTRIBUTIVE JUSTICE CONSTRAINTS THAT PREDATE THE ALGORITHMS

It is worth remembering that organ scarcity, not algorithmic design, is the underlying condition that makes every allocation choice ethically fraught in the first place; no formula, however carefully audited, resolves the basic fact that there are more people who need organs than organs available [21,22]. Classical accounts of organ allocation ethics distinguish medical utility (allocating to maximize expected benefit), equity (giving everyone a fair chance regardless of factors unrelated to medical need), and urgency (prioritizing those who will die soonest without an organ), and note that these principles frequently conflict rather than converge on a single ranking [23,24]. The point of naming these tensions explicitly is that they existed long before any allocation system used a neural network, and they will persist after machine-learning tools are more fully incorporated -- an algorithm cannot resolve a disagreement about which ethical principle should govern allocation, it can only encode whichever answer its designers gave it, often implicitly.

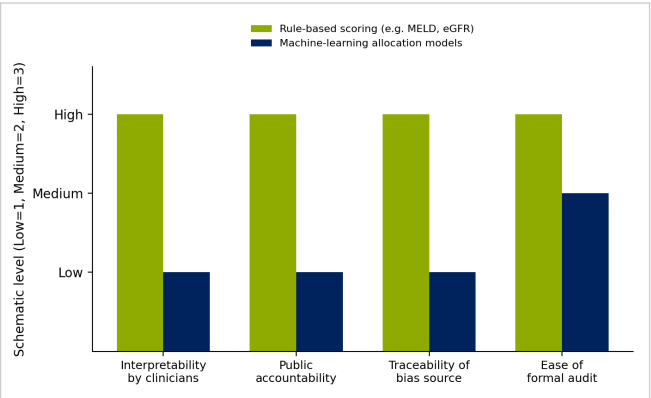
Longstanding disparities in referral for transplant evaluation, insurance coverage for post-transplant immunosuppression, and geographic access to transplant centers all predate and independently compound whatever a specific allocation formula does at the point of organ offer [25]. This matters directly for how much weight to put on algorithmic reform: fixing the allocation formula, however important, addresses only one stage of a much longer pathway along which access to transplantation narrows for disadvantaged patients. Table 1 situates the eGFR correction and the newer machine-learning proposals within that longer pathway, and Figure 1 shows schematically how the "location" of the fairness problem has shifted from a

single miscalibrated input variable toward a more diffuse and harder-to-audit set of model behaviors as allocation tools have grown more complex.

Table 1. Ethical principles commonly invoked in organ allocation, and how each is affected as allocation tools move from simple scoring rules toward machine-learning models.

Principle	Under rule-based scoring (e.g. MELD, KAS)	Under machine-learning allocation models	Main risk if unaddressed
Utility	Maximizes expected survival benefit using a small, published set of variables	Maximizes predicted survival using many variables learned from historical outcomes	Training data may embed access disparities as if they were medical differences
Equity	Explicit point adjustments for waiting time, sensitization, prior donation	Equity constraints must be deliberately engineered into the model objective	Unconstrained models tend to reproduce, not correct, historical inequities
Urgency	Directly encoded via mortality-risk score components	Can be modeled more finely, but weighting against utility is less transparent	Trade-off between urgency and predicted success becomes harder to audit
Accountability	Rule changes go through public policy review	Model behavior can drift with retraining without a discrete policy change	Loss of a clear, reviewable record of why a decision rule changed

Figure 1. Schematic comparison of rule-based scoring systems and machine-learning allocation models on interpretability, accountability, traceability of bias, and ease of formal audit (illustrative, not derived from a single empirical study).



CONCLUSION

The organ allocation field has learned, somewhat the hard way, that a formula's apparent objectivity is not evidence of its fairness. The race-based eGFR correction was simple enough, by the standards of a

modern predictive model, that its flaw could eventually be identified, debated in public, and corrected through an open policy process -- and even then, correcting it narrowed rather than eliminated a real disparity in access [7]. Machine-learning allocation tools raise the same underlying question -- what assumptions has this tool encoded about who should receive a scarce organ? -- while making the answer considerably harder to extract, because the tools are more complex, frequently proprietary, and trained on data that already reflects the disparities their advocates hope algorithmic sophistication will reduce. Genuine algorithmic fairness in this domain will likely require more than better mathematics: it needs publicly reviewable model documentation, external audits that check for disparate impact rather than only aggregate accuracy, and an honest acknowledgment that no allocation algorithm, however well designed, can substitute for addressing the upstream inequities in referral and access that shape who reaches the waiting list in the first place.

DECLARATIONS

Acknowledgements

The authors are grateful to colleagues in their respective bioethics and health policy institutes for helpful discussion during the drafting of this review.

Funding

This review received no external funding.

Conflict of Interest

The authors declare no conflicts of interest relevant to this work.

Author Contributions

Sofia Benitez conceived the review, coordinated the literature search, and drafted the initial manuscript. Mette Novak developed the analysis of the eGFR case study and the sections on machine-learning allocation models, and critically revised the manuscript. Claire Muller contributed the discussion of explainability, accountability, and distributive justice, and provided final approval of the submitted version. All authors read and approved the final manuscript.

Ethics Statement

As a literature review synthesizing previously published, publicly available research and policy documents, this article did not involve new data collection from human or animal subjects and did not require institutional ethics committee approval.

Data Availability

No new data were generated or analyzed for this review. All sources discussed are cited in the reference list and are publicly available through their respective publishers.

REFERENCES

1. Ethics of allocation of donor organs (2023) Current Opinion in Organ Transplantation.
2. Kidney Organ Allocation System: How to Be Fair (2022) Seminars in Nephrology.
3. Inconsistent values and algorithmic fairness: a review of organ allocation priority systems in the United States (2024) BMC Medical Ethics, DOI:[10.1186/s12910-024-01116-x](https://doi.org/10.1186/s12910-024-01116-x).
4. Reassess Inclusion of Race in Estimated Glomerular Filtration Rate (eGFR) Equation (2021) HRSA/OPTN Public Comment Proposal.
5. Separate and Unequal: Race-Based Algorithms and Implications for Nephrology (2021) PMC.
6. Ethics and Algorithms to Navigate AI's Emerging Role in Organ Transplantation (2025) Journal of Clinical Medicine.
7. Removing race from kidney function algorithm helped more Black patients access transplants (2026) STAT News.
8. Effects of race-neutral estimated glomerular filtration rate policies on access to kidney transplantation (2025) American Journal of Transplantation.
9. Reassessing the Use of Race in Clinical Algorithms: An Interactive, Case-Based Session for Medical Students Using eGFR (2023) MedEdPORTAL, DOI:[10.15766/mep_2374-8265.11412](https://doi.org/10.15766/mep_2374-8265.11412).
10. Transforming liver transplant allocation with artificial intelligence and machine learning: a systematic review (2025) BMC Medical Informatics and Decision Making, DOI:[10.1186/s12911-025-02890-3](https://doi.org/10.1186/s12911-025-02890-3).
11. FairAlloc: Learning Fair Organ Allocation Policy for Liver Transplant (2025) Journal of Healthcare Informatics Research, DOI:[10.1007/s41666-025-00206-8](https://doi.org/10.1007/s41666-025-00206-8).
12. Beyond generic principles: a framework for the ethical analysis of artificial intelligence in organ transplantation (2026) Transplant International, DOI:[10.3389/ti.2026.16525](https://doi.org/10.3389/ti.2026.16525).
13. "Just" accuracy? Procedural fairness demands explainability in AI-based medical resource allocations (2022) AI & Society, DOI:[10.1007/s00146-022-01614-9](https://doi.org/10.1007/s00146-022-01614-9).
14. Prince S, Savulescu J (2025) When is black-box AI justifiable to use in healthcare? Big Data & Society, DOI:[10.1177/20539517251386037](https://doi.org/10.1177/20539517251386037).
15. El-Azab S, Nong P (2023) Clinical algorithms, racism, and "fairness" in healthcare: a case of bounded justice. Big Data & Society, DOI:[10.1177/20539517231213820](https://doi.org/10.1177/20539517231213820).
16. Racial Bias in Clinical Tools and Impact on Organ Transplantation (2024) American Society of Transplantation.
17. Bias in Artificial Intelligence: A Basic Primer (2024) Clinical Journal of the American Society of Nephrology.
18. On the practical, ethical, and legal necessity of clinical Artificial Intelligence explainability: an examination of key arguments (2025) BMC Medical Informatics and Decision Making.
19. One Size Doesn't Fit All: A Review of International Deceased Donor Kidney Allocation Algorithms (2025) PubMed-indexed review.
20. Allocating organs through algorithms and equitable access to transplantation -- a European human rights law approach (2023) Journal of Law and the Biosciences, DOI:[10.1093/jlb/lbad004](https://doi.org/10.1093/jlb/lbad004).
21. Organ Transplantation (2023) The Hastings Center Briefing Book.
22. Allocating a Future: Organ Transplantation (2023) Markkula Center for Applied Ethics.
23. Ethical Principles in the Allocation of Human Organs (2023) HRSA/OPTN.
24. Bioethics in Practice: Ethical Issues in Organ Allocation for Transplantation -- Whose Life Is Worth Saving More? (2015) PMC.
25. Disparities in Access to Solid Organ Transplant Services -- Past, Present, and Future (2023) American Health Law Association, Journal of Health Law.