

Fairness Over Time: Evaluating How Fairness Objectives Shape Health Disparities in AI-Driven Clinical Decision Support Systems

Kaitlyn E. Wade¹, Alice C.W. Huang^{1, 2, 3}, Daniel J. Lizotte^{1, 4}

¹ Department of Computer Science, Western University, London, ON, Canada

² Department of Philosophy, Western University, London, ON, Canada

³ Schwartz Reisman Institute for Technology and Society, Toronto, ON, Canada

⁴ Department of Epidemiology and Biostatistics, Western University, London, ON, Canada

kwade4@uwo.ca, alice.huang@uwo.ca, dlizotte@uwo.ca

Abstract

Fairness is not a single concept and may be operationalized using a variety of definitions reflecting different ethical commitments, yet these are rarely compared systematically in healthcare AI. Fairness is often assessed using static metrics, even though clinical decisions are sequential and evolve over time. In chronic disease management, small inequities early in a care pathway may accumulate into long-term disparities. We develop a Markov cohort microsimulation model to examine how different fairness operationalizations align with or diverge from their intended goals. Using our simulation model, we compare five heuristic allocation policies based on Utilitarian, Egalitarian, Sufficentarian, Prioritarian, and Needs-Based principles against a Status Quo baseline. Outcomes are evaluated longitudinally using quality-adjusted life years, between-group inequality, threshold attainment, and cost-effectiveness. No policy performs best across all criteria; most achieve their intended objective while performing worse on metrics aligned with other frameworks, though the policies with the strongest local equity commitments, Egalitarian and Sufficentarian, notably underperform even on their own aligned metrics. These same two policies also produce the largest cumulative disparities over time. We show that this may come from a structural access mechanism, where locally fair allocation rules do not offset the access disparity that determines whether resources are actually received. These findings show that the choice of fairness definition can have consequential and sometimes counterintuitive effects, reinforcing the need for explicit, longitudinal fairness evaluation.

Code — <https://github.com/kwade4/fairness-over-time>

Introduction

Artificial intelligence (AI) is increasingly being integrated into clinical decision support systems for applications such as diagnostic screening, risk stratification, treatment recommendation, and resource allocation (Elhaddad and Hamam 2024). As these systems move closer to deployment in real-world settings, questions of fairness and equity have become increasingly important. It is well-documented that models trained on historical data can reproduce or amplify existing disparities in healthcare access, treatment, and outcomes

across demographic groups (Obermeyer et al. 2019; Cross, Choma, and Onofrey 2024; Naderalvojjoud et al. 2025).

These considerations are especially important for chronic disease management, where care is shaped by a sequence of interdependent decisions rather than a single prediction. In sequential settings, each action can influence future health states, available treatments, and downstream outcomes, meaning that small differences early in a care pathway can accumulate and influence long-term access to treatment and health outcomes (Jayasooriya et al. 2023; Cross, Choma, and Onofrey 2024).

There is broad consensus in the fairness research community that there is no single definition of fairness, and each one reflects different moral and ethical commitments. Several definitions are even incompatible, with different fairness criteria requiring different and sometimes contradictory constraints on model performance that cannot be simultaneously satisfied (Chouldechova 2017; Kleinberg, Mullainathan, and Raghavan 2017). Critiques from philosophy, law, and sociology have argued that the choice of metric in algorithmic fairness is itself a statement of values about whose interests matter and how benefits and burdens should be distributed (Selbst et al. 2019). This normative dimension is often underexplored in computer science literature, and fairness operationalizations and their effects over time are rarely compared systematically or justified explicitly.

In this paper, we develop a microsimulation framework and use it as an analytic lens to examine how different fairness definitions shape outcomes in sequential clinical decision-making. This allows us to systematically compare fairness definitions in a controlled setting where full patient trajectories are observable. We make three contributions. First, we develop a Markov cohort microsimulation of disease progression based on the Healthy-Sick-Sicker-Dead model that includes inequality at multiple levels, including baseline health status, access to care, and treatment success rates across demographic groups. We include these sources of heterogeneity in an effort to represent real-world disparities in healthcare systems and to ensure fairness objectives are evaluated under realistic structural inequality. Second, we compare five fairness paradigms: Utilitarian, Egalitarian, Sufficentarian, Prioritarian, and Needs-Based, that are implemented as heuristic allocation policies and compared against a Status Quo baseline. These paradigms are grounded in core ideas

from bioethics and distributive justice, and correspond to different views of what it means for a healthcare system to be fair, including maximizing total health, reducing inequality, prioritizing the worst-off, and addressing unmet needs. Third, we show that locally fair policies do not necessarily lead to fair outcomes, and that the link between allocation rules and realized equity depends on the structural barriers present in the environment. In particular, Egalitarian and Sufficiency policies, despite their strong local fairness commitments, produce the largest cumulative disparities over time. We show this is driven by an access mechanism that shapes whether allocated resources are actually received. Our results contribute to a growing body of work on fairness in sequential decision-making and dynamic settings (D'Amour et al. 2020; Wen, Bastani, and Topcu 2021; Henzinger et al. 2023; Deng et al. 2024). We show that different fairness definitions can lead to systematically different patterns of inequality over time, and that they vary in how well they achieve their intended fairness goals.

Background and Related Work

Fairness as a Normative Choice

Fairness in healthcare is complex, multi-faceted, and defines principles for allocating benefits and burdens. These principles are grounded in ideas of distributive justice (how resources should be distributed) and procedural justice (how decisions are made and justified). In healthcare settings, fairness can encode assumptions about whose outcomes matter, how much weight to assign to different health states, and how to prioritize patients with different levels of need and baseline health. We focus on five philosophical conceptions of fairness relevant to healthcare resource allocation.

Utilitarianism aims to maximize aggregate welfare across all individuals and is often summarized as promoting “the greatest good for the greatest number” (Bentham 1789; Mill 2016). As a fairness objective for healthcare resource allocation, it corresponds to policies that maximize overall health benefit, even when doing so may not reduce disparities between groups. Philosopher John Harris argues that individuals typically value three things in healthcare: maximizing life expectancy, maximizing quality of life, and maximizing the chance of achieving both (Harris 1987, 2006).

Egalitarian approaches emphasize equality across individuals or groups and are often considered more normatively acceptable in healthcare than pure Utilitarianism (Dolan and Cookson 2000; Nord et al. 1995; Hynninen, Vilkkumaa, and Salo 2021). Egalitarianism is based on distributive justice and holds that fairness involves giving equal consideration to individuals, often in the form of equal treatment or equal access to opportunities and outcomes (Rawls 1971; Anderson 1999). It emphasizes that individuals should be treated equally with respect to their relevant needs, with similar cases treated similarly. In the algorithmic fairness literature, these ideas are often approximated using group fairness metrics such as demographic parity, which enforce similarity in outcome distributions across demographic groups.

Sufficiency holds that justice does not require equality per se, but that all individuals should reach a minimum

threshold of well-being (Frankfurt 2018; Crisp 2003; Shields 2020). In healthcare ethics, the “fair innings” argument (Bognar 2015) captures a similar intuition, and focuses on ensuring that individuals are provided with enough resources to allow them to reach a “normal” or “sufficient” lifespan (Daniels 1985). This notion is particularly relevant in resource-constrained settings where achieving full equality may be impossible.

Prioritarianism holds that benefits matter more when they go to those who are worse off (Parfit 1995). Unlike Sufficiency, which prioritizes bringing people up to a threshold of well-being and gives less weight to additional gains beyond that threshold, Prioritarianism gives greater moral weight to benefits received by worse-off individuals across the entire distribution. In healthcare, this means that individuals with worse underlying health trajectories may be prioritized because improvements to their health generate greater moral value under a Prioritarian weighting, even when they are not currently the least healthy.

Needs-Based allocation distributes resources based on current clinical need regardless of group membership or underlying social position. It is widely used in health policy and aligns with standard equity principles in public healthcare systems (Daniels 1985; Whitehead 1992). Unlike Prioritarianism, it does not explicitly weigh benefits according to baseline disadvantage or lifetime deprivation, but instead prioritizes individuals based on their current health status.

These five frameworks represent different views of what fairness entails, and each encodes a distinct normative commitment about whose interests matter and how benefits and burdens should be distributed (Selbst et al. 2019). This highlights the need for transparency about the normative assumptions underlying fairness objectives and how they shape resource allocation strategies.

Fairness in Machine Learning

Fairness in machine learning has primarily been studied in static prediction settings, using metrics such as demographic parity, equalized odds, predictive parity, and calibration (Hardt, Price, and Srebro 2016; Verma and Rubin 2018; Barocas, Hardt, and Narayanan 2023) to evaluate whether model outputs and the prediction errors they produce (false positives, false negatives, miscalibration) are balanced across groups at a single point in time. These statistical criteria are often mutually incompatible (Chouldechova 2017; Kleinberg, Mullainathan, and Raghavan 2017). In binary classification settings with observable ground-truth outcomes, differences in base rates create unavoidable trade-offs between properties such as calibration and equalized error rates. We use this as a motivating analogy, recognizing that fairness objectives may encode competing priorities that cannot always be simultaneously satisfied, and examine whether similar trade-offs emerge empirically in our resource allocation setting. Rather than evaluating prediction errors at a single time point, we evaluate how allocation decisions distribute health benefits and burdens over time under stochastic health dynamics.

These static definitions do not extend naturally to settings where decisions have downstream consequences. If feedback effects of decisions are ignored, imposing fairness con-

straints such as demographic parity or equality of opportunity can exacerbate unfairness over time (Wen, Bastani, and Topcu 2021). Thus, because decisions affect future outcomes, enforcing fairness constraints at a single step may lead to unintended consequences. In healthcare, for example, fair treatment allocation at one time point can alter patients' subsequent health and risk profiles, which then affects future allocation decisions and may reinforce disparities over time. This raises the question of how fairness should even be evaluated when decisions affect future states. Existing approaches typically evaluate fairness either locally, at individual time-steps, or globally, aggregated over entire trajectories (Wen, Bastani, and Topcu 2021; Reuel and Ma 2024). These two views can conflict (Deng et al. 2024). A policy may appear fair at each step while producing unequal long-term outcomes or achieve fairness in aggregate while introducing disparities at specific points along the way.

Recent work decomposes inequality into components arising from environmental structure, historical effects, and decision-making (Deng et al. 2024), allowing disparities to be attributed to the decisions being made, the structure of the environment, or historical inequities carried forward into the current state. Since fairness in sequential settings also depends on the history of decisions, rather than the current state alone, there is recognition that fairness may be non-Markovian, and should be evaluated throughout the decision-making process (Alamdari et al. 2024). Studies showing that fairness interventions can degrade with shifts in data distributions further motivate longitudinal monitoring and the evaluation of fairness as a dynamic property rather than a single point-in-time assessment (Castelnovo et al. 2021). Despite these insights, few studies systematically compare the impacts and trade-offs of these choices over time.

Fairness in Sequential Healthcare Decision-Making

Care is delivered through a series of decisions about whether to intervene, which treatment to use, when to adjust therapy, and how closely to monitor progress. These decisions are made in response to a patient's current condition, treatment history, and changing risks. This is especially relevant in chronic disease management, where patients interact with the healthcare system repeatedly over months or years, and each encounter can influence subsequent options and outcomes. This sequential structure has important implications for fairness. Disparities in care at one stage can persist and accumulate over time, creating inequities that may not be visible in cross-sectional analyses. For example, delayed diagnosis or less effective treatment early in disease progression may lead to poorer health, fewer treatment options, and greater risk of adverse outcomes later (Jayasooriya et al. 2023; Cross, Choma, and Onofrey 2024; Naderalvojud et al. 2025). Thus, fairness metrics that evaluate individual decisions in isolation may overlook how disadvantages compound across care trajectories.

Despite growing interest in fairness in AI for healthcare, the literature lacks systematic and controlled comparisons of how different fairness frameworks perform in sequential healthcare settings. Fairness is often treated as an emergent property of model training rather than a deliberate design

decision, and the question of which values are embedded in a given metric is rarely made explicit (Selbst et al. 2019).

Using Microsimulation to Evaluate Fairness

Microsimulation is widely used in public health and health economics to evaluate health policies and interventions. It is especially useful for studying how outcomes are distributed across populations, including inequalities that arise across multiple and overlapping social dimensions (Speybroeck et al. 2013; Horvath et al. 2023; Yang et al. 2026). It also enables low-cost testing of interventions and policy design before real-world deployment. These characteristics make it well-suited for studying fairness in sequential clinical AI systems. Microsimulation enables prospective evaluation of fairness interventions in controlled settings, where patient trajectories and counterfactual outcomes under different fairness definitions can be observed. This is not possible with retrospective analysis of real-world data, where interventions are confounded, trajectories are incomplete, and counterfactuals are unobservable. Furthermore, it allows multiple sources of inequality to be modelled simultaneously, including baseline health, access to care, and treatment response, which better reflects the multi-dimensional nature of health disparities than models that treat inequality as a single variable.

Limitations of Existing Work

While these studies highlight the importance of temporal dynamics for fairness in sequential decision-making, several gaps remain. First, although philosophical and ethical frameworks for fairness are well established, they are rarely systematically compared in computational models of clinical decision-making. The ethical values encoded by fairness operationalizations are often left implicit rather than explicitly stated or justified. Second, prior work shows that fairness at individual decision points can diverge from fairness over full trajectories, and that fairness in sequential settings may depend on past decisions rather than the current state alone. However, much of this literature focuses on developing algorithms or proving guarantees for a single fairness objective. Less attention has been paid to whether different fairness definitions, when treated as objectives, actually produce the kinds of outcomes they aim to achieve when applied over time. This distinction is especially important in healthcare, where sequential decisions shape long-term outcomes.

Our work addresses these gaps by asking a complementary question: given that multiple fairness definitions exist, each encoding different ethical commitments, what are the consequences of choosing one over another in a clinical setting, and how do these consequences unfold over time?

Methods

Simulation Environment

We model clinical decision-making as a Markov Decision Process (MDP) using a microsimulation of chronic disease progression. This provides a controlled environment to examine how different fairness objectives shape outcomes over time across heterogeneous populations. The allocation policies evaluated are pre-specified heuristics rather than learned

policies, a design chosen to provide transparent implementations of the five ethical frameworks under study and isolate the impact of each fairness objective from the influence of a particular learning algorithm. We view this as a necessary first step before studying how learned policies may approximate or operationalize these objectives.

Health State Model Our simulation is inspired by SIMA, an open-source healthcare simulation (Tikka et al. 2021) environment. Unlike SIMA, our state space includes both discrete health status and a continuous latent health score that tracks within-state disease progression and accumulated treatment effects. The discrete component of the state space uses a Healthy-Sick-Sicker-Dead structure, commonly used in healthcare economics to model progressive deterioration in health (Siebert et al. 2012; Iskandar 2018; Alarid-Escudero et al. 2023). Individuals begin in a healthy state and may move to worse health states over time, with death as an absorbing state.

The continuous component is a latent health score bounded in $[0, 1]$. At each step, the score is updated stochastically as a function of baseline health resilience, current disease state, and treatment receipt. Discrete states are derived from this score via fixed thresholds: scores above 0.8 correspond to Healthy, between 0.4 and 0.8 to Sick, below 0.4 and above 0 to Sicker, and 0 to Dead. Because the score evolves continuously before crossing a threshold, the model captures gradual within-state deterioration that influences transition risk before any formal state change occurs. For example, a person in the Sick state with a score of 0.79 is close to recovery, while one with a score of 0.41 is close to becoming Sicker, even though both are formally Sick.

With this augmented state representation, the transition dynamics are Markovian, in the sense that future states depend only on the current state and action, rather than the full history. This is a deliberate choice, as individuals in the same discrete health state can still carry different risks due to prior disease progression and treatment exposure. Including the latent score therefore preserves Markov structure while allowing more realistic, history-sensitive disease progression than a memoryless discrete-state model.

Population Structure We simulate a population of $N=100,000$ individuals organized into cohorts defined by deprivation quintiles (Q0 to Q4), where Q0 represents the most deprived and Q4 the least. A latent socioeconomic score sampled within each quintile generates three correlated individual-level attributes that remain fixed throughout the simulation:

- **Care access:** the probability that an allocated resource is received, representing structural barriers such as insurance coverage, clinic availability, and transportation.
- **Treatment efficacy:** the degree to which an individual benefits from a received treatment, representing biological or contextual variation in treatment response.
- **Healthiness:** a resilience parameter governing the rate of health deterioration without treatment, representing baseline vulnerabilities from factors such as comorbidities, environmental exposure, or age.

Algorithm 1: Simulation Loop

```

for each time step  $t = 1, \dots, T$  do
  Allocate treatment budget across living individuals according to policy
  for each living individual  $i$  do
    Treatment:
    if Bernoulli( $a_i$ ) = 1 then
       $\text{health}_i \leftarrow \text{health}_i + \Delta_i^{\text{treat}}$ 
    end if
    Baseline decay:
     $\text{health}_i \leftarrow \text{health}_i - \Delta_i^{\text{age}}$ 
    Stochastic transition:
     $\text{health}_i \leftarrow \text{health}_i - \Delta_i^{\text{stochastic}}(\rho_i)$ 
     $\text{health}_i \leftarrow \text{clip}(\text{health}_i, 0, 1)$ 
    Derive discrete state  $s_i \in \{H, S_1, S_2, D\}$  from  $\text{health}_i$ 
  end for
  Accrue QALYs and costs for all living individuals
end for

```

MDP Formulation The environment is defined by the tuple (S, A, R, T, γ) . States represent an individual’s current health status, disease history, and sociodemographic characteristics. Actions represent available clinical decisions at each time point, such as diagnostic testing or treatment initiation. The reward is based on health outcomes, including QALYs, treatment costs, disease progression, and mortality, and is designed so that fairness objectives can be incorporated as modifications to the base reward. Transition dynamics follow those specified in the SIMA model (Tikka et al. 2021) and depend on individual characteristics, prior decisions, and stochastic disease processes. The simulation runs over a 50-year horizon with monthly decision points, letting us study how disparities emerge and compound across the full course of care. At each step t , the simulation proceeds as described in Algorithm 1.

Validation

We validate the model using the Status Quo policy, which allocates resources based on current access (Figure 1). This provides a baseline check that the simulation produces plausible disease progression and inequality dynamics. We also test the robustness of the simulation by perturbing transition parameters by $\pm 20\%$, varying the resource allocation budget, population size, and time horizon, and by running multiple random seeds. Across these settings, the relative ordering of policies and the main qualitative results remain stable.

Decision Policies

The simulation operates under limited treatment capacity at each decision point and requires policies to allocate a fixed resource budget across individuals. We evaluate heuristic decision policies that apply pre-specified clinical allocation rules, analogous to protocol-driven clinical care. The Status Quo allocation strategy is the baseline comparator and allocates resources in proportion to existing care access, reflecting underlying structural inequality without any corrective

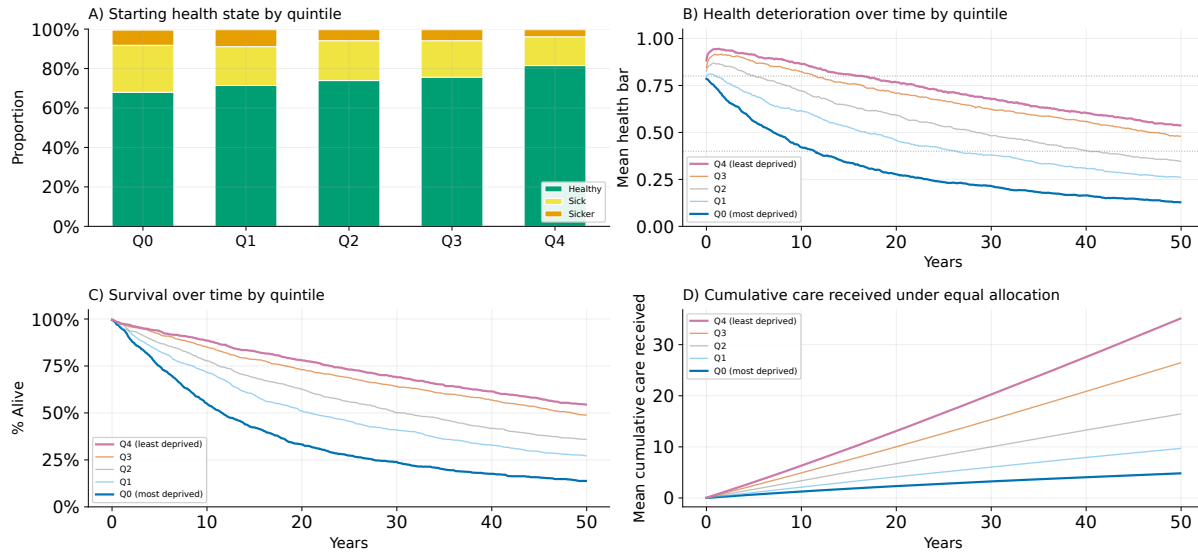


Figure 1: Validation of the simulation model under the Status Quo policy, showing A) the state distribution, B) health deterioration, C) survival analysis, and D) cumulative care received by each quintile.

adjustments. We then compare five allocation policies, each corresponding to one fairness approach.

- **Utilitarian:** prioritizes individuals with the highest expected health gain, defined as the product of health deficit and treatment efficacy.
- **Egalitarian:** allocates an equal share of resources to every living individual, regardless of health state or group membership. This reflects equality of resources rather than equality of outcomes. A luck egalitarian view would instead give more to those in worse health due to circumstances outside their control, compensating for brute luck rather than equalizing inputs or outcomes uniformly across the population (Dworkin 1981; Cohen 1989; Arneson 1989). The Needs-Based policy approximates this logic and allocates resources based on current health deficits, which are outside of an individual's control.
- **Sufficientarian:** prioritizes individuals below a minimum health threshold and allocates remaining resources to individuals above it only once those needs are met. When resources are insufficient to bring everyone up to the threshold, the worst-off are prioritized first to reduce the overall gap to the threshold. If all individuals are above the threshold, resources are distributed equally.
- **Prioritarian:** weights allocation by current health severity and structural disadvantage, giving a currently healthy individual from a deprived group higher priority than an equally healthy individual from a less deprived group.
- **Needs-Based:** allocates proportional to current health deficit regardless of group membership, with no threshold and no lifetime weighting.

Translating philosophical fairness theories into allocation rules requires simplifying their normative commitments. Our

aim is not to fully formalize each theory, but to define transparent policies capturing their central principles for comparison. These policies should therefore be interpreted as approximations of each framework's core commitments, not complete representations of the underlying theories.

Experimental Design

To study the contribution of each axis, we run experiments across all possible on/off combinations of the three sources of disparity described earlier: care access, treatment efficacy, and healthiness, along with a fully equal baseline with no active inequality. Within each condition, we evaluate all five policies using the same underlying population and transition dynamics. This isolates which sources of structural disadvantage drive divergence between fairness definitions, and whether some fairness objectives are more robust or more sensitive to particular forms of disparity.

Evaluation

We evaluate each strategy using five metrics that measure health outcomes, distributional fairness, threshold equity, resource allocation, and cost-effectiveness. We evaluate all metrics across decision strategies and heterogeneity conditions over the full 50-year horizon.

Quality-Adjusted Life Years (QALYs) QALYs combine length and quality of life into a single utility-weighted measure, where one QALY represents one year lived in perfect health (Zeckhauser and Shepard 1976; Pliskin, Shepard, and Weinstein 1980; Drummond et al. 2015). Individuals in the Healthy, Sick, and Sicker states accrue QALYs at weights of 1.0, 0.75, and 0.5 per year, respectively, while the Dead state accrues none. We report mean QALYs overall and by

deprivation group, including the Q4–Q0 QALY gap between the least and most deprived groups, to assess health gains and their distribution. Although widely used in healthcare economics, QALY maximization alone may disadvantage populations with higher treatment costs or lower expected gains, including disadvantaged groups (Whitehead and Ali 2010).

Distributional Fairness We measure population-level inequality in QALY outcomes using the Gini coefficient, a summary measure of inequality across a population (Ceriani and Verme 2012). This operationalizes an Egalitarian objective by capturing disparities in the overall QALY distribution. The Gini coefficient is conceptually related to demographic parity in the algorithmic fairness literature, which seeks to reduce disparities in outcomes across groups (Barocas, Hardt, and Narayanan 2023). However, unlike demographic parity, it does not require predefined groups and instead measures inequality across the population as a whole.

Threshold-Based Equity We measure threshold-based equity as the proportion of individuals in each quintile whose cumulative QALYs meet or exceed a prespecified minimum threshold (floor attainment). This reflects a Sufficiency objective by evaluating whether individuals achieve an acceptable health level rather than whether outcomes are equal across groups. This metric is conceptually related to fairness criteria that compare group-level rates of binary outcomes (Hardt, Price, and Srebro 2016), but differs in that the binary outcome represents attainment of a health threshold rather than a classification outcome.

Resource Allocation and Outcomes We measure disparities in resource allocation using mean cumulative care received by the most deprived group (Q0 Care) and the Q4–Q0 care gap between the least and most deprived groups.

Incremental Cost-Effectiveness Ratio (ICER) ICER measures the additional cost required to generate one additional QALY, relative to the Status Quo baseline. It is calculated as the difference in total costs divided by the difference in total QALYs over the simulation horizon. Lower values indicate higher cost-effectiveness (Drummond et al. 2015). ICER is inherently utilitarian in that it evaluates interventions based on total QALYs per dollar without accounting for how those QALYs are distributed across groups. We report ICER alongside distributional and threshold-based metrics to make this trade-off explicit.

Relationship to Statistical Fairness Criteria Although our evaluation framework is grounded in distributive and philosophical accounts of fairness, we relate our metrics to statistical fairness criteria (Barocas, Hardt, and Narayanan 2023) to situate them within the broader AI fairness literature. Metrics such as demographic parity and equalized odds were developed for prediction settings with defined outcomes and, in some cases, ground-truth labels, whereas our framework considers sequential resource allocation with trade-offs in resource distribution and continuous health outcomes. We therefore interpret our metrics as fairness-inspired analogues: the Q4–Q0 care gap measures disparities in allocation,

the Q4–Q0 QALY gap measures disparities in realized outcomes, the Gini coefficient measures population-level inequality, and floor attainment evaluates minimum outcome attainment rather than group parity.

Results

No Single Definition of Fairness Dominates

No single policy outperformed the others across all metrics and conditions. Three of the five policies ranked first or second on the metric aligned with their own normative goal, while two ranked in the bottom half. In general, policies that performed well on their own aligned metric underperformed on metrics reflecting other fairness goals. This reflects a broader tension in fairness research, where competing definitions of fairness can require different and sometimes incompatible trade-offs (Chouldechova 2017; Kleinberg, Mullainathan, and Raghavan 2017).

Table 1 evaluates each policy using the metric that aligns with its underlying normative goal. Under the full inequality condition, Utilitarianism produces the highest aggregate QALYs, and Prioritarianism allocates the most care to the most deprived quintile (Q0), each ranking first on their own aligned metric. Needs-Based allocation ranks second on the Q4–Q0 care gap, the metric most closely aligned with its objective, narrowly behind Prioritarianism.

Sufficiency and Egalitarianism are notable exceptions. Despite aiming to provide a minimum acceptable outcome for the worst-off, Sufficiency ranks fourth of five policies on threshold attainment for Q0, ahead of only Egalitarianism. Similarly, despite allocating resources equally across individuals, Egalitarianism performs worst on the Gini coefficient, the metric aligned with its objective of population-level equality. This indicates that equal allocation does not necessarily produce equal outcomes when baseline vulnerability and access to care differ across groups, a mechanism we examine further in the subsequent section.

Across all six metrics, Utilitarianism performs strongly, ranking first on four criteria and third on both care-allocation measures. Prioritarianism also performs consistently well beyond its aligned metric, ranking second in total QALYs and the Gini coefficient in addition to first in Q0 Care. In contrast, Egalitarianism and Sufficiency perform poorly across most criteria, including their own aligned objectives. This pattern appears to emerge from the interaction between its weighting scheme and the simulation's disease dynamics, and warrants further investigation.

These results should not be interpreted as an overall ranking of fairness frameworks. Each metric encodes a distinct normative assumption about fair allocation, and no single metric provides a complete evaluation. Instead, different fairness concepts lead to different behaviours and outcomes depending on the normative objective. We next examine potential mechanisms behind these trade-offs.

Short-Term Fairness Does Not Guarantee Long-Term Equity

Over a 50-year time horizon, Utilitarian, Prioritarian, and Needs-Based policies have the smallest QALY gaps between

Policy	Total QALYs (↑)	Care received Q0 (↑)	Care gap Q4-Q0 (↓)	Q0 Above Floor (↑)	Gini (↓)	QALY Gap Q4-Q0 (↓)
Utilitarian	17927* (1)	2.21 (3)	0.14 (3)	0.77 (1)	0.088 (1)	0.79 (1)
Prioritarian	17804 (2)	4.01* (1)	-2.98 (1)	0.755 (2)	0.094 (2)	1.05 (3)
Needs-Based	17754 (3)	2.52 (2)	-0.51* (2)	0.752 (3)	0.097 (3)	1.02 (2)
Sufficientarian	16891 (4)	1.80 (4)	2.58 (4)	0.625* (4)	0.134 (4)	1.89 (5)
Egalitarian	16844 (5)	1.77 (5)	2.63 (5)	0.610 (5)	0.136* (5)	1.88 (4)

Table 1: Policy performance across six evaluation criteria under full inequality. Arrows indicate the direction of better performance. Ranks (1 = best, 5 = worst) are shown in parentheses, and * marks each policy’s aligned normative objective. Q0 denotes the most deprived quintile and Q4 the least deprived. Care gap and QALY gap are calculated as $Q4 - Q0$, where negative values indicate greater allocation or outcomes favouring Q0 and positive values favour Q4.

the most and least advantaged individuals (Figure 2). In contrast, the Egalitarian and Sufficientarian policies perform worse, with the Egalitarian policy performing similarly to the Status Quo baseline and Sufficientarian showing the largest gap. At earlier time points, these policies start with relatively small gaps, but over longer time horizons, differences become more pronounced.

All five policies show growing divergence between local and global fairness over the simulation horizon. The rate of divergence differed substantially across policies, by a factor of approximately 2.1 between the fastest and slowest (Table 2). Sufficientarian (0.935) and Egalitarian (0.907) diverge fastest despite applying locally fair rules at every decision point, illustrating that local fairness does not guarantee global fairness in sequential settings (Alamdari et al. 2024). This pattern is reflected in Table 3, where both Egalitarian and Sufficientarian show high care access-received correlations (0.85). In these cases, resources are more often received by individuals with higher access, even though allocation is locally equal. This suggests that the observed disparities arise not from the allocation rules themselves, but from structural differences in access that affect whether resources are actually received.

Utilitarian (0.446), Prioritarian (0.472), and Needs-Based (0.534) diverge more slowly and show negative or near-zero access-care correlations, meaning they actively directed resources toward lower-access individuals. Utilitarian’s stability likely reflects that concentrating resources where expected health gains are highest incidentally favors more deprived individuals under these simulation conditions, rather than a normative property of Utilitarian allocation itself.

These results suggest locally fair allocation rules alone do not ensure long-term equity when access barriers remain. The weaker performance of Egalitarian and Sufficientarian does not reflect a flaw in either framework, but rather that, in this implementation, neither addresses the barriers that determine whether allocated resources are received.

Disparities Compound Over Time and Are Sensitive to Their Source

Table 4 reports each policy’s performance on its own objective under three isolated inequality sources (access, efficacy, and health) and the full inequality setting. Each source uses the same underlying inequality parameterization (i.e., the

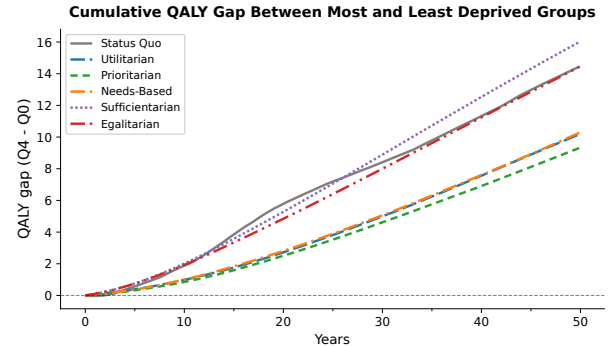


Figure 2: QALY difference between the least (Q4) and most (Q0) deprived groups.

Policy	Gap ($t = 0$)	Gap (final)	Divergence
Prioritarian	0.000	0.472	0.472
Needs-Based	0.002	0.536	0.534
Utilitarian	0.002	0.448	0.446
Sufficientarian	0.002	0.936	0.935
Egalitarian	0.002	0.909	0.907

Table 2: Divergence between local and global fairness over time. Divergence growth is defined as the change in the QALY gap between $t = 0$ and the final time point.

Policy	Access-Care corr.	Q0 Care	Q4 Care
Prioritarian	-0.416	4.01	1.03
Needs-Based	-0.115	2.52	2.02
Utilitarian	-0.016	2.21	2.35
Sufficientarian	0.849	1.80	4.38
Egalitarian	0.850	1.77	4.40

Table 3: Access barriers under full inequality. The access–receipt correlation indicates whether care is concentrated among higher-access (positive) or lower-access (negative) individuals. Q0 and Q4 care values represent mean cumulative care over the simulation horizon.

same magnitude of disparity), so differences across conditions are less likely to be explained by differences in inequality magnitude alone and may instead reflect the distinct role

each source plays in the treatment process. The results show that the source of inequality shapes how well each policy achieves its intended goal, and in some cases alters the relative ordering of policies across conditions. Overall, allocation performance is not consistent across environments and depends on which structural constraint is active.

Access-only inequality presents a substantial challenge for allocation policies to address. Under this setting, Sufficientarian floor attainment is 0.615, which is lower than under efficacy-only (0.795) and health-only (0.777) conditions. Because the underlying inequality parameterization is constant, this pattern is consistent with an access-barrier mechanism rather than differences in inequality magnitude alone. When inequality is driven by differences in access, allocation rules that do not explicitly counteract the access gradient may fail to ensure that resources are received, even when they are allocated appropriately. This is consistent with the high access–care correlations observed for Egalitarian and Sufficientarian policies in the previous section, where locally equal allocation did not prevent disparities in resource receipt.

Needs-Based allocation shows a similar sensitivity to inequality source. Under access-only and full inequality, the Q4–Q0 care gap is strongly negative (−0.552 and −0.506, respectively), indicating greater care directed toward Q0. Under efficacy-only inequality, the gap remains negative but is much smaller in magnitude (−0.036), indicating only weak alignment with Q0’s need. Under health-only inequality, the gap becomes slightly positive (0.034), meaning Q4 receives marginally more care than Q0, the opposite of Needs-Based’s intended direction. This suggests that allocating based on current health deficit may be less effective when disparities originate from baseline health differences alone. One possible explanation is that baseline differences can build up over time, even if they do not always result in greater need at each decision point.

The full inequality condition is challenging across policies, likely due to interactions between sources of inequality, rather than their independent effects alone. These results suggest that no policy is optimal across all structural conditions, and that relative effectiveness depends on which sources of inequality are present in the environment.

Cost-Effectiveness Across Policies

ICER is reported as a supplementary measure for comparability with health economic evaluations. Because it reflects aggregate QALY gains relative to costs without accounting for how benefits are distributed, it is not used as a basis for evaluating fairness across policies (Drummond et al. 2015).

ICER results differ from the welfare- and equity-based evaluations above. Needs-Based allocation is the most cost-effective (−13,729), followed by Egalitarian (−12,895), Prioritarian (−12,587), Utilitarian (−11,687), and Sufficientarian (−11,340). One possible explanation is that prioritizing individuals with greater current health deficits may direct resources toward larger marginal health gains per unit cost, although this mechanism requires further investigation. Overall, ICER may diverge from fairness-based evaluations because it prioritizes efficiency without accounting for how health benefits

are distributed across groups.

Discussion

In this paper, we use microsimulation as an analytic tool to study how different fairness definitions affect outcomes over time in a sequential clinical setting. Our framework allows fairness interventions to be evaluated prospectively, in a controlled environment where the consequences of different design choices can be traced across full patient trajectories. This enables a prospective fairness audit that is not possible with retrospective analysis of real-world data, where interventions are confounded, and trajectories are incomplete.

Allocation Policies Cannot Eliminate Structural Access Barriers

Across all conditions, disparities between the most and least deprived groups persist regardless of the allocation policy used. Policies that do not actively counteract the access gradient, namely Egalitarian and Sufficientarian, show high access–receipt correlations of approximately 0.85. This indicates that individuals with structural disadvantages tend to receive less of the care that is formally allocated to them. Even under access-only inequality, Sufficientarian floor attainment remains limited at 0.615.

This finding highlights an important limitation of fairness interventions that operate only at the level of allocation. Allocation rules that are fair at the decision stage are not sufficient to produce equitable outcomes when structural barriers affect whether resources are actually received. Thus, improving equity in clinical AI systems may require changes not only to decision policies, but also to the broader healthcare infrastructure that determines whether patients can receive care in the first place.

Empirical Evidence for the Local-Global Fairness Tradeoff

Prior work has shown that static analyses of fairness do not capture the long-term consequences of ML-based decision systems, and that simulation offers a tractable way to study these dynamics (D’Amour et al. 2020). We extend this by examining how these dynamics unfold across ethical frameworks and sources of structural inequality in a clinical allocation setting. We show that while local and global fairness diverge across policies, the rate of divergence depends on how each policy interacts with structural inequality, particularly whether it counteracts or reinforces access disparity.

Egalitarian applies a locally fair rule at every decision point, yet produced the second-fastest divergence (growth = 0.907), and Sufficientarian diverged fastest (0.935). In both cases, structural inequalities dominate outcomes, and “fair” inputs produced “unfair” outputs because of structural access, not the allocation rule, that determined who received care. Conversely, Prioritarian introduces deliberate local inequality at each step and allocates more resources to the most deprived individuals, and produces the second-slowest divergence (0.472). A cross-sectional fairness assessment would characterize its allocation as unequal, while a longitudinal view suggests it is among the more effective policies

Impact of Disparity Sources on Fairness Policy Outcomes

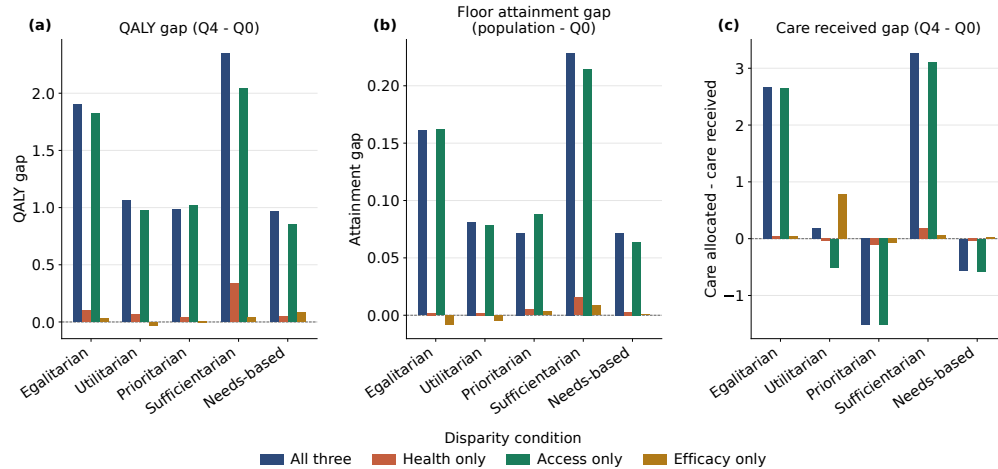


Figure 3: Impact of disparity sources on A) QALY gap between least (Q4) and most deprived (Q0) individuals, B) floor attainment gap between the population average and Q0 under various structural inequalities, and C) Gap between amount of care allocated and received across policies under different sources of structural inequalities.

Policy	Metric	Access	Efficacy	Health	All Axes
Utilitarian	Total QALY ($\uparrow$)	17,862	17,945	18,028	17,927
Prioritarian	Care Q0 ($\uparrow$)	4.04	5.16	5.26	4.01
Needs-Based	Care gap Q4-Q0 ($\downarrow$)	-0.552	-0.036	0.034	-0.506
Sufficientarian	% Q0 above floor ($\uparrow$)	0.615	0.795	0.777	0.625
Egalitarian	Gini ($\downarrow$)	0.133	0.120	0.118	0.136

Table 4: Policy-specific metric performance by inequality axis. Each policy is evaluated on its own target metric under each single-axis and full inequality condition.

at limiting cumulative disparity. Neither pattern is visible through single timepoint evaluation. Overall, the results suggest that longitudinal fairness depends less on whether a policy appears locally fair and more on whether it interacts with structural mechanisms that determine whether allocated resources are received.

Tradeoffs are Inevitable and Should be Made Explicit

No single policy consistently performed best across all conditions, metrics, and time points. This reflects a broader tension in fairness research, where competing objectives can lead to different and sometimes incompatible trade-offs (Chouldechova 2017; Kleinberg, Mullainathan, and Raghavan 2017). Several results challenged intuitive assumptions. Decisions that appeared fair locally could produce unequal long-term outcomes. Egalitarian and Sufficientarian policies do not always perform best for the worst-off groups over time. The Needs-Based policy, despite allocating in proportion to current health deficit, directed marginally more care toward the least deprived group under health-only inequality (Table 4). This suggests that health need and deprivation are not always aligned under different sources of inequality. Furthermore, the policy with no explicit equity objective, Utilitarian, di-

verged slowest and performed consistently across varying inequality conditions. The cost-effectiveness results reinforce this point. ICER, which reflects aggregate QALY maximization per unit cost, ranks Needs-Based first and Utilitarian fourth. A decision-making model that relies on ICER as the primary evaluation criterion would therefore be likely to reach different conclusions than one based on distributional metrics, reflecting that these measures encode different assumptions about what constitutes a good allocation.

These trade-offs appear to reflect structural tensions between competing normative objectives rather than artifacts of parameter tuning or algorithmic design. Balancing aggregate benefit, distribution, and priority setting requires explicit choices about whose outcomes matter most. Therefore, fairness definitions cannot be selected purely empirically and should be explicitly justified during policy design.

Implications for Healthcare AI Design

Our findings have several implications for the design and deployment of AI-driven clinical decision support systems.

1. **Fairness should be incorporated from the outset.** Fairness should be treated as a design objective rather than an emergent property of training. Our results show that optimizing for health outcomes without an explicit fair-

ness objective does not automatically produce equitable results and may reinforce pre-existing structural inequalities. This aligns with broader critiques of AI development (Selbst et al. 2019), which argue that fairness requires active normative choices rather than passive optimization.

2. **Fairness definitions should be explicitly chosen and justified.** Each framework evaluated here encodes a different ethical commitment and produces different patterns of benefit and disadvantage. The operationalization of these frameworks as allocation rules necessarily simplifies their full philosophical content. Developers, clinicians, decision-makers, and affected communities should collaboratively discuss this choice and recognize that decisions about fairness definitions should be made with both short- and long-term consequences in mind.
3. **Fairness evaluation should be longitudinal.** Point-in-time assessments can miss how disparities emerge and compound over time. Although Egalitarian and Sufficentarian policies appeared locally equitable, they produced the largest cumulative disparities. Systems should monitor fairness over time to identify and mitigate emerging inequities before they become entrenched.
4. **Structural barriers require structural solutions.** Allocation policies alone cannot address access barriers that arise outside the decision-making process. Equitable clinical AI systems should therefore be paired with interventions that address the underlying conditions influencing whether allocated resources are ultimately received.

Limitations and Future Directions

This work provides a controlled framework for studying how fairness objectives shape longitudinal health outcomes, but several limitations should be considered. First, the Markov assumption underlying the simulation is a simplification of clinical decision-making. In reality, health states are only partially observed, and factors such as patient behaviour and treatment adherence introduce uncertainty. Although our model incorporates some latent dynamics, a partially observable MDP (POMDP) formulation (Kaelbling, Littman, and Cassandra 1998) would better represent these complexities. Second, our simulation model is intentionally simplified. Treatment is modelled as having immediate positive effects, whereas real treatments may have delayed or adverse outcomes. Therefore, results should be interpreted as demonstrating relative policy trade-offs rather than predictions of clinical outcomes. Additionally, the evaluated policies are heuristic rather than learned, meaning our findings characterize how fairness objectives behave under fixed decision rules, rather than how learned policies optimizing these objectives may behave. Finally, the synthetic population is not calibrated to real-world data. While the simulated heterogeneity reflects plausible sources of health disparity, future work should calibrate and validate the model using longitudinal clinical data to assess whether these trade-offs generalize. We release the simulation code as open source to support reproducibility and future extensions.

There are also several directions for future work. Our work examines heuristic policies as a first step towards understand-

ing how fairness definitions shape longitudinal outcomes. Training reinforcement learning agents under each fairness objective would test whether learned policies reproduce these trade-offs or exploit ambiguities in objective specification. The simulation could also be extended with richer disease models, non-linear progression dynamics, and partial observability. Our simulation framework could also be applied to other sequential decision-making settings, including social services and education, where similar questions about fairness over time arise.

Most importantly, this work provides a simulation-based framework for studying how fairness definitions shape longitudinal equity outcomes. From this framework, we derive practical guidelines for the development and evaluation of fair AI systems, including the need for prospective assessment of fairness over time, explicit justification of fairness definitions, and monitoring beyond single decision points. These guidelines can inform the development of fairness monitoring tools, enabling developers to anticipate how design choices propagate over time and reframing fairness evaluation as an ongoing component of responsible AI development rather than a post hoc audit.

Conclusions

This work develops a microsimulation framework to model sequential clinical decision-making. We use our simulation to examine how the choice of fairness definition shapes health disparities over time, and whether outcomes align with each fairness paradigm's intended goal. We find that no single framework dominates across all conditions and metrics; most policies achieve their intended goal while underperforming on metrics aligned with other frameworks, though Egalitarian and Sufficentarian underperform even on their own aligned objectives. Notably, these same two policies, which have the strongest local equity commitments, produced faster longitudinal divergence between groups, while Utilitarian, whose allocation rule embeds no explicit equity consideration, diverged slowest under the simulated structural conditions. This suggests that allocation rules designed to promote equity at the point of decision may not translate into equitable outcomes when structural barriers limit whether resources are actually received.

Our findings demonstrate that the choice of fairness definition has consequential and often counterintuitive effects. Fairness in healthcare AI cannot be treated as an emergent property of training or evaluated only post hoc; it requires explicit consideration throughout development and deployment. The trade-offs identified here reflect normative choices that cannot be resolved by algorithms alone. They reflect substantive disagreements about values, and the choice of fairness definition directly shapes who benefits and who does not, in ways that may not be apparent in cross-sectional evaluation. Making these choices visible through longitudinal simulation is a necessary step toward more deliberate, transparent, and equitable design of AI systems in healthcare.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Discovery Grant program (RGPIN-2024-05673). This work was also supported by the Canadian Institutes of Health Research (CIHR) through the Canada Graduate Scholarships—Doctoral (CGS D) award program (536598).

References

- Alamdari, P. A.; Klassen, T. Q.; Creager, E.; and McIlraith, S. A. 2024. Remembering to be fair: non-Markovian fairness in sequential decision making. In *Proceedings of the 41st International Conference on Machine Learning*, 906–920.
- Alarid-Escudero, F.; Krijkamp, E.; Enns, E. A.; Yang, A.; Hunink, M. G. M.; Pechlivanoglou, P.; and Jalal, H. 2023. A tutorial on time-dependent cohort state-transition models in R using a cost-effectiveness analysis example. *Medical Decision Making*, 43(1): 21–41.
- Anderson, E. S. 1999. What is the Point of Equality? *Ethics*, 109(2): 287–337.
- Arneson, R. J. 1989. Equality and equal opportunity for welfare. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 56(1): 77–93.
- Barocas, S.; Hardt, M.; and Narayanan, A. 2023. *Fairness and machine learning: Limitations and opportunities*. MIT press.
- Bentham, J. 1789. An Introduction to the Principles of Morals and Legislation. In Harrison, W.; and Burns, J. H., eds., *The Collected Works of Jeremy Bentham*. Oxford: Clarendon Press. 1996.
- Bognar, G. 2015. Fair innings. *Bioethics*, 29(4): 251–261.
- Castelnovo, A.; Malandri, L.; Mercurio, F.; Mezzanzanica, M.; and Cosentini, A. 2021. Towards fairness through time. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 647–663. Springer.
- Ceriani, L.; and Verme, P. 2012. The origins of the Gini index: extracts from *Variabilità e Mutabilità* (1912) by Corrado Gini. *The Journal of Economic Inequality*, 10(3): 421–443.
- Chouldechova, A. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2): 153–163.
- Cohen, G. A. 1989. On the currency of egalitarian justice. *Ethics*, 99(4): 906–944.
- Crisp, R. 2003. Equality, priority, and compassion. *Ethics*, 113(4): 745–763.
- Cross, J. L.; Choma, M. A.; and Onofrey, J. A. 2024. Bias in medical AI: Implications for clinical decision-making. *PLOS Digital Health*, 3(11): e0000651.
- D’Amour, A.; Srinivasan, H.; Atwood, J.; Baljekar, P.; Sculley, D.; and Halpern, Y. 2020. Fairness is not static: deeper understanding of long term fairness via simulation studies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 525–534.
- Daniels, N. 1985. *Just health care*. Cambridge University Press.
- Deng, Z.; Jiang, J.; Long, G.; and Zhang, C. 2024. What hides behind unfairness? Exploring dynamics fairness in reinforcement learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 3908–3916.
- Dolan, P.; and Cookson, R. 2000. A qualitative study of the extent to which health gain matters when choosing between groups of patients. *Health Policy*, 51(1): 19–30.
- Drummond, M. F.; Sculpher, M. J.; Claxton, K.; Stoddart, G. L.; and Torrance, G. W. 2015. *Methods for the economic evaluation of health care programmes*. Oxford University Press.
- Dworkin, R. 1981. What is equality? Part 1: Equality of welfare. *Philosophy & Public Affairs*, 10(3): 185–246.
- Elhaddad, M.; and Hamam, S. 2024. AI-driven clinical decision support systems: an ongoing pursuit of potential. *Cureus*, 16(4): e57728.
- Frankfurt, H. 2018. Equality as a moral ideal. In *The notion of equality*, 367–389. Routledge.
- Hardt, M.; Price, E.; and Srebro, N. 2016. Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, 29.
- Harris, J. 1987. QALYfying the value of life. *Journal of Medical Ethics*, 13(3): 117–123.
- Harris, J. 2006. *The value of life: an introduction to medical ethics*. Routledge.
- Henzinger, T.; Karimi, M.; Kueffner, K.; and Mallik, K. 2023. Runtime monitoring of dynamic fairness properties. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 604–614.
- Horvath, T.; Leoni, T.; Reschenhofer, P.; and Spielauer, M. 2023. Socio-economic inequality and healthcare costs over the life course—A dynamic microsimulation approach. *Public Health*, 219: 124–130.
- Hynninen, Y.; Vilkkumaa, E.; and Salo, A. 2021. Operationalization of utilitarian and egalitarian objectives for optimal allocation of health care resources. *Decision Sciences*, 52(5): 1169–1208.
- Iskandar, R. 2018. A theoretical foundation for state-transition cohort models in health decision analysis. *PLOS One*, 13(12): e0205543.
- Jayasooriya, N.; Baillie, S.; Blackwell, J.; Bottle, A.; Petersen, I.; Creese, H.; Saxena, S.; Pollok, R. C.; and study group, P. O. P.-I. B. D. 2023. Systematic review with meta-analysis: time to diagnosis and the impact of delayed diagnosis on clinical outcomes in inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics*, 57(6): 635–652.
- Kaelbling, L. P.; Littman, M. L.; and Cassandra, A. R. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2): 99–134.
- Kleinberg, J.; Mullainathan, S.; and Raghavan, M. 2017. Inherent Trade-Offs in the Fair Determination of Risk Scores. In Papadimitriou, C. H., ed., *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 1–23. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. ISBN 978-3-95977-029-3.

- Mill, J. S. 2016. Utilitarianism. In *Seven masterpieces of philosophy*, 329–375. Routledge.
- Naderalvojud, B.; Curtin, C.; Asch, S. M.; Humphreys, K.; and Hernandez-Boussard, T. 2025. Evaluating the impact of data biases on algorithmic fairness and clinical utility of machine learning models for prolonged opioid use prediction. *JAMIA Open*, 8(5): ooaf115.
- Nord, E.; Richardson, J.; Street, A.; Kuhse, H.; and Singer, P. 1995. Maximizing health benefits vs egalitarianism: an Australian survey of health issues. *Social Science & Medicine*, 41(10): 1429–1437.
- Obermeyer, Z.; Powers, B.; Vogeli, C.; and Mullainathan, S. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464): 447–453.
- Parfit, D. 1995. *Equality or priority?* University of Kansas Kansas.
- Pliskin, J. S.; Shepard, D. S.; and Weinstein, M. C. 1980. Utility functions for life years and health status. *Operations Research*, 28(1): 206–224.
- Rawls, J. 1971. An egalitarian theory of justice. *Philosophical ethics: An Introduction to Moral Philosophy*, 365–370.
- Reuel, A.; and Ma, D. 2024. Fairness in reinforcement learning: A survey. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, volume 7, 1218–1230.
- Selbst, A. D.; Boyd, D.; Friedler, S. A.; Venkatasubramanian, S.; and Vertesi, J. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68.
- Shields, L. 2020. Sufficiencyarianism 1. *Philosophy Compass*, 15(11): 1–10.
- Siebert, U.; Alagoz, O.; Bayoumi, A. M.; Jahn, B.; Owens, D. K.; Cohen, D. J.; and Kuntz, K. M. 2012. State-transition modeling: a report of the ISPOR-SMDM modeling good research practices task force–3. *Medical Decision Making*, 32(5): 690–700.
- Speybroeck, N.; Van Malderen, C.; Harper, S.; Müller, B.; and Devleesschauwer, B. 2013. Simulation models for socioeconomic inequalities in health: a systematic review. *International Journal of Environmental Research and Public Health*, 10(11): 5750–5780.
- Tikka, S.; Hakanen, J.; Saarela, M.; and Karvanen, J. 2021. Sima—an open-source simulation framework for realistic large-scale individual-level data generation. *International Journal of Microsimulation*, (3).
- Verma, S.; and Rubin, J. 2018. Fairness definitions explained. In *Proceedings of the International Workshop on Software Fairness*, 1–7.
- Wen, M.; Bastani, O.; and Topcu, U. 2021. Algorithms for fairness in sequential decision making. In *International Conference on Artificial Intelligence and Statistics*, 1144–1152. PMLR.
- Whitehead, M. 1992. The concepts and principles of equity and health. *International Journal of Health Services*, 22(3): 429–445.
- Whitehead, S. J.; and Ali, S. 2010. Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*, 96(1): 5–21.
- Yang, D.; Cao, Z.; Byun, J.; Ban, J.-W.; and Kim, D. D. 2026. A Systematic Review of Microsimulation Models in Cardiometabolic Disease: Model Calibration and Validation. *Applied Health Economics and Health Policy*, 1–20.
- Zeckhauser, R.; and Shepard, D. 1976. Where now for saving lives. *Law & Contemp. Probs.*, 40: 5.