Appendix of "Quantifying Health Inequalities Induced by Data and AI Models"

Ethics Statement

Permission was granted by the data controllers to use the MIMIC-III and HiRID datasets. No personal data was processed in this study.

Algorithm 1: Left Boundary Adjustment

input : E: learned KDE;
lb: the lower bound;
ub: the upper bound;
t: value to adjust;
t_p: $\arg\max(\{v|v \in M : v < t\})$ when M is discrete and t is not boundary, otherwise t;
ε: a small constant like 1^{-10};
V: an empty array.
output: $\hat{t}$: the adjusted value for t

if $len(V) = 0$ **then**
  /* get an evenly spaced numbers between lb and ub with a relatively big number n, e.g., $n = 20 \times (ub - lb)$. */
  $a \leftarrow gen(lb, ub, n)$;
  $s \leftarrow (ub - lb)/n$;
  for $i \leftarrow 1$ **to** *len(a)* **do**
    $x_p \leftarrow lb$;
    if $i > 1$ **then**
      $x_p \leftarrow a[i-1]$;
    end
    $x \leftarrow a[i]$;
    $p \leftarrow exp(E(x))$;
    while $p \geq \varepsilon$ *and* $x > x_p$ **do**
      $x \leftarrow (x - s)$;
      $p \leftarrow exp(E(x))$;
    end
    if $exp(E(x)) < \varepsilon$ **then**
      $V.add(x)$;
    end
  end
end
$\hat{t} \leftarrow \arg\max(\{v|v \in V : v < t\})$;
if $\hat{t} \leq t_p$ **then**
  $\hat{t} \leftarrow t$;
end
return $\hat{t}$;

Algorithm 2: Approximate A-D Curve

input : P: the patient cohort;
d: the deterioration index function;
m: the measurement;
a: the AI model as an allocator;
l: a constant for smoothing;
n: a constant for specifying the number of points to be generated;
ν: a threshold for the minimal numbers of patients for deterioration estimation.
output: the curve

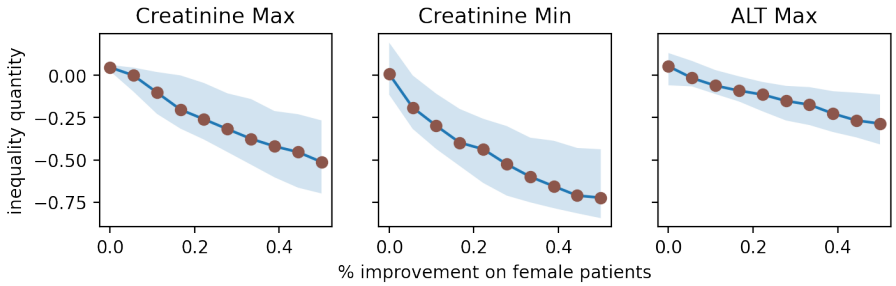
$C \leftarrow []$;
$X \leftarrow gen(0, 1, n)$;
for x *in* X **do**
  $\hat{P} \leftarrow \{p|p \in P : (x - l) \leq a(p) < (x + l)\}$;
  if $|\hat{P}| \geq \nu$ **then**
    $C.add((x, d(\hat{P}; m)))$;
  end
end
return C

Health inequality assessments on synthetic datasets

Measurement	mean [95% CI]	p-value
Creatinine max	0.044 [-0.083, 0.130]	0.0664
Creatinine min	0.024 [-0.266, 0.302]	0.7084
ALT max	0.033 [-0.157, 0.182]	0.4231

Table 3: Overall inequality of **female vs male** quantified on 10 synthetic datasets, where there should be no inequality overall.

Figure 4: Inequality Quantification Evaluation on synthetic data: y-axis is the inequality quantity of female vs male. x-axis is the percentage of controlled improvements on readings of the female subcohort. Y-value of each point is the mean value of 10 runs on the same x-value, i.e., % of improvement. Shaded areas denote 25-75% quantile regions.



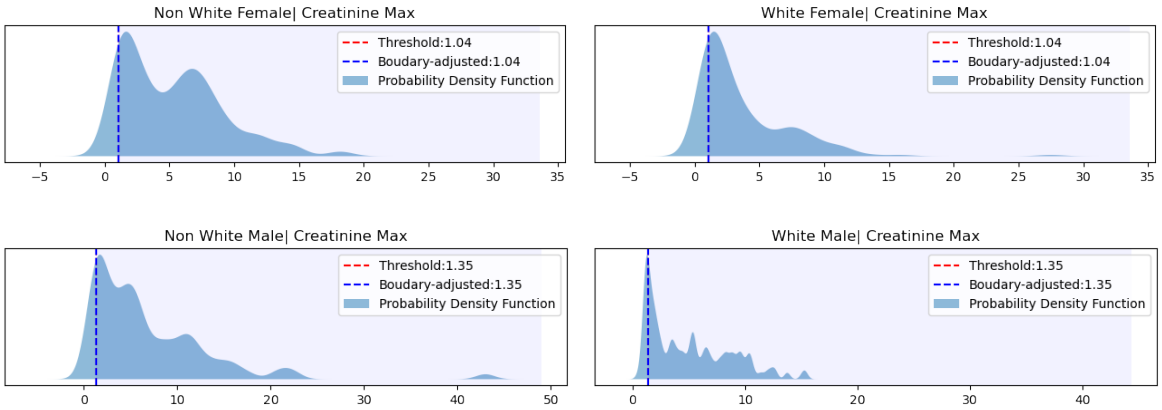


Figure 5: Probability density functions for quantifying inequalities of **non-White vs White** in the scenario of kidney operations in MIMIC-III dataset. Dashed lines denote thresholds (i.e., boundary values of abnormal readings) for computing deterioration index. Shaded area are regions where the probability integral happens for getting the deterioration index. The above two figures are females, which illustrate an inequality of 35.06%. The bottom two are males, where there is an inequality of 19.94%.

	Kidney operation				Renal Autotransplantation			
	Creatinine Max		Normalised MM		Creatinine Max		Normalised MM	
DB inequality	29.10%		7.62%		16.08%		2.58%	
Models	LR	RF	LR	RF	LR	RF	LR	RF
Inequality at Decision Region	**37.58%**	22.15%	**10.52%**	4.54%	9.13%	3.51%	2.45%	**23.36%**
Inequality at the whole area	16.17%	30.21%	-11.8%	9.65%	14.73%	22.70%	-26.10%	0.20%

Table 4: Inequality of **non-White vs White patients** channelled and exacerbated by AI models in two decision-making scenarios of kidney related operations in the MIMIC-III dataset. *DB inequality* row gives the DB embedded inequality quantities of relevant measurements. *Inequality at Decision Region* is the area between A-D curves within the region where a model suggesting surgery, while *Inequality at the whole area* is the area between two curves overall.

Attributes	Details
Feature List	['age', 'Chronic kidney disease', 'gender', 'Leukemia', 'cirrhosis', 'Infection']

Random Forest Hyper-parameters	tuned_parameters = {
	'n_estimators': [50, 100, 200],
	'max_depth': [5, 10, 20, 50]
	}

Logistic Regression Hyper-parameters	tuned_parameters = {
	'penalty': ['l1', 'l2'],
	'C': [#.001, .01,
	.1, 1, 10, 100, 1000],
	'max_iter': [100, 150],
	'solver': ['liblinear']
	}

Random state	1

Table 5: AI Model's hyperparameters and other reproducible setups

Renal Autotransplantation		
	Case	Control
N	146	438
Gender(male)	83 (56.8%)	286 (65.3%)
Age	53.31 [47.00-60.75]	53.47 [47.00-61.00]
Clinical attributes		
Length of Stay(days)	10.88 [6.00-14.00]	8.07 [3.00-11.00]
Death	5 (3.4%)	29 (6.6%)
CKD	145 (99.3%)	157 (35.8%)
Cirrhosis	25 (17.1%)	35 (8.0%)
Infection	37 (25.3%)	90 (20.5%)
Number of multimorbidities	4.27 [3.00-5.00]	2.83 [1.00-4.00]

Table 6: Characteristics of the study cohorts for the Renal Autotransplantation prediction task. The case cohort is identified from the MIMIC III database using ICD-9 code 55.61 and the control cohort is matched using similar age (+/- 3) with 1:3 ratio.

Operations on Kidney		
	Case	Control
N	584	1752
Gender(male)	293 (50.2%)	1,018 (58.1%)
Age	58.78 [49.00-69.00]	58.91 [49.00-70.00]
Clinical attributes		
Length of Stay(days)	10.43 [5.00-14.00]	8.24 [3.00-11.00]
Death	34 (5.8%)	165 (9.4%)
CKD	537 (92.0%)	665 (38.0%)
Cirrhosis	35 (6.0%)	117 (6.7%)
Infection	219 (37.5%)	366 (20.9%)
Number of multimorbidities	3.74 [2.00-5.00]	3.18 [1.00-5.00]

Table 7: Baseline Characteristics of the study cohorts for the Operations on Kidney prediction task. The case cohort is identified from the MIMIC III database using ICD-9 codes of 55.xx and the control cohort is matched using similar age (+/- 3) with 1:3 ratio.

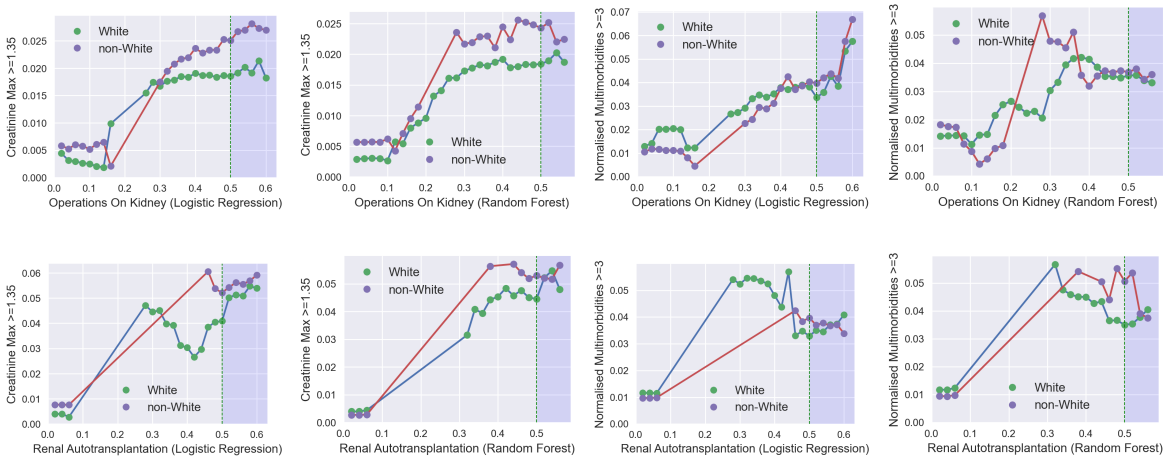


Figure 6: Allocation-Deterioration Indices of four models trained for predicting the needs of kidney related surgeries. The top row is for a generic *Operations on Kidney* and the bottom is for a particular *Renal Autotransplantation*. The left two columns are those using *deterioration index* defined on renal functions, while the right two are those using multimorbidities. In all cases, non-White patients are consistently more severe within the decision region (shaded area, allocation index > 0.5).