

Investigating Ethical Tradeoffs in Crisis Standards of Care through Simulation of Ventilator Allocation Protocols

Jonathan Herington^{1,2*}, Jessica Shand^{1,3}, Jeanne Holden-Wiltse⁴, Anthony Corbett⁴, Richard Dees², Chin-Lin Ching⁵, Margie Shaw¹, Xueya Cai⁶, Martin Zand^{4,7}.

¹ Department of Health Humanities and Bioethics, University of Rochester, Rochester, NY, USA

² Department of Philosophy, University of Rochester, Rochester, NY, USA

³ Department of Pediatrics, University of Rochester, Rochester, NY, USA

⁴ Clinical and Translational Sciences Institute, University of Rochester, Rochester, NY, USA

⁵ Department of Medicine, University of Rochester, Rochester, NY, USA

⁶ Department of Biostatistics, University of Rochester, Rochester, NY, USA

⁷ Division of Nephrology, Department of Medicine, University of Rochester, Rochester, NY, USA

* Corresponding author

Email: jonathan.herington@rochester.edu (JH)

Investigating Ethical Tradeoffs in Crisis Standards of Care through Simulation of Ventilator Allocation Protocols

Abstract

Introduction: Arguments over the appropriate Crisis Standards of Care (CSC) for public health emergencies often assume that there is a tradeoff between saving the most lives, saving the most life-years, and preventing racial disparities. However, these assumptions have rarely been explored empirically. To quantitatively characterize possible ethical tradeoffs, we aimed to simulate the implementation of five proposed CSC protocols for rationing ventilators in the context of the COVID-19 pandemic.

Methods: A Monte Carlo simulation was used to estimate the number of lives saved and life-years saved by implementing clinical acuity-, comorbidity- and age-based CSC protocols under different shortage conditions. This model was populated with patient data from 3707 adult admissions requiring ventilator support in a New York hospital system between April 2020 and May 2021. To estimate lives and life-years saved by each protocol, we determined survival to discharge and estimated remaining life expectancy for each admission.

Results: The simulation demonstrated stronger performance for age- and comorbidity-sensitive protocols. For a capacity of 1 bed per 2 patients, ranking by age bands saves approximately 28.7 lives and 3408 life-years per thousand patients, while ranking by Sequential Organ Failure Assessment (SOFA) bands saved the fewest lives (13.2) and life-years (416). For all protocols, we observed a positive correlation between lives saved and life-years saved. For all protocols except lottery and the banded SOFA, significant disparities in lives saved and life-years saved were noted between White non-Hispanic, Black non-Hispanic, and Hispanic sub-populations.

Conclusion: While there is significant variance in the number of lives saved and life-years saved, we did not find a tradeoff between saving the most lives and saving the most life-years. Moreover, concerns about racial discrimination in triage protocols require thinking carefully about the tradeoff between enforcing equality of survival rates and maximizing the lives saved in each sub-population.

1.0 Introduction

The allocation of scarce resources during public health emergencies is often presumed to require balancing several high-level ethical goals. For scarce resource allocation, potential ethical goals may include (i) saving the most lives,(1) (ii) saving the most life-years,(2,3) (iii) respecting principles of non-discrimination,(4) or (iv) promoting health equity.(5) Crisis Standard of Care (CSC) protocols, which allocate scarce hospital resources in the setting of public health emergencies, are one such area of policymaking. Published and widely recognized protocols, such as the New York State (NYS) Ventilator Allocation Guidelines, the Pennsylvania Crisis Standards of Care, and the Colorado Crisis Standards of Care Plan (1,3,6) are taken to encode judgements about how to weight competing ethical values. The NYS guidelines use the Sequential Organ Failure Assessment (SOFA) with the stated goal of maximizing survival to discharge,(1) while the Colorado guidelines use SOFA plus quantitative adjustments for comorbidities to maximize life-years saved.(6) These protocols make an implicit assumption that the chosen clinical criteria will promote the intended ethical goal. Moreover, they assume that tradeoffs exist between goals--for example between saving the most lives and ameliorating health inequities. Because these policies are implemented in complex, dynamic systems, these assumptions may turn out to be false. Modelling of COVID-19 vaccine allocation strategies, for example, suggested no tradeoff between vaccination strategies that saved the most lives and life-years because of the strong association between age and COVID-19 mortality.(7) Rigorous and reproducible methodologies to test ethical assumptions is therefore essential if CSC protocols are to serve their intended ethical goals.

Recent work (8–10) has attempted to simulate the implementation of CSC protocols in the setting of COVID-related scarcity. Miller et. al.(9) and Wunsch et. al.(10) assessed the predictive value and potential racial bias of popular protocols, but used a convenience data set of non-COVID positive individuals without simulating demand or resource scarcity. Bhavani et. al.(8) used simulation methods to

investigate allocation protocols in the setting of scarcity, but simulated triage only for COVID+ patients and did not estimate post-discharge life expectancy. Finally, Kim et. al.(11) used a discrete event simulation approach, with accompanying demand models for COVID-19 patients, to train a bespoke machine learning model for prioritizing critical care beds, but they did not assess the performance of proposed triage protocols. None of the previously reported simulation methods attempted to evaluate the tradeoff between saving the most lives to discharge and saving the most life-years.

We expand on this work to develop a simulation of CSC protocols that estimates tradeoffs between lives saved, life-years saved, and equal allocation of resources amongst racial groups. We ground our analysis in a more realistic context than prior work by evaluating previously published and adopted CSC protocols in the setting of a mixed population of COVID+ and COVID- patients. Our hope is that our application of quantitative simulation to a real-world clinical dataset can be broadly adapted to other bioethical debates about the appropriate mechanism for resource allocation in health systems.

2.0 Methods

We sought to simulate the implementation of five proposed ventilator allocation protocols: 1) lottery; 2) age-banding; 3) pure SOFA 4) SOFA bands (New York), 5) SOFA plus comorbidity (Colorado), and 6) a different SOFA plus comorbidity model, with age as a tie breaker (Bhavani). The age-banding protocol assigns priority by 10-year age bands, with ties broken by lottery. The NY SOFA protocol assigns priority tiers by SOFA score (SOFA ≤ 7 = Tier 1; SOFA 8-11 = Tier 2; SOFA ≥ 12 = Tier 3), with ties broken through a lottery. The Colorado protocol assigns points by SOFA band (≤ 5 SOFA = 1 pt; 6-9 SOFA = 2 pts; 10-12 SOFA = 3 pts, > 12 SOFA = 4 pts), and by a modified Charlson comorbidity index, with ties broken by lottery.(6) The Bhavani protocol (8) assigns points by SOFA bands (< 9 SOFA = 1 pt; 9-11 SOFA = 2 pts; 12-14 SOFA = 3 pts; > 14 SOFA = 4 pts) and adds +3 points to the SOFA score of any patient with a “severe” comorbidity, understood as a van Walraven acute Elixhauser score > 4 (i.e. the upper tertile of Elixhauser scores with a “90% predicted 1 year mortality” in (12)). Ties between patients in the Bhavani protocol are then broken by course-grained age-bands (0-49, 50-69, 70-84, ≥ 85) and then by lottery. To simplify this study, we ignored protocol instructions that involved re-assessment and re-assignment of ventilators.

Our study population included all critically-ill adult patients who received mechanical ventilation between April 2020 and May 2021 within the University of Rochester Medical Center system (including three hospitals in the Rochester metropolitan area). The URM Coronavirus Ethics Response Group (CERG) convened in March 2020 to adapt the 2015 NYS Ventilator Allocation Guidelines into a usable triage algorithm.⁽¹³⁾ Using the Clinical Translational Science Institute BLIS database and in collaboration with hospital IT, the group created a robust data workflow that allowed for the extraction and database capture of clinical information from the electronic medical record system at 30 minute intervals on all inpatient that could then be used as the basis for allocation decisions. For each patient, the data included their ventilator status, their sequential organ failure (SOFA) score, and their admission date. Through a data broker, we received deidentified data from the database combined with additional data for COVID infection status, age, sex, self-identified race and ethnicity, acute comorbidities (ICD-10 diagnosis), and discharge status from electronic health records for all subjects. Missing values relevant to SOFA score calculation within the first 24 hours of ventilation were assumed to be normal (rather than imputed), aligning with common practice and the necessity for timely data in a resource shortage.^(10,14)

Our study involved secondary re-use of a de-identified version of this dataset. It was deemed by the University of Rochester Research Subject Review Board as exempt from review and a waiver of informed consent was granted. The dataset was fully anonymized before being accessed by the study team, and was provided to us on the 15th of January 2022.

4604 admissions involving mechanical ventilation, from 4147 unique patients, were identified during the study period. Of these, 897 (19.5%) were excluded because they (i) lacked the lab results required to calculate an initial SOFA score within ± 24 hours of their initial intubation (n=320), (ii) had no discharge disposition, ICD-10 codes or COVID test data (n=245), or (iii) had an unclear pattern of admissions and intubations (n=332). The distribution of age, sex, race and COVID positivity amongst the excluded patients was similar to the included patients. There was no significant difference in survival rates amongst the excluded (72%) and included (72%) encounters (Table S5). Demographic features of the final modelling cohort of 3707 encounters from 3512 unique patients are described in Table 1.

Table 1: Descriptive Statistics for Mechanically Ventilated Adult Patients, March 2020 to April 2021

Feature	n	%	Survival % (CI)	Age-Adjusted Survival % (CI)
Overall	3707		72 (71-74)	84 (82-86)
Sex				
Male	2197	59	73 (71 - 75)	84 (80 - 87)
Female	1510	41	72 (69 - 74)	84 (80 - 86)
Age				
<25	104	03	93 (86 - 97)	
25-34	235	06	90 (85 - 93)	
35-44	247	07	83 (77 - 87)	
45-54	456	12	81 (77 - 85)	
55-64	868	23	76 (73 - 79)	
65-74	980	26	69 (66 - 73)	
75-84	629	17	60 (54 - 64)	
>85	188	05	50 (38 - 59)	
Race				
AAPL, non-Hispanic	50	01	78 (61 - 89)	87 (24 - 97)
AIAN, non-Hispanic	3	>01	1.0	1.0
Black, non-Hispanic	577	16	76 (72 - 80)	82 (74 - 87)
Hispanic	152	04	82 (73 - 88)	89 (78 - 94)
White, non-Hispanic	2752	74	71 (69 - 74)	85 (81 - 87)
>1 Race, non-Hispanic	16	>01	75 (36 - 93)	87 (23 - 96)
Unknown, non-Hispanic	157	04		79 (81 - 87)
COVID Status				
Negative	2431	66	73 (71 - 75)	84 (81 - 86)
Positive	1276	34	71 (68 - 74)	83 (76 - 88)
SOFA Band				
1-7	3192	86	76 (74 - 77)	87 (84 - 89)
8-11	372	10	56 (49 - 63)	58 (15 - 80)
11-24	143	04	48 (34 - 59)	75 (00 - 83)

We simulated the implementation of each protocol under different levels of scarcity using a Monte Carlo method. Extending the approach of Bhavani et al.,(8) our model simulates an $n/20$ (i.e. 5%, 10%...95% capacity) ventilator shortage by (1) randomly sampling twenty patients $\langle A, B, \dots, T \rangle$ from our dataset, and assigning each patient to one of ten decision “pairs” randomly numbered 1 through 10 (e.g. 1:[A, R], 2:[H, C]...10: [Q, L]) , (2) ranking the patients in each pair for priority based on the relevant protocol (e.g. 1:[A > R], 2:[C > H]...10:[Q>L]), (3) allocating beds to both patients in the first $n-10$ pairs (i.e. if $n=15$, then assign [A=1, R=1], [C=1, H=1] etc. to first five pairs), and not allocating beds to either patient in the last $10-n$ pairs (i.e. if $n=15$, then no pairs miss out on beds), 4) allocating a bed to the highest priority patients in each remaining pair (e.g. [Q=1, L=0]), and 4) repeating this process until all patients in our dataset have been allocated (or not) a bed. We estimated survival in each simulation by

observing the actual (scarcity-free) survival to discharge of patients assigned beds. We assumed that patients who were not allocated a bed in our simulation did not survive. We repeated the simulation 1,000 times for the 50% capacity results and 250 times for all other capacity levels.

For each simulation, we calculated (i) survival rate, (ii) age-adjusted survival rate, and (iii) aggregate comorbidity-adjusted life expectancy (in years). Due to COVID's severe age-associated mortality and the distribution of age within racial groups, survival rates for sub-populations were age-adjusted, and confidence intervals for both raw and age-adjusted rates were calculated using the modified Gamma method of Tiwari et al.(15,16) Raw life expectancy was calculated for each subject from the corresponding National Vital Statistics System (NVSS) life tables for their age, sex and race. The impact of comorbidities on life expectancy was estimated by applying the adjustments previously calculated by Cho et. al for Medicare recipients without a history of cancer.(17) Subjects were placed into one of three comorbidity bands (none, low/medium, high) and the corresponding age adjustment in Cho et. al was applied, before re-calculating the subject's life expectancy using the NVSS tables. The comorbidity adjustment for a 65yr old in Cho et. al was applied to all subjects in our cohort <65yrs (we explore the limitations of this approach in the discussion).

We then defined "lives saved" and "years of life saved" per patient for each protocol, p , and beds-per-patient capacity level, $(0 < c < 1)$, by first identifying the survival rate, S_B , and expected years of life lived post-discharge per patient, LY_B , of our population in the baseline no scarcity scenario. Next, for each capacity level, we defined the expected survival rate ($S_B \times c$) and expected years of life per patient ($LY_B \times c$) as the product of the baseline rates and the relevant capacity constraint (e.g. 0.5 beds per patient). Lives saved per patient, $LS_{p,c}$, is the difference between the simulated survival rate for the protocol and the expected survival rate for that capacity. Years of life saved per patient, $LYS_{p,c}$, is the difference between the simulated number of life-years lived per patient for that protocol, and the expected number of life-years lived for that capacity.

All data analysis was performed in Python v3.11 using the pandas, matplotlib, numpy, scipy, seaborn and statsmodels packages. De-identified data is available at

<https://figshare.com/s/fb7e6ac1702b5e9a95fd> . Simulation and analysis code is available at <https://github.com/jcherington/triage-sim>.

3. Results

Our population of 3707 ventilator encounters with 3512 unique subjects had a mean (\pm std) age of 62 (± 16) (Table 1). Among them, Black and Hispanic patients were younger (56 ± 17 and 53 ± 18) compared to non-Hispanic White patients (64 ± 15). Black patients were more likely (rate [95% CI]) to be diagnosed with COVID (.45 [.41-.49]), compared to Hispanic (.39, [.32-.47]) and non-Hispanic White patients (.33, [.31-.34]). In addition, SOFA scores at the initial point of intubation were higher for White subjects (mean \pm std: 3.69 ± 0.07) than for both Black and Hispanic subjects (3.46 ± 0.15 and 3.14 ± 0.29).

Contrary to prior reported results(5,8,10) in our baseline population, we did not observe significant racial disparities in age-adjusted survival amongst White, Black and Hispanic patients who received mechanical ventilation (Table 2). Neither did we observe significant disparities in age-adjusted survival (rate [95% CI]) for patients who were COVID positive (.84 [.81-.87]) and negative (.83 [.76-.88]). In a multiple variable logistic regression model, SOFA score at the point of initial intubation (OR: .896, [.878-.916]), age in years (OR: .964, [.958-.969]), and Elixhauser score (OR: 1.054, [1.033-1.074]) were all predictive of survival, whereas sex, race and COVID-positivity were not significantly predictive (Table S7).

3.1 Allocation Inequalities

When simulating a shortage of ventilators where only one bed is available for two patients ($c=0.5$), and without controlling for other factors (i.e. clinical acuity), all protocols except Lottery exhibited significant racial disparities in allocation of ventilators (Table 2). Contrary to prior findings with simulated populations of COVID positive patients,(8) in our simulations with mixed populations of COVID positive and negative patients, allocation of ventilators favored non-Hispanic Black and Hispanic patients when compared to non-Hispanic Whites (Fig S9). Allocation disparities were most pronounced for the Age-based protocol and least pronounced in the context of the NY SOFA protocol.

3.2 Survival and Lives Saved

In a simulation of moderate scarcity (0.5 beds per patient), the Age-based protocol had significantly higher overall survival rates (.3916 [.3914-.3918]) than any other protocol (Table 2). On the other hand, the NY SOFA (.3760 [.3758-.3762]) protocol resulted in lower overall survival rates than all other protocols except Lottery (.3629 [.3727-.3631]), both in overall and in age-adjusted terms for every racial/ethnic sub-population. The Bhavani, Colorado, and Pure SOFA protocols exhibited significantly lower rates of age-adjusted survival for White patients compared to both their Black and Hispanic counterparts, with the biggest racial inequalities in age-adjusted survival occurring for the Colorado protocol (White: .488, Black: .524). The Age-based and Lottery protocols exhibited significantly lower age-adjusted survival for Black patients, compared to the White or Hispanic population. Unsurprisingly, the Age-based protocol heavily skewed survival towards younger Age groups, while NY SOFA and Lottery resulted in roughly similar distributions of survival across age-groups (Table S8 and Fig S10).

Table 2: Allocation and Survival by Protocol and Race and Ethnicity.

Protocol	Overall Survival (%)	Allocation Rate by Race and Ethnicity (%)			Age-Adjusted Survival Rate (%)		
		Non-Hispanic, Black	Hispanic, All Races	Non-Hispanic, White	Non-Hispanic, Black	Hispanic, All Races	Non-Hispanic, White
Baseline (100% capacity)	72.57 (70.83-74.23)	100	100	100	82.3 (74.5-87.4)	89.1 (78.4-94.1)	85 (81.4-87.4)
Lottery	36.29 (36.27-36.31)	50.1 (50.0-50.2)	50.3 (50.1-50.5)	50.0 (49.9-50.0)	41.1 (40.9-41.3)	45.0 (44.7-45.4)	42.5 (42.3-42.6)
Age	39.16 (39.14-39.18)	59.7 (59.6-59.8)	64.7 (64.5-64.9)	46.6 (46.6-46.7)	68.2 (68.1-68.2)	73.6 (73.5-73.8)	71.2 (71.1-71.2)
Pure SOFA	37.96 (37.94-37.97)	51.8 (51.7-51.9)	55.3 (55.1-55.5)	49.7 (49.7-49.7)	51.2 (51.0-51.4)	50.4 (50.1-50.7)	48.3 (48.2-48.5)
NY SOFA	37.60 (37.58-37.62)	50.5 (50.3-50.6)	51.0 (50.8-51.3)	50.0 (50.0-50.1)	44.5 (44.3-44.7)	47.1 (46.8-47.4)	44.5 (44.3-44.7)
Colorado	37.76 (37.74-37.77)	52.0 (51.9-52.1)	55.7 (55.5-55.9)	49.6 (49.6-49.6)	52.4 (52.2-52.5)	51.2 (51.0-51.5)	48.8 (48.7-48.9)
Bhavani	38.08 (38.06-38.10)	53.3 (53.2-53.4)	58.0 (57.8-58.2)	49.1 (49.1-49.2)	55.2 (55.0-55.3)	54.8 (54.6-55.0)	52.7 (52.6-52.8)

*Mean overall survival rate and 95% CIs are reported for 1000 Monte Carlo simulations at 50% scarcity.

Because survival rates are heavily influenced by the distribution of survival amongst sub-populations in our underlying cohort, we also estimate the number of lives saved by each protocol: i.e. the additional

number of lives saved by employing a protocol instead of using a lottery (Table 3 and Fig 1). In a simulation of $c=0.5$ beds per patient, the Age-based protocol saved more than twice as many lives per thousand patients (28.7 [28.6-28.9]) as the nearest competing protocol, Pure SOFA (16.7 [16.5-16.9]). Again, the NY SOFA protocol (13.2 [13.0-13.4]) saved the least lives per thousand patients of any protocol, both overall and for the Black (16.5 [15.4-17.5]) and Hispanic (9.9 [7.8-12.1]) sub-populations. This effect was reversed for the White sub-population, for whom the NY SOFA protocol saved the most lives (13.6, [13.3-13.9]) and the Age protocol the least (4.3 [4.0-4.6]). All protocols except NY SOFA exhibited significantly lower numbers of lives saved for White patients compared to both their Black and Hispanic counterparts, with the biggest racial inequalities in lives saved occurring for the Age protocol (White: 4.3 [4.0-4.6], Black: 99.9 [99.0-100.7], Hispanic: 143.4 [141.7-145.1]).

Table 3: Lives Saved by Protocol and Ethnicity/Race at 50% capacity

Protocol	Lives Saved per 1000 patients (95% CI)			Overall
	Non-Hispanic, Black	Hispanic, All Races	Non-Hispanic, White	
Age	99.9 (99.0-100.7)	143.4 (141.7-145.1)	4.3 (4.0-4.6)	28.7 (28.6-28.9)
Pure SOFA	40.8 (40.0-41.7)	46.3 (44.5-48.1)	12.5 (12.2-12.7)	16.7 (16.5-16.9)
NY SOFA	16.5 (15.4-17.5)	9.9 (7.8-12.1)	13.6 (13.3-13.9)	13.2 (13.0-13.4)
Colorado	41.3 (40.4-42.2)	48.3 (46.5-50.1)	9.2 (8.9-9.5)	14.7 (14.5-14.9)
Bhavani	41.3 (40.4-42.2)	48.3 (46.5-50.1)	9.2 (8.9-9.5)	18 (17.8-18.2)

Fig 1. Lives saved per patient by protocol: A. Overall increase in lives saved per patient, for each protocol, at 50% scarcity. **B.** Lives saved per patient stratified by racial group, for each protocol, at 50% scarcity.

After simulating different capacity shortages, we noted that the magnitude of the dominance of the Age protocol was sensitive to the degree of scarcity (Table S9 & Figure 2). At 0.5 beds per patient the Age based protocol saves approximately 28.7 additional lives per 1000 patients (relative to the Lottery

protocol), while at 0.9 beds/patient, the protocol saves only 5.6 lives per 1000 patients. The differences between NY SOFA, Colorado and the Bhavani protocols at capacities of greater than 0.85 beds per patient are not statistically significant (Table S9 & Figure 2).

Fig 2. Lives saved at different levels of scarcity, for each protocol. The greatest differences between protocols occur at moderate levels of scarcity (i.e. ~0.5 beds per patient), and differences between protocols decline at both high and low levels of scarcity.

3.3 Life-years saved

With respect to life-years saved, for a scarcity level of 0.5 beds per patient, the Age-based protocol saved more than twice as many life-years per thousand patients (3408 [3402-3413]) than the nearest competing protocol, Bhavani (1454 [1447-1461]) (Table 4 and Fig 3). The NY SOFA protocol saved fewer life-years than all other protocols, both overall (416 [407-424]) and for every racial/ethnic sub-population (White: 377 [366-388], Hispanic: 450 [366-534], Black: 626 [591-661]). All five non-random protocols resulted in significantly fewer life-years saved for White patients when compared to both their Black and Hispanic counterparts (Table 4 and Fig 3). After simulating different capacity shortages, we again noted that the difference between Age and other protocols was sensitive to the degree of scarcity, but less so than lives saved (Table S9). For instance, at 0.9 beds/patient, the Age protocol saves only 681 [674-688] life-years per thousand patients, but continue to outperform its nearest competing protocol, Pure SOFA (299 [290-308]), by a factor of two (Table S9).

Table 4: Life-years Saved by Protocol and Ethnicity/Race at 50% capacity

Protocol	Life-years Saved per 1000 patients (95% CI)			Overall
	Non-Hispanic, Black	Hispanic, All Races	Non-Hispanic, White	
Age	5517 (5494-5540)	8531 (8478-8584)	2553 (2545-2561)	3408 (3402-3413)
Pure SOFA	1957 (1928-1985)	2245 (2172-2317)	584 (574-592)	839 (831-846)
NY SOFA	626 (591-661)	450 (366-534)	377 (366-388)	416 (407-424)

Colorado	2226 (2198-2253)	2596 (2529-2662)	698 (689-708)	999 (992-1007)
Bhavani	2916 (2890-2942)	3748 (3688-3808)	1048 (1040-1057)	1454 (1447-1461)

Fig 3: Life years saved per patient. A. Overall increase in life years saved per patient, for each protocol, at 50% scarcity. B. Life years saved per patient stratified by racial group, for each protocol, at 50% scarcity.

4.0 Discussion

In this simulation study, we found that Age-based protocols significantly outperformed all other protocols with respect to both lives saved and life-years saved. At moderate levels of scarcity (i.e. 0.5 beds per patient), we estimate that selecting Age over Lottery could save approximately 29 lives per thousand patients requiring ventilation. When scaled to a national level, this level of effectiveness compares very favorably to classic public health interventions such as seatbelt use (14,955 lives saved per year)(18) or flu vaccination (7,200 lives saved from 36 million infections in 2020, or 0.2 lives saved per 1000 infections).(19) Because age-based protocols discriminate on the basis of age, they have an uncertain legal status.. Among the other four protocols, the difference between the best (Bhavani) and worst performing (NY SOFA) protocols is less pronounced but still significant, saving approximately 5 lives and 1040 life-years per thousand patients at moderate levels of scarcity.

Notably, the level of scarcity appears to have a dramatic influence on the effect size of choosing different protocols. While the Age based protocol dominates the other protocols for both severe (0 – 0.2 beds per patient), moderate (0.4-0.6 beds per patient) and low levels of scarcity, the size of the effect diminishes substantially at low and severe scarcity. At very low levels of scarcity (e.g. 0.9 beds per patient), choosing Age over NY SOFA still saves 2.9 lives per 1000 patients. The differences between NY SOFA, Pure SOFA, Colorado and Bhavani at that level of scarcity are not statistically significant, suggesting that other ethical considerations such as complexity of implementation, transparency, and social acceptability may be decisive.

One possible explanation for the poor performance of the NY SOFA protocol is that it has many fewer prioritization bands (i.e. only 3) than the other protocols (Age = 8 bands, Colorado = 10 bands, Pure SOFA = 23 bands, Bhavani = 28 bands). In principle, more bands add greater discriminatory power to a protocol and so enhance its ability to save lives at the margins. This is borne out by the significant but small improvement offered by the Pure SOFA protocol. This empirical result suggests that the value of treating clinically similar cases alike (by having relatively “wide” priority bands) must be weighed against the value of maximizing the overall numbers of lives saved.

4.1 Tradeoffs Between Lives and Life-Years

Much prior work has assumed that there is an empirical tradeoff between lives saved and numbers of life-years saved, thus necessitating a difficult normative decision about prioritizing different values in the context of disagreement.⁽²⁾ In our simulations, that tradeoff is not evident. Not only does the Age-based protocol save the most lives, it also saves the most life-years. Indeed, this result appears to be robust for all protocols. While there is a high degree of variance in both survival and longevity within each protocol, protocols which save more lives on average also save more life-years on average.

This result call into question the assumed tension between maximizing survival to discharge and maximizing life-years saved, but it also illuminates the different justifications one might have for focusing on the distinction between lives and life-years. For utilitarians, who aim simply to maximize the aggregate good, this result suggests that the distinction between lives and life-years is empirically unimportant when choosing a pandemic policy. For those who hold a “fair innings” view for the importance of allocating ventilators to younger patients, this result is less important – since saving people with long life expectancies is an ancillary result of prioritizing patients who have yet to progress through the different stages of life. Alongside the performance of the Age protocol, this result shows that prioritizing the young is at least compatible with maximizing the number of lives saved.

Fig 4: Tradeoffs between life years saved and lives saved. While there is significant variance in numbers of lives and life years saved, there is a strong positive correlation between each statistic for all protocols.

4.2 Tradeoffs between Performance and Racial Disparities

Of those protocols examined, the protocol that offered the most equal distribution of survival rates over the three racial groups had the least favorable overall performance (excluding a simple Lottery). In our simulations, at all levels of scarcity and for all protocols except the NY SOFA and Lottery, the White sub-population experienced significantly lower allocation rates, lives saved, and life-years saved. This finding is in *prima facie* tension with previously reported simulations of triage protocols with COVID-19 positive cohorts,(8) pre-pandemic cohorts,(9,10) and with the lived experience of actual COVID mortality in intensive care units.(5,20) Age-adjusted rates of survival remained significantly lower for White patients, and thus it is unlikely that the observed racial disparities are solely explained by the different age distributions amongst White non-Hispanic, Hispanic and Black non-Hispanic patients in our sample. Rather, we suspect that hospital-level population bias may explain this difference. In particular, there are two major hospital systems that offer adult critical care services in our student cohort's metropolitan area, and we cannot assume that the distribution of patients between these systems is the same for each sub-population. Hospitals serve particular patient populations, not national averages. This highlights that implementation of protocols may result in different aggregate outcomes in different contexts (i.e. with national-level results coming apart from regional or hospital level outcomes).

More interestingly, our results suggest a tradeoff between equalizing lives saved across racial groups, and maximizing the number of lives saved for each racial group. In our results, protocols other than NY SOFA greatly improve the number of lives saved for Black and Hispanic populations while making only marginal differences to the number of lives saved for the White population. Thus, while NY SOFA secures equality in survival rates, it would deprive Black and Hispanic populations of very large gains for little benefit in lives or life-years saved for White patients. Choosing NY SOFA in such cases would thus be a classic case of 'levelling down' – a situation where inequality makes little difference to the absolute welfare of the worst off group, but could markedly improve the welfare of other groups.(21) Notably, this tradeoff is only observable by simulating the set of alternative policies in a single population, and hence discussion of the value of equality in survival rates benefits from the use of simulation methods.

4.3 Limitations

This study has a number of important limitations. The first is that our dataset is limited to a single hospital system in a region with two major healthcare institutions. As we note above, this may partially explain our surprising findings regarding racial disparities in age-adjusted survival, because Black and Hispanic patients with poorer social determinants of health may have been more likely to have been admitted elsewhere. In future research, we hope to extract a region-wide dataset that eliminates hospital-level effects.

Another limitation is that we did not collect, nor adjust for, the existence of a “do-not-resuscitate” order or a decision to undertake compassionate withdrawal. In principle, some patients subject to DNRs may have otherwise survived their admission and, given well characterized racial differences in the uptake of DNR orders, this may have artificially depressed survival rates for White patients relative to other racial sub-populations.⁽²²⁾ While this is true for survival rates, it is less likely to distort differences in the lives saved and life-years saved statistics since it accounts for variation in actual mortality (including withdrawal of treatment and natural expiration) and estimates deviations from “expected” survival rates for each population.

A final limitation of our simulation is that it does not allow for “re-allocation” of ventilators. Most existing protocols (including New York’s, Maryland’s and Colorado’s) allow patients a “timed trial” of ventilation, and contemplate mechanisms for removing ventilators from patients who are not improving at specified re-assessment timepoints.⁽²³⁾ The criteria for re-allocation (and the timing of re-assessments) must therefore strike a balance between giving each individual a chance to recover and preventing “wastage” of ventilators on patients who will not recover.⁽²⁴⁾ In this respect, the criteria for reallocation may have significant effects on the survival of individual patients and the overall numbers of lives saved.

Unfortunately, our existing dataset and simulation strategy cannot accommodate reallocation of ventilators, since it only includes SOFA scores collected at point of intubation, and thus envisages allocation until survival or natural expiration. Future work should incorporate time-series datasets of patient trajectories while intubated, and simulation designs that allow for “pools” of ventilators to be allocated and re-allocated. We leave this extension for future work.

5.0 Conclusion

In this paper we have empirically tested a number of assumptions that have been at the forefront of bioethical debates over the appropriate CSC protocols. In our simulation using actual patient data, we show several findings. First, there are clear performance differences between protocols, and age- or comorbidity-sensitive protocols appear to save more lives to discharge than protocols which rely on SOFA as an indicator of clinical acuity. Second, using comorbidity-adjusted estimates of post-discharge life expectancy, we show that there is unlikely to be a tradeoff between saving lives and saving life-years in the aggregate. Third, we identify that while there was tradeoff between equalizing lives saved between sub-populations and overall performance, those protocols with high levels of inequality were sometimes better for the least racial groups than protocols with more equal distributions of lives saved. These three findings cut to the heart of important bioethical debates, and should inform both philosophical and implementation work to improve CSC policy.

Moreover, we have reproduced and extended a methodology for prospectively and retrospectively analyzing the performance of different CSC protocols, under different levels of scarcity. Building on earlier work by Bhavani et al, we show not only that these simulation methods can be used to estimate survival, but also that they can also provide information about estimated life expectancy, inequalities between sub-populations, and the optimal parameterization of models (i.e. numbers of bands, precise threshold for clinical acuity scores). Nonetheless, the work of empirically exploring CSC protocols is not complete. So far, no simulation protocol has been generated that includes (i) all critical care patients, (ii) is sensitive to duration of resource use, and (iii) accommodates re-allocation of resources. These components are critical to many currently-existing CSC protocols, and they may yield important empirical results that alter or contradict the results from the static allocation simulations conducted here and elsewhere.(8) Without a generalizable methodology for either post hoc or ex ante testing of the ethical assumptions underpinning re-allocation of resources, we will continue to lack evidence that future allocation policies achieve their intended ethical goals. We hope that the preliminary work we present in this paper spurs others to develop more robust, dynamic models that can inform CSC policy choices for the next public health emergency.

Acknowledgements

The authors would like to acknowledge the efforts of Data Science Masters' students - Derek Caramella, Ezgi Siir Kibris, Nefle Nesli Oruc, and Walter Burnett - for assisting with initial exploration of the dataset under the supervision of JH. Final data acquisition, cleaning, analysis and simulation work was conducted by JH. The authors would like to thank Michael Nabozny, William Parker, Lainie Ross and the members of the URMCAfrican American Patient and Family Advisory Council for comments on earlier versions of the project. The project described in this publication was partially supported by the University of Rochester CTSA award number UL1 TR002001 from the National Center for Advancing Translational Sciences of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. New York State Task Force on Life and the Law. Ventilator Allocation Guidelines [Internet]. New York, N.Y.: NYS Department of Health; 2015 Nov [cited 2021 Jul 4]. Available from: https://www.health.ny.gov/regulations/task_force/reports_publications/docs/ventilator_guidelines.pdf
2. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair Allocation of Scarce Medical Resources in the Time of Covid-19. *New England Journal of Medicine*. 2020 Mar 23;382(21):2049–55.
3. White DB, Lo B. A Framework for Rationing Ventilators and Critical Care Beds During the COVID-19 Pandemic. *JAMA*. 2020 May 12;323(18):1773.
4. Silva DS. Ventilators by Lottery: The Least Unjust Form of Allocation in the Coronavirus Disease 2019 Pandemic. *Chest*. 2020 Sep;158(3):890.
5. Schmidt H, Roberts D, Eneanya N. Rationing, racism and justice: advancing the debate around "colourblind" COVID-19 ventilator allocation. *Journal of medical ethics* [Internet]. 2022 Feb [cited 2022 Mar 7];48(2). Available from: <https://pubmed.ncbi.nlm.nih.gov/33408091/>
6. Department of Public Health & Environment. Colorado Crisis Standards of Care [Internet]. 2022 [cited 2023 Aug 1]. Available from: <https://cdphe.colorado.gov/colorado-crisis-standards-care>
7. Goldstein JR, Cassidy T, Wachter KW. Vaccinating the oldest against COVID-19 saves both the most lives and most years of life. *Proc Natl Acad Sci U S A*. 2021 Mar 16;118(11):e2026322118.
8. Bhavani SV, Luo Y, Miller WD, Sanchez-Pinto LN, Han X, Mao C, et al. Simulation of Ventilator Allocation in Critically Ill Patients with COVID-19. *Am J Respir Crit Care Med*. 2021 Nov 15;204(10):1224–7.

- 379 9. Miller WD, Han X, Peek ME, Charan Ashana D, Parker WF. Accuracy of the Sequential Organ
380 Failure Assessment Score for In-Hospital Mortality by Race and Relevance to Crisis Standards of
381 Care. JAMA Network Open. 2021 Jun 18;4(6):e2113891.
- 382 10. Wunsch H, Hill AD, Bosch N, Adhikari NKJ, Rubenfeld G, Walkey A, et al. Comparison of 2 Triage
383 Scoring Guidelines for Allocation of Mechanical Ventilators. JAMA Network Open. 2020 Dec
384 14;3(12):e2029250–e2029250.
- 385 11. Kim J, Lim H, Ahn JH, Lee KH, Lee KS, Koo KC. Optimal Triage for COVID-19 Patients Under
386 Limited Health Care Resources With a Parsimonious Machine Learning Prediction Model and
387 Threshold Optimization Using Discrete-Event Simulation: Development Study. JMIR Med Inform.
388 2021 Nov 2;9(11):e32726.
- 389 12. Snow GL, Bledsoe JR, Butler A, Wilson EL, Rea S, Majercik S, et al. Comparative evaluation of the
390 clinical laboratory-based Intermountain risk score with the Charlson and Elixhauser comorbidity
391 indices for mortality prediction. PLoS One. 2020;15(5):e0233495.
- 392 13. University of Rochester Coronavirus Ethics Response Group. Ventilator Allocation Guidelines
393 [Internet]. Rochester, NY: University of Rochester Medical Center; 2020 May [cited 2022 Mar 9].
394 Available from:
395 [https://www.raom.org/resources/Documents/URMedicine%20Ventilator%20Allocation%20Guidelines](https://www.raom.org/resources/Documents/URMedicine%20Ventilator%20Allocation%20Guidelines_2020.pdf)
396 [_2020.pdf](https://www.raom.org/resources/Documents/URMedicine%20Ventilator%20Allocation%20Guidelines_2020.pdf)
- 397 14. Vincent JL, de Mendonca A, Cantraine F, Moreno R, Takala J, Suter PM, et al. Use of the SOFA
398 score to assess the incidence of organ dysfunction/failure in intensive care units: Results of a
399 multicenter, prospective study. Critical Care Medicine. 1998 Nov;26(11):1793.
- 400 15. Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. Stat Methods
401 Med Res. 2006 Dec;15(6):547–69.
- 402 16. Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the
403 gamma distribution. Stat Med. 1997 Apr 15;16(7):791–801.
- 404 17. Cho H, Klabunde CN, Yabroff KR, Wang Z, Meekins A, Lansdorp-Vogelaar I, et al. Comorbidity-
405 Adjusted Life Expectancy: A New Tool to Inform Recommendations for Optimal Screening Strategies.
406 Ann Intern Med. 2013 Nov 19;159(10):667–76.
- 407 18. National Center for Statistics and Analysis. Lives saved in 2017 by restraint use and minimum-
408 drinking-age laws. [Internet]. Washington, DC: National Highway Traffic Safety Administration; 2019
409 Mar [cited 2024 Jan 18]. (Traffic Safety Facts Crash•Stats). Report No.: DOT HS 812 683. Available
410 from: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812683>
- 411 19. CDC. Centers for Disease Control and Prevention. 2023 [cited 2024 Jan 18]. Benefits of Flu
412 Vaccination During 2019-2020 Flu Season. Available from: [https://www.cdc.gov/flu/about/burden-](https://www.cdc.gov/flu/about/burden-prevented/2019-2020.htm)
413 [prevented/2019-2020.htm](https://www.cdc.gov/flu/about/burden-prevented/2019-2020.htm)
- 414 20. Schmidt H, Roberts DE, Eneanya ND. Sequential organ failure assessment, ventilator rationing and
415 evolving triage guidance: new evidence underlines the need to recognise and revise, unjust allocation
416 frameworks. J Med Ethics. 2022 Feb;48(2):136–8.
- 417 21. Temkin L. Equality, Priority, and the Levelling-Down Objection. In: Clayton M, Williams A, editors.
418 The Ideal of Equality. Macmillan; 2000. p. 126–61.
- 419 22. Shepardson LB, Gordon HS, Ibrahim SA, Harper DL, Rosenthal GE. Racial Variation in the Use of
420 Do-Not-Resuscitate Orders. J Gen Intern Med. 1999 Jan;14(1):15–20.

23. Truog RD. Ventilator Allocation Protocols: Sophisticated Bioethics for an Unworkable Strategy. Hastings Cent Rep. 2021 Sep;51(5):56–7.
24. Dees RH, Herington J, Chiafery M, Shand JC, D'Angio CT, Ching CL, et al. The Ethics of Implementing Emergency Resource Allocation Protocols. The Journal of Clinical Ethics. 2023 Mar;34(1):58–68.

Supporting Information

Table S5: Descriptive Statistics for INCLUDED and EXCLUDED Encounters, April 2020 to May 2021

Fig S5: Violin plot of age distribution by race and ethnicity and COVID-positivity status (AAPI, AIAN, Multi-Racial and Unknown sub-populations are not shown).

Table S6: Distribution of SOFA Scores

Fig S6: Histogram of SOFA score distribution. Coloring indicates NY SOFA scoring bad, Red = first priority, Yellow = second priority, Blue = third priority.

Fig S7: Survival rates by initial SOFA score and age. Single explanatory variable logistic regressions and 95% CI plotted. **A.** In these simple univariable logit models (Survival ~ SOFA" and "Survival ~Age"), each SOFA score point at intubation is associated with a 9.25% (7.38 – 11.09%) decrease in the probability of survival to discharge (Logit coeff. = -0.0971, std err = 0.010). **B.** In our population, each year of subject age is associated with a 3.65% (3.11 – 4.18%) decrease in the probability of survival to discharge (Logit coeff. = -0.0372, std err = 0.003).

Fig S8: Initial SOFA Score and survival by race/ethnicity. A. Distribution of SOFA score by race/ethnicity **B.** In a single explanatory variable logistic regression analyzing the relationship between SOFA score and survival for each racial/ethnic sub-population, each SOFA score point at intubation is associated with decreases in the probability of survival to discharge of 13.15% for non-Hispanic, Black patients, 3.29% (-13.2% to +8.55%) for Hispanic patients of any race, and 8.71% for non-Hispanic, White patients. Logistic regressions and 95% CI plotted.

Table S7: Odds ratio of survival for whole population and racial sub-populations. A logistic regression model was used to calculate the relationship between survival and six explanatory variables:

race (ref: White, non-Hispanic), sex-assigned-at-birth (ref: female), SOFA score, Age, COVID positivity (ref:neg) and Elixhauser comorbidity index. The model was also run for each racial/ethnic sub-population. Odds-ratio are reported alongside 95% confidence intervals, significant effects ($p < 0.01$) in bold.

Fig S9: Racial disparities in allocation. For each protocol, the allocation rate for each race/ethnicity are shown. All rates reported are means of the underlying rates in 1000 simulations of 3700 encounters at 50% scarcity.

Fig S10: Survival rate by Age and Protocol.

Table S8: Survival by Age Group at 50% capacity. Mean survival and 95% CI is reported for 1000 Monte Carlo simulations at the indicated level of scarcity.

Table S9: Survival by Protocol at different Capacities. Mean survival and 95% CI is reported for 250 Monte Carlo simulations at the indicated level of scarcity. All findings within and across capacity-levels are statistically significant to $P < 0.05$.

Table S10: Lives Saved by Protocol at different Capacities. Mean survival and 95% CI is reported for 250 Monte Carlo simulations at the indicated level of scarcity. All findings within and across capacity-levels are statistically significant to $P < 0.05$.

Table S11: Life-years Saved by Protocol at different Capacities. Mean survival and 95% CI is reported for 250 Monte Carlo simulations at the indicated level of scarcity. All findings within and across capacity-levels are statistically significant to $P < 0.05$.

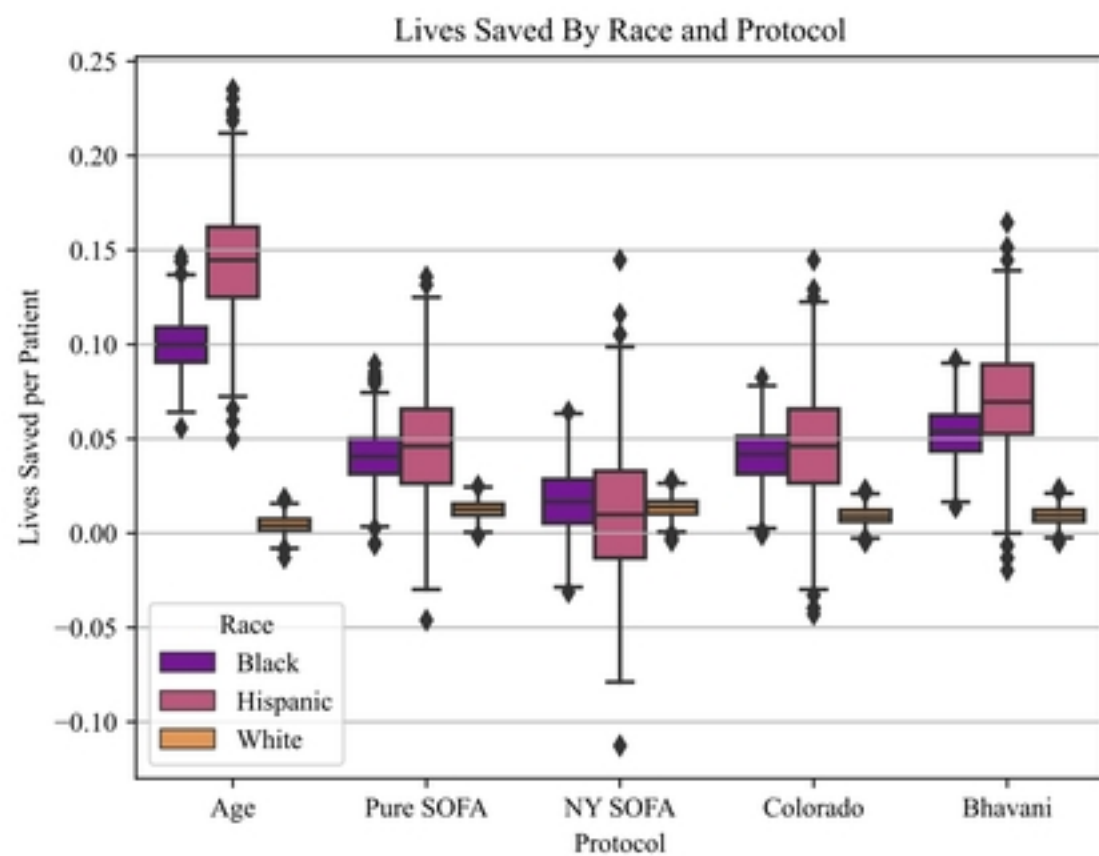
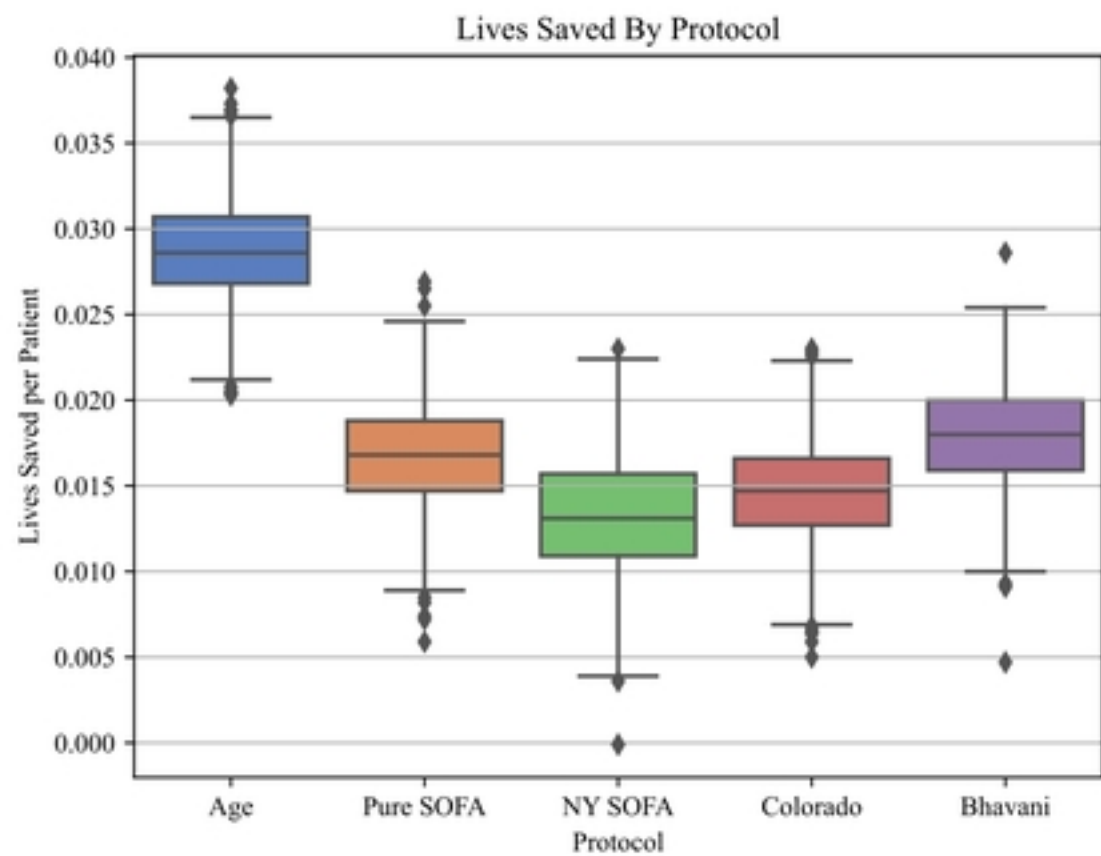


Fig 1

Lives Saved at Different Levels of Scarcity

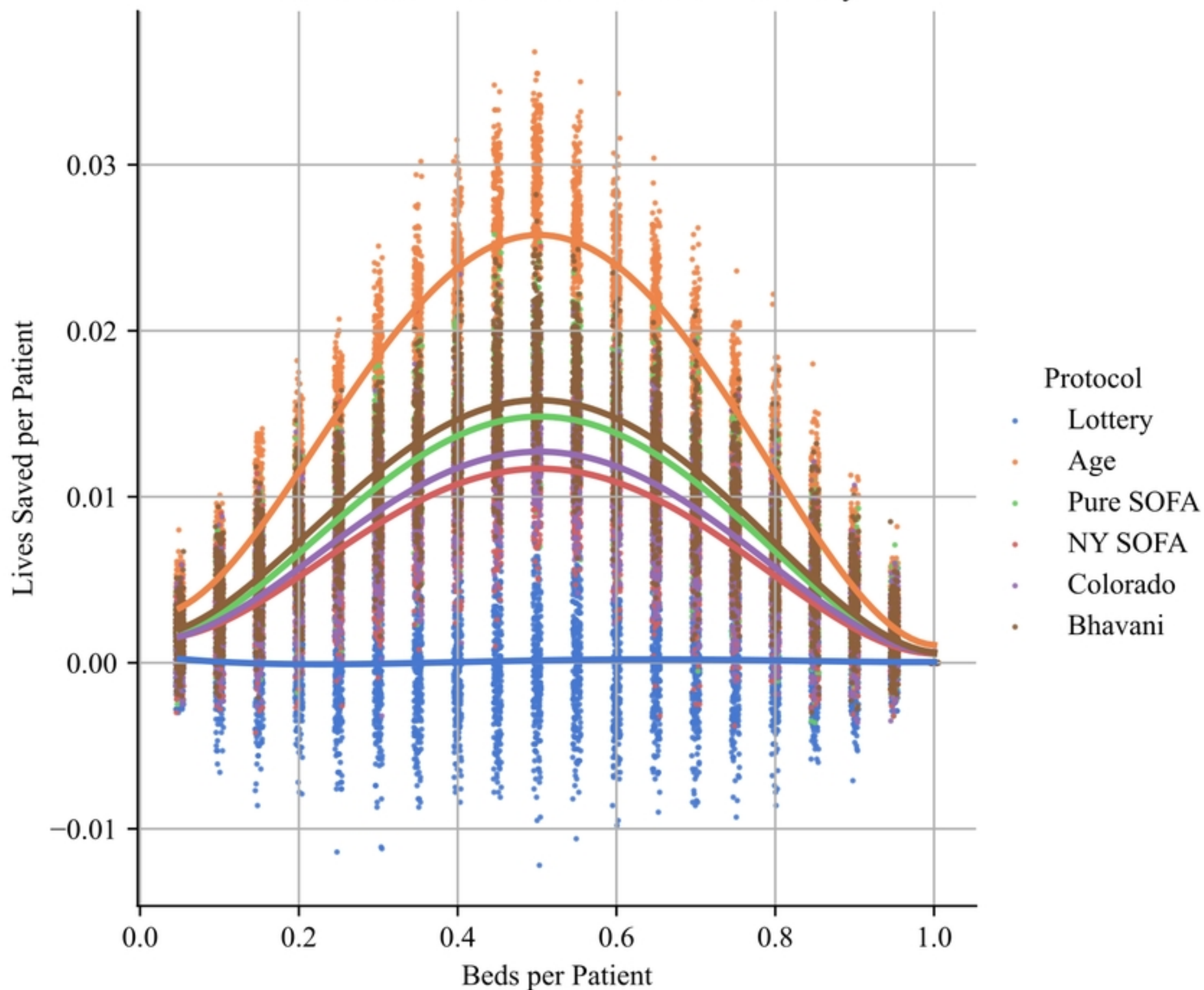
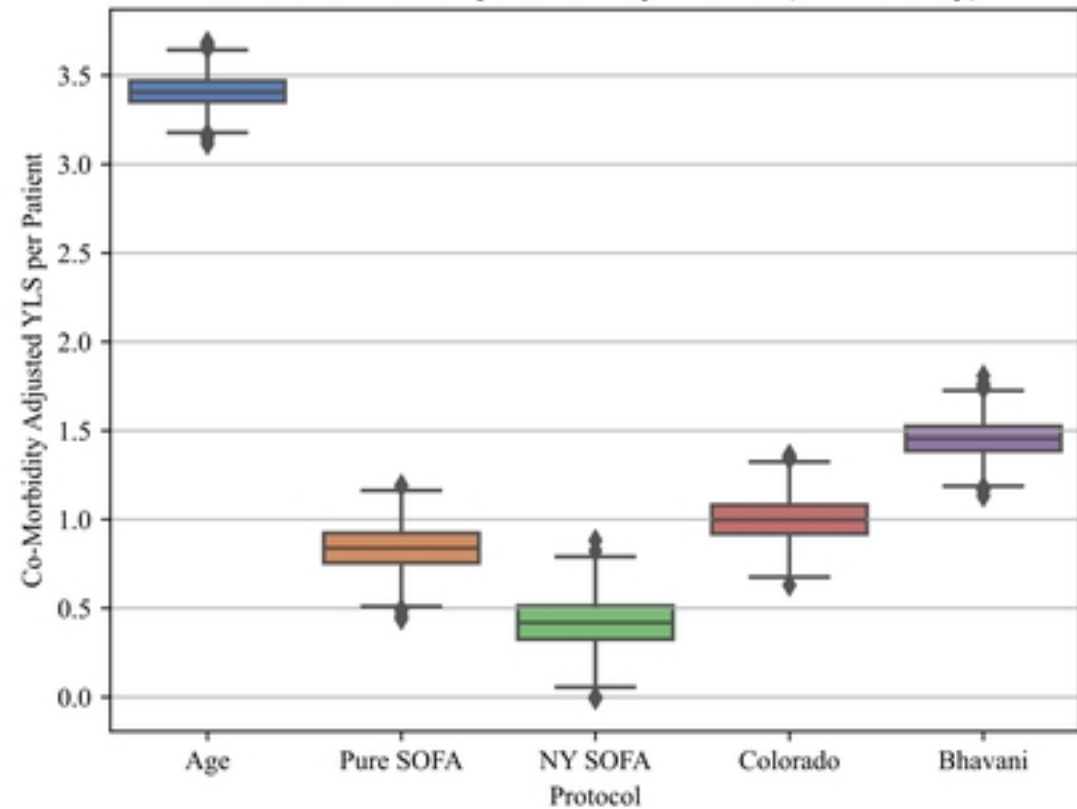


Fig 2

Years of Life Saved per Patient by Protocol (50% Scarcity)



Years of Life Saved per Patient by Race and Protocol (50% Scarcity)

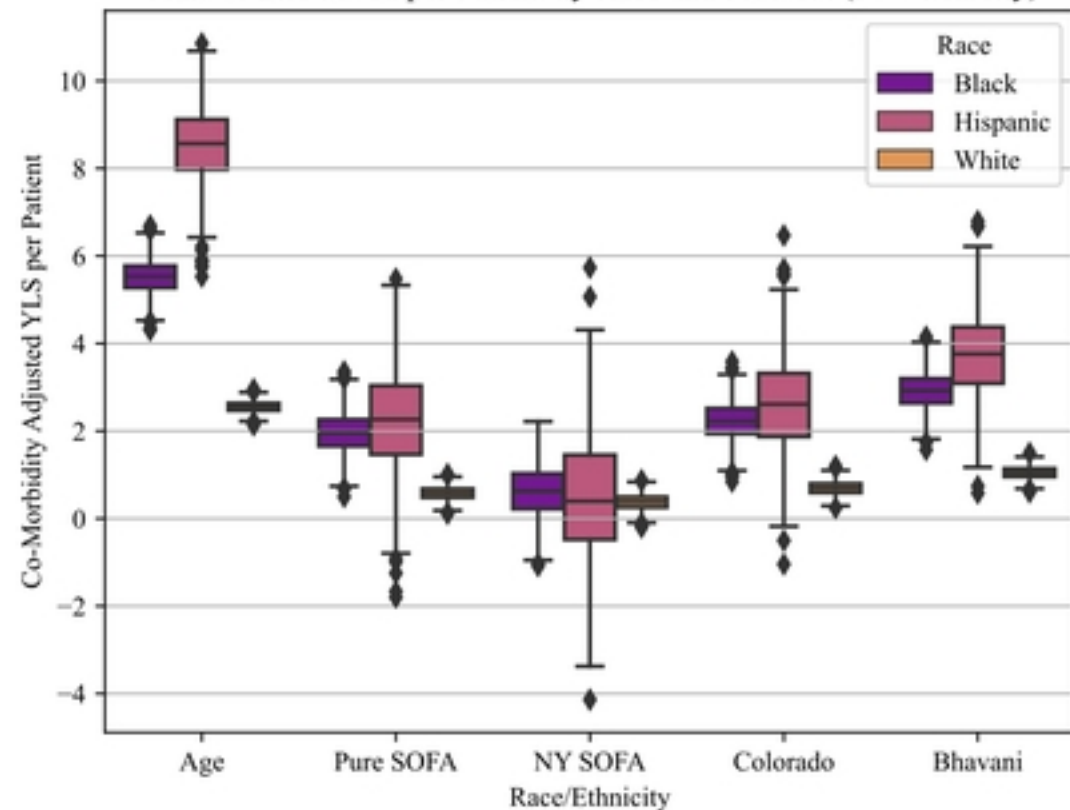


Fig 3

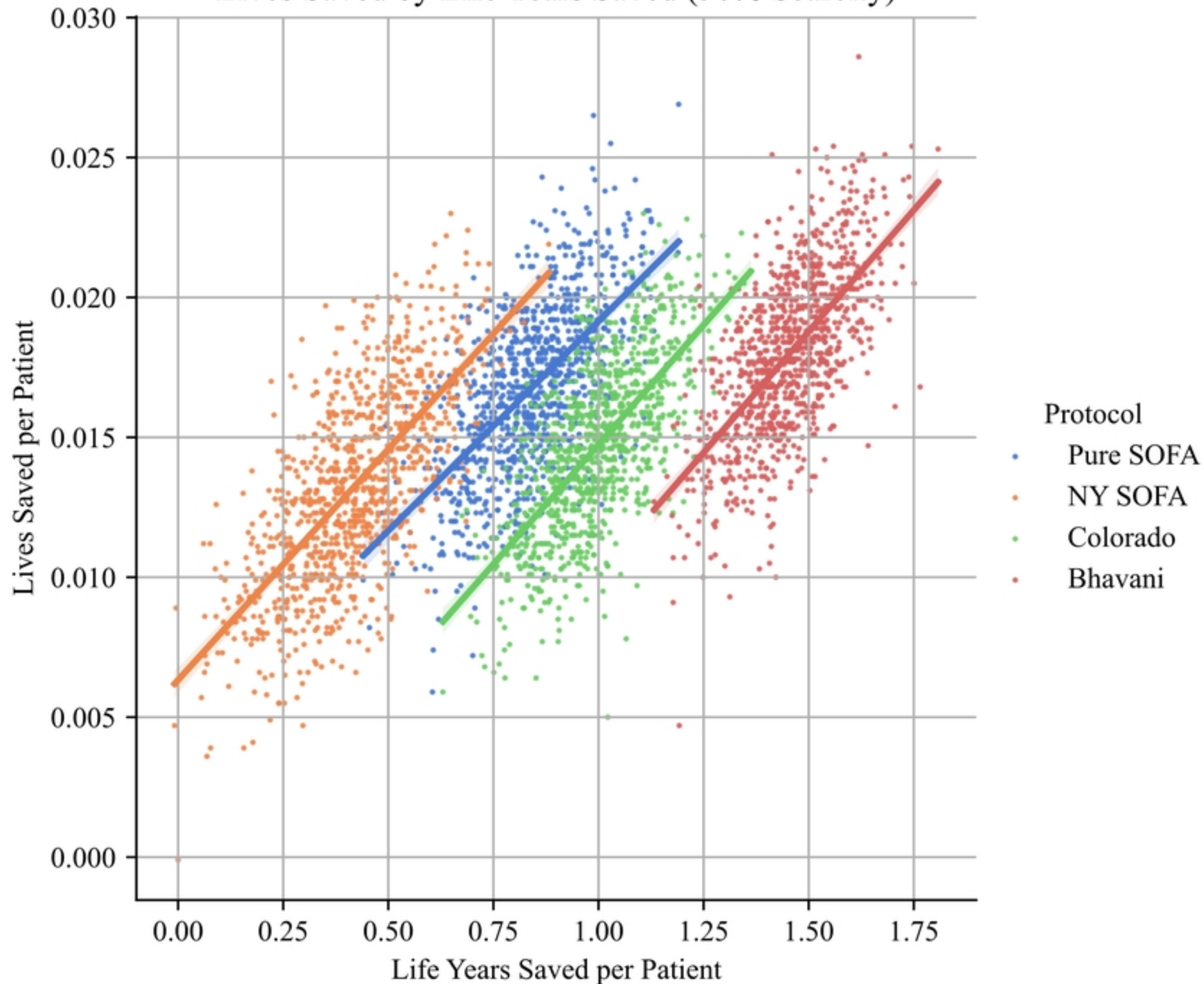


Fig 4