

Care Phenotypes In Critical Care

LL. Weishaupt¹ T. Wang², J. Schamroth³, P. Morandini⁴, J. Matos⁵, LM Hampton⁶, J. Gallifant⁷, A. Fiske⁸, N. Dundas⁹, K. David¹⁰, LA. Celi,¹¹ A.Carrel¹², J. Byers¹⁰, G. Angelotti¹³

1. Harvard & MIT, Health Sciences and Technology, Cambridge, MA, United States
2. Ronald Reagan UCLA Medical Center, Los Angeles, United States
3. Faculty of Population Health Sciences, UCL, London, UK
4. IRCCS Humanitas Research Hospital, Artificial Intelligence Center, Milan, Italy
5. Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, UK
6. Massachusetts Institute of Technology, Electrical Engineering and Computer Science, Cambridge, MA, United States,
7. Massachusetts Institute of Technology, Laboratory for Computational Physiology, Cambridge, MA, United States
8. Institute of History and Ethics in Medicine, Department of Preclinical Medicine, TUM School of Medicine and Health, Technical University of Munich
9. UC Berkeley Department of Bioengineering, CA, United States
10. Beth Israel Deaconess Medical Center, Respiratory Therapy, Boston, MA, United States
11. Imperial College London, London, UK
12. Beth Israel Deaconess Medical Center, Respiratory Therapy, Boston, MA, United States,
13. Dalle Molle Institute for Artificial Intelligence, USI-SUPSI, Lugano, Switzerland

ABSTRACT

The Social Determinants of Health (SDoH) have long been recognised as significant drivers of health inequalities. Within healthcare settings, large EHR datasets have increasingly enabled the use of machine learning (ML) to explore how patient background and demographic factors mediate and predict clinical outcomes. The intensive care unit (ICU) in particular provides a rich source of data for such research. However, major limitations with current approaches persist, including (i) overreliance on individual demographic labels or measures of difference, (ii) the impracticality of highly intersectional patient groups and (iii) that the underlying accuracy and validity of these demographic constructs is low.

The main objective of this study was to take a novel approach, to first understand who within the ICU setting receives sub-standard care and use this to create new, objective labels based on quality of care and outcomes ('Care Phenotypes') when different patients interface with the health system. Using the MIMIC-IV database, we focused on highly protocolised, essential care procedures (turning, mouth care) in mechanically ventilated ICU patients. We performed a series of regression analyses to understand in which patients treatment deviated from these protocols.

In a cohort of 8,919 ICU patients undergoing IMV, consistent patterns in sup-optimal protocol adherence for certain groups, notably heavier patients. Compared to equivalently

sick peers, for every extra weight decile, a patient can expect a reduction of one percentile in frequency of turning care (0.0760 turning interval percentile per weight percentile, $p < 0.05$). Furthermore, patients who receive fewer turnings should also expect to receive fewer mouth care procedures, in a quantile ratio of 1 to 5 (0.2055 mouth care interval quantile per turning interval quantile, $p < 0.05$).

The findings in this initial analysis provide support to the concept of first looking at the actual care delivered to patients to illuminate the relationship between patient demographics and outcomes of interest. The 'Care Phenotypes' approach has the potential to improve fairness evaluations for machine learning in healthcare, support causal inference research and play a larger role in research into healthcare disparities.

INTRODUCTION

Social determinants of health in the era of machine learning

The Social Determinants of Health (SDoH) have long been recognised as significant drivers of health inequalities. The SDoH include socioeconomic status, sex and gender, ethnic background and education levels, all of which can have significant impacts on health outcomes.[1] Adverse structural and social factors including poverty, racism and discrimination by gender, ethnicity or class can also be understood as adverse social determinants of health that are important drivers of health outcomes.[2]

One lens for understanding how these demographic factors lead to inequities in health outcomes is to explore how they manifest within a healthcare setting. An advantage of many healthcare settings (albeit largely those in rich, Western nations) is that electronic health records (EHRs) and associated data collection provides a rich source of information for exploring the associations between a patients' background, environment, disease and health outcomes. The intensive care unit (ICU) in particular is, at least in theory, an optimal setting for research exploring disparities in care delivery and outcomes across different patient demographics.[3] ICU care typically involves invasive treatment with continuous monitoring of physiological status through medical equipment, generating vast amounts of granular clinical data relating to all aspects of care, much of it captured within EHR.

As the use of machine learning (ML) applied to biomedical and health systems research has expanded, researchers have adopted these tools to help probe the relationship between background SDoH factors and care outcomes within the ICU [24]. ML offers potentially powerful tools for eliciting relationships and predicting outcomes by iteratively training models on ever greater volumes of clinical data.

The limitations of demographic labels

To date, there has been no shortage of research into how and why disparities manifest within the ICU. Typically, research has explored differences in key health outcomes as well as differences in the type, or amount, of care ICU patients receive, according to different demographic factors. A wide range of prospective and retrospective studies have explored the relationship between ICU care, health outcomes and i. patient race or ethnicity, [4][5][6][7][8][9] ii. patient sex or gender, [10][11][12][13][14] and iii. socioeconomic status [15] [16][17]. Furthermore, there has increasingly been recognition that the way data is collected within the ICU clinical setting is itself inherently prone to bias along demographic lines which can impact outcomes. One notable example is that patients with darker skin pigmentation are at risk of real-world harm through lower oxygen levels that go undetected due to poorly calibrated pulse-oximeter devices producing inaccurate SpO2 readings. [18]

Similar calibration issues have been identified in other ICU medical devices, which risks encoding bias into clinical data. This in turn becomes more problematic when the large EHR datasets generated from routine ICU data are used to train machine learning models.

Despite progress in this area in recent years, it is apparent from the literature that the use of existing demographic categories for decades has not meaningfully progressed our understanding of health disparities nor facilitated the development of effective tools for mitigating them. The current use of demographic labels for SDoH research within the ICU has major limitations, which are brought into focus when we consider the data required for training machine learning algorithms:

- I. Currently, we rely too heavily on existing metadata categories. Investigating discrepancies in care or outcomes only by individual measures of difference (e.g. race) are too blunt, arbitrary and insufficient for exploring complex, subtle and intersectional questions of disparities that can occur across numerous axes (race, sex, gender, weight, income level, geography). [25,26,27,31] There is mounting evidence that it is the intersections of these factors that will provide the real insights.
- II. Furthermore, attempts at intersectional analyses considering numerous demographic factors also pose major challenges, slicing this pie more finely results in smaller sample sizes which are cumbersome, impractical and often inadequate for research purposes.
- III. Finally, a more fundamental problem exists at the heart of such research efforts, which is that the underlying accuracy and validity of these demographic constructs captured within EHR data may be low. Inaccurate or incomplete data capture about patient background is rife throughout EHR data, posing major issues in data quality. [26, 27,31] The resulting labels relating to patient ethnicity, gender and class may lack fidelity, which can seriously undermine utility and application of this data in health systems research, including for machine learning.

Care Phenotypes: A new approach

Here, we take a novel approach for exploring how differences in ICU care link to individual patient backgrounds. Our aim was to start by understanding who within the ICU setting receives sub-standard care, and then use ML techniques to provide new, objective labels based on quality of care and outcomes ('Care Phenotypes') when different patients interface with the health system. Care phenotypes are based on the sub-type of care patients actually receive. Recognising that we cannot entirely ignore the use of demographic labels in health services research, in this paper we outline how research can move from a demographic-first approach to an outcomes and care quality-first approach.

Within the ICU, strict guidelines are designed to standardize and optimize patient care, ensuring that every patient receives high quality, evidence-based care. The effectiveness of these guidelines hinges not just on their scientific validity but also on their uniform

application across diverse patient groups. Adherence to guidelines should, in principle, be equitable across all patients. In practice, clinical factors, human systems factors as well socioeconomic factors, ethnic background or the healthcare provider's implicit biases can lead to variations in adherence to these guidelines, with some receiving less or lower quality care than others. [4,28,29,30] Focusing on adherence to specific ICU protocols therefore provides a valuable yardstick to scrutinise care quality and how an individual patient's care was actually delivered.

Invasive Mechanical Ventilation (IMV) is a critical intervention in ICU for patients experiencing severe respiratory failure. While life-saving, IMV necessitates rigorous care protocols, including regular patient repositioning (turning frequency) and comprehensive oral care (mouth care) to mitigate complications, such as pressure ulcers and ventilator associated pneumonia respectively. Inconsistent adherence to these protocols can result in adverse outcomes, underscoring the importance of examining care consistency across different patient groups.[19] From a data analysis perspective, protocol adherence vis-a-vis common and repetitive tasks such as turning provide valuable, granular data for analysis, yet remain highly clinically salient, because if overlooked, low adherence to the turning protocol can have serious effects.

These routine procedures (turning, mouth care) were also chosen because they are clear and unambiguous to derive from EHRs. High degrees of certainty regarding treatment timings and magnitude in the dataset allows for minimization of the effect of unobserved confounders, providing some reassurance that results obtained were due to the variable under investigation rather than another unadjusted behavior in how the treatment was recorded.

Through discussion amongst an interdisciplinary team of ICU clinicians, machine learning experts and health service researchers, hypotheses were established regarding which demographic factors may be of greatest relevance given our chosen measure of turning. The team chose to focus on patient weight (kg) for two main reasons. Firstly, the logic of a possible causal link between weight and the policy under investigation (turnings) is well defined from an operational standpoint and supported by clinical experience in discussions with nursing and other clinical staff. Secondly, patient weight is a common predictor used in ML model development to characterize patients and has advantages as a continuous variable that allows for stratification and grouping.

METHODS

Data Sources

The MIMIC-IV database has been used as the primary data source for this study. The MIMIC-IV is a large database containing de-identified ICU electronic health record (EHR) data collected between 2008 and 2019 from the Beth Israel Deaconess Medical Center in

Boston, MA. MIMIC contains demographic profiles of admitted patients, such as self-identified race and ethnicity, English fluency, and insurance. These features allow us to investigate whether disparities linked to weight are stable across diverse sub-populations. Similarly, although de-identified and aggregated, MIMIC retains partial temporal traceability, allowing us to examine whether weight disparities evolve as clinical protocols change. MIMIC-IV contains data for over 65,000 patients admitted to an ICU and over 200,000 patients admitted to the emergency department.[20]

Inclusion and Exclusion Criteria

We included all patients who underwent Invasive Mechanical Ventilation. We excluded all patients under IMV for less than 24 hours, with missing weights and with weights falling outside the 10-250Kg range, see Figure 1.

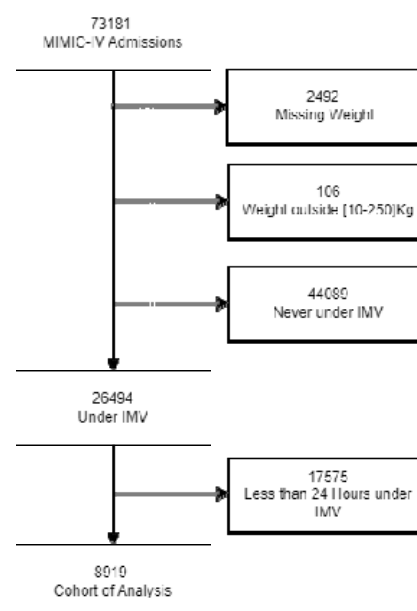
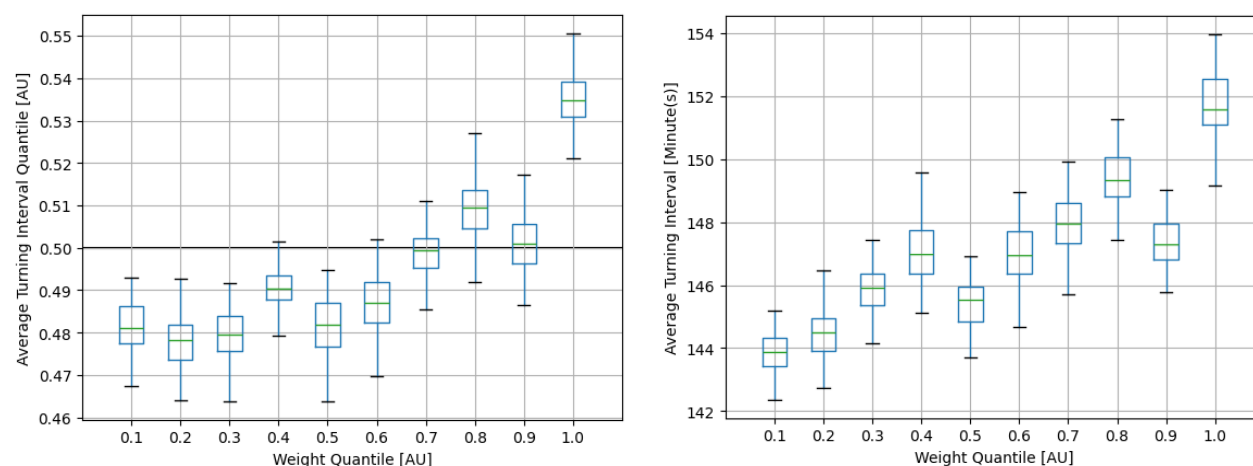


Figure 1. Flow Diagram

RESULTS



Main Results

(a)

(b)

Figure 2: Bootstrap estimates of weight quantile against frequency of turnings (b) and frequency of turnings quantiles (a). (a) In a perfect scenario we would expect the behavior represented with the black line: all patients receive the same amount of care. In reality, we observe that patients on the right end of the spectrum in terms of weight quantiles can expect to receive turning care less frequently than those in weight quantiles towards the left. (b) Absolute average turning intervals. Patients are turned significantly less often than once every 120 minutes, with a marked skew towards heavier quantiles.

We investigated volumes of turnings and mouth care in a cohort of 8,919 Intensive Care Unit (ICU) patients undergoing Invasive Mechanical Ventilation (IMV) into the MIMIC-IV database. Each patient was characterized by disease severity [14], comorbidities [15] as well as various non-clinical demographic factors, such as ethnicity, type of health insurance and first spoken language. We further characterized patients by their period of hospital admission (from 2008-2019) to track possible clinical policy changes across the years, hospital admission location and type of intensive care unit [4]. Volumes of turnings and mouth care were characterized via average intervals between consecutive administrations. Patients undergoing IMV should be turned regularly, regardless of circumstances, in order to prevent serious consequences [23]. For IMV patients in this MIMIC-IV cohort, ICU protocols for turning and mouth care required/recommended turning every 2 hours and mouth care every 2 to 3 hours. These policies remained in place throughout the 2008-2019 period.

		Overall	Frequency of Turnings				p
			Q1: (0.83, 2.1] Hour(s)	Q2: (2.1, 2.3] Hour(s)	Q3: (2.3, 2.6] Hour(s)	Q4: (2.6, 24.0] Hour(s)	
Admissions (n)		8919	2234	2226	2230	2229	
Race or ethnicity, n (%)	Asian	238 (2.7)	52 (2.3)	66 (3.0)	56 (2.5)	64 (2.9)	0.01
	Black	906 (10.2)	210 (9.4)	249 (11.2)	225 (10.1)	222 (10.0)	
	Hispanic	303 (3.4)	86 (3.8)	69 (3.1)	70 (3.1)	78 (3.5)	
	Other	493 (5.5)	155 (6.9)	131 (5.9)	104 (4.7)	103 (4.6)	
	Unknown	1518 (17.0)	388 (17.4)	400 (18.0)	361 (16.2)	369 (16.6)	
Sex, n (%)	White	5461 (61.2)	1343 (60.1)	1311 (58.9)	1414 (63.4)	1393 (62.5)	
	Female	3746 (42.0)	937 (41.9)	920 (41.3)	977 (43.8)	912 (40.9)	0.213
	Male	5173 (58.0)	1297 (58.1)	1306 (58.7)	1253 (56.2)	1317 (59.1)	
BMI, n (%)	Normal	1682 (18.9)	483 (21.6)	395 (17.7)	418 (18.7)	386 (17.3)	0.006
	Obesity	2666 (29.9)	615 (27.5)	686 (30.8)	664 (29.8)	701 (31.4)	
	Overweight	1940 (21.8)	488 (21.8)	500 (22.5)	489 (21.9)	463 (20.8)	
	Severely Underweight	54 (0.6)	14 (0.6)	9 (0.4)	18 (0.8)	13 (0.6)	
	Underweight	153 (1.7)	38 (1.7)	49 (2.2)	37 (1.7)	29 (1.3)	
	Unknown or Unavailable	2424 (27.2)	596 (26.7)	587 (26.4)	604 (27.1)	637 (28.6)	
Weight, mean (SD)		84.8 (25.3)	82.6 (23.2)	84.5 (24.1)	84.5 (25.8)	87.5 (27.7)	<0.001
Language, n (%)	English	7921 (88.8)	1971 (88.2)	1982 (89.0)	1986 (89.1)	1982 (88.9)	0.79
	Other	998 (11.2)	263 (11.8)	244 (11.0)	244 (10.9)	247 (11.1)	
Insurance, n (%)	Medicaid	820 (9.2)	191 (8.5)	204 (9.2)	217 (9.7)	208 (9.3)	0.408
	Medicare	4119 (46.2)	1021 (45.7)	1032 (46.4)	1060 (47.5)	1006 (45.1)	
	Other	3980 (44.6)	1022 (45.7)	990 (44.5)	953 (42.7)	1015 (45.5)	
Care Unit, n (%)	Cardiac	1842 (20.7)	324 (14.5)	325 (14.6)	479 (21.5)	714 (32.0)	<0.001
	Medical-Surgical	5371 (60.2)	1208 (54.1)	1401 (62.9)	1429 (64.1)	1333 (59.8)	
	Neuro-Trauma	1706 (19.1)	702 (31.4)	500 (22.5)	322 (14.4)	182 (8.2)	
Age (admission), mean (SD)		63.5 (16.4)	63.5 (16.9)	64.0 (16.3)	63.8 (16.3)	62.9 (16.1)	0.099
Admission Location, n (%)	Emergency Room	4898 (54.9)	1180 (52.8)	1237 (55.6)	1284 (57.6)	1197 (53.7)	0.004
	Other	495 (5.5)	149 (6.7)	131 (5.9)	103 (4.6)	112 (5.0)	
	Physician Referral	1173 (13.2)	324 (14.5)	287 (12.9)	272 (12.2)	290 (13.0)	
	Transfer From Hospital	2353 (26.4)	581 (26.0)	571 (25.7)	571 (25.6)	630 (28.3)	
Admission Year, n (%)	2008 - 2010	3547 (39.8)	739 (33.1)	804 (36.1)	943 (42.3)	1061 (47.6)	<0.001
	2011 - 2013	2215 (24.8)	518 (23.2)	544 (24.4)	622 (27.9)	531 (23.8)	
	2014 - 2016	1975 (22.1)	543 (24.3)	557 (25.0)	457 (20.5)	418 (18.8)	
	2017 - 2019	1182 (13.3)	434 (19.4)	321 (14.4)	208 (9.3)	219 (9.8)	
SOFA, mean (SD)		7.3 (4.2)	6.9 (4.0)	7.2 (4.1)	7.4 (4.2)	7.8 (4.3)	<0.001
CCI, mean (SD)		5.1 (3.0)	4.9 (3.1)	5.0 (3.0)	5.2 (2.9)	5.1 (2.8)	0.079
IMV Days, mean (SD)		4.4 (5.0)	3.2 (3.2)	4.7 (5.1)	5.5 (5.7)	4.4 (5.6)	<0.001
Caregivers, mean (SD)		2.8 (0.5)	2.8 (0.5)	2.8 (0.5)	2.8 (0.4)	2.6 (0.5)	<0.001
In-Hospital Mortality, n (%)	Deceased	2537 (28.4)	596 (26.7)	614 (27.6)	632 (28.3)	695 (31.2)	0.006
	Survivor	6382 (71.6)	1638 (73.3)	1612 (72.4)	1598 (71.7)	1534 (68.8)	

Table 1A. Cohort Description, stratified by frequency of turnings.

Caption: BMI, body mass index; SOFA, sequential organ failure assessment score; CCI, Charlson comorbidity score; GCS, Glasgow coma score

		Overall	Frequency of Mouth Care				p
			Q1: (1.411, 3.429] Hour(s)	Q2: (3.429, 3.832] Hour(s)	Q3: (3.832, 4.238] Hour(s)	Q4: (4.238, 24.0] Hour(s)	
Admissions (n)		8892	2735	1711	2223	2223	
Race or ethnicity, n (%)	Asian	238 (2.7)	61 (2.2)	48 (2.8)	60 (2.7)	69 (3.1)	0.014
	Black	901 (10.1)	303 (11.1)	162 (9.5)	208 (9.4)	228 (10.3)	
	Hispanic	303 (3.4)	89 (3.3)	72 (4.2)	66 (3.0)	76 (3.4)	
	Other	493 (5.5)	138 (5.0)	101 (5.9)	142 (6.4)	112 (5.0)	
	Unknown	1514 (17.0)	458 (16.7)	254 (14.8)	384 (17.3)	418 (18.8)	
	White	5443 (61.2)	1686 (61.6)	1074 (62.8)	1363 (61.3)	1320 (59.4)	
Sex, n (%)	Female	3731 (42.0)	1117 (40.8)	722 (42.2)	956 (43.0)	936 (42.1)	0.482
	Male	5161 (58.0)	1618 (59.2)	989 (57.8)	1267 (57.0)	1287 (57.9)	
BMI, n (%)	Normal	1679 (18.9)	498 (18.2)	331 (19.3)	426 (19.2)	424 (19.1)	0.072
	Obesity	2659 (29.9)	802 (29.3)	537 (31.4)	689 (31.0)	631 (28.4)	
	Overweight	1932 (21.7)	582 (21.3)	373 (21.8)	496 (22.3)	481 (21.6)	
	Severely Underweight	54 (0.6)	14 (0.5)	10 (0.6)	12 (0.5)	18 (0.8)	
	Underweight	153 (1.7)	40 (1.5)	27 (1.6)	34 (1.5)	52 (2.3)	
	Unknown or Unavailable	2415 (27.2)	799 (29.2)	433 (25.3)	566 (25.5)	617 (27.8)	
Weight, mean (SD)		84.8 (25.3)	85.4 (25.5)	85.2 (25.0)	85.3 (26.0)	83.1 (24.5)	0.005
Language, n (%)	English	7896 (88.8)	2437 (89.1)	1497 (87.5)	1971 (88.7)	1991 (89.6)	0.209
	Other	996 (11.2)	298 (10.9)	214 (12.5)	252 (11.3)	232 (10.4)	
Insurance, n (%)	Medicaid	817 (9.2)	245 (9.0)	189 (11.0)	196 (8.8)	187 (8.4)	0.118
	Medicare	4106 (46.2)	1247 (45.6)	778 (45.5)	1034 (46.5)	1047 (47.1)	
	Other	3969 (44.6)	1243 (45.4)	744 (43.5)	993 (44.7)	989 (44.5)	
Care Unit, n (%)	Cardiac	1834 (20.6)	463 (16.9)	267 (15.6)	509 (22.9)	595 (26.8)	<0.001
	Medical-Surgical	5356 (60.2)	1717 (62.8)	1080 (63.1)	1268 (57.0)	1291 (58.1)	
	Neuro-Trauma	1702 (19.1)	555 (20.3)	364 (21.3)	446 (20.1)	337 (15.2)	
Age (admission), mean		63.5	62.8 (16.5)	63.0 (16.4)	64.7 (16.1)	63.7 (16.5)	<0.001

(SD)		(16.4)					1
Admission Location, n (%)	Emergency Room	4884 (54.9)	1700 (62.2)	1042 (60.9)	1109 (49.9)	1033 (46.5)	<0.001
	Other	492 (5.5)	95 (3.5)	70 (4.1)	169 (7.6)	158 (7.1)	
	Physician Referral	1170 (13.2)	286 (10.5)	175 (10.2)	317 (14.3)	392 (17.6)	
	Transfer From Hospital	2346 (26.4)	654 (23.9)	424 (24.8)	628 (28.3)	640 (28.8)	
Admission Year, n (%)	2008 - 2010	3524 (39.6)	1252 (45.8)	774 (45.2)	808 (36.3)	690 (31.0)	<0.001
	2011 - 2013	2215 (24.9)	711 (26.0)	467 (27.3)	520 (23.4)	517 (23.3)	
	2014 - 2016	1972 (22.2)	525 (19.2)	325 (19.0)	537 (24.2)	585 (26.3)	
	2017 - 2019	1181 (13.3)	247 (9.0)	145 (8.5)	358 (16.1)	431 (19.4)	
SOFA, mean (SD)		7.3 (4.2)	7.0 (4.2)	7.5 (4.2)	7.3 (4.0)	7.6 (4.3)	<0.001
CCI, mean (SD)		5.1 (3.0)	4.9 (2.9)	4.9 (3.0)	5.2 (2.9)	5.2 (3.0)	0.002
GCS, mean (SD)		13.1 (3.6)	12.9 (3.7)	12.9 (3.7)	13.2 (3.4)	13.1 (3.6)	0.003
Frequency of Turnings, mean (SD)		2.5 (1.0)	2.4 (0.8)	2.4 (0.5)	2.4 (0.8)	2.8 (1.6)	<0.001
IMV Days, mean (SD)		4.5 (5.0)	3.6 (4.2)	6.7 (6.2)	4.4 (5.2)	3.8 (4.3)	<0.001
Caregivers, mean (SD)		2.8 (0.5)	2.8 (0.5)	2.8 (0.3)	2.7 (0.5)	2.7 (0.5)	0.001
In-Hospital Mortality, n (%)	Deceased	2534 (28.5)	716 (26.2)	501 (29.3)	643 (28.9)	674 (30.3)	0.009
	Survivor	6358 (71.5)	2019 (73.8)	1210 (70.7)	1580 (71.1)	1549 (69.7)	

Table 1B. Cohort Description, stratified by frequency of mouth care.

Caption: BMI, body mass index; SOFA, sequential organ failure assessment score; CCI, Charlson comorbidity score; GCS, Glasgow coma score

Our findings indicate that 50% of patients were turned significantly less than once every two hours (Median: 2.3 Hours, IQR: 2.1 - 2.6 Hours, $p < 0.05$), and that adherence to the protocol adherence was uneven - see figure 2. Focusing on weight as our variable of interest, we observe that heavier patients can expect to be turned less often, even compared to equivalently sick peers, in a predictable manner: for every extra weight decile, a patient can expect a reduction of one percentile in frequency of turning care (0.0760 turning interval percentile per weight percentile, $p < 0.05$, Figure 3a). More worryingly, patients who receive fewer turnings should also expect to receive fewer mouth care procedures, in a quantile ratio of 1 to 5 (0.2055 mouth care interval quantile per turning interval quantile, $p < 0.05$, Figure 3b).

See Appendix A for an extensive description of the methodology and results.

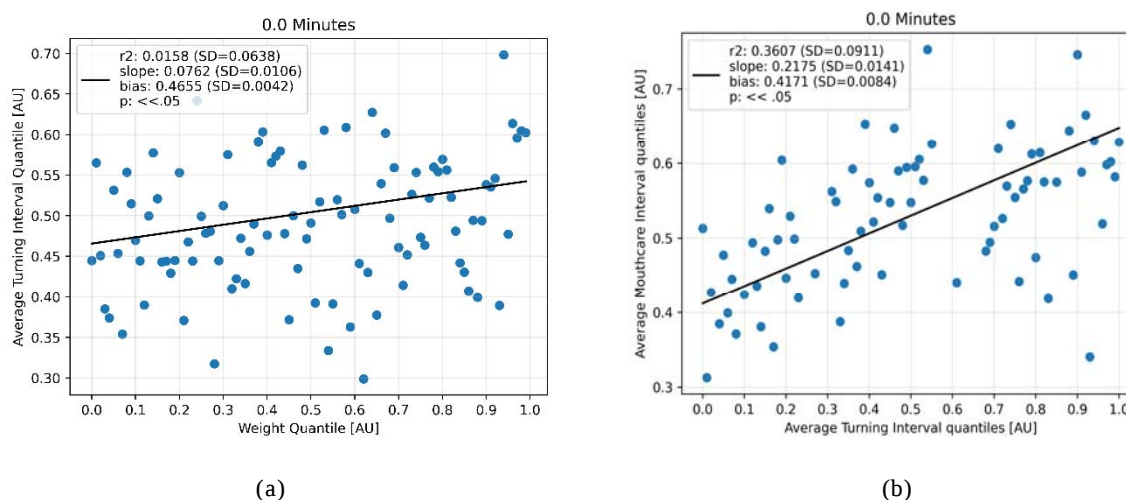


Figure 3: Univariate regression between weight quantiles and frequencies of turnings (a) and quantiles of turning and mouth care frequencies with no additive noise (b). There appears to be a significant positive correlation in both, suggesting that heavier patients who already receive fewer turns also receive less mouth care.

DISCUSSION

Key Results

We investigated volumes of turnings and mouth care in a cohort of 8,919 Intensive Care Unit patients undergoing Invasive Mechanical Ventilation (IMV) identified within the MIMIC-IV database. Each patient was characterized by severity [20], comorbidity [21] and demography plus other non-clinical factors, namely ethnicity, type of health insurance and first spoken language. This resulted in a cohort with a majority of English speaking elderly white males (Mean Age: 63.5 years old at admission, 58% Males, 88.8% reported English as their first language).

In this cohort we observe that many IMV patients are turned significantly less often than required by standard protocols and the volume of care (measured by turning frequency) progressively diminishes as the patient's weight increases. In addition, patients receiving less frequent turning also appeared to receive reduced mouth care. As such, our care phenotype here can be understood as those receiving suboptimal turning and mouthcare during invasive mechanical ventilation, with patient weight emerging as a key mediator of this observation. That worse care and reduced protocol adherence in one aspect (turning) then in turn spills over into another (mouth care) suggests a compounding effect of subtle diminishing care.

Besides the inherent practical importance of the specific finding itself, its implications are far-reaching. Our findings highlight the existence of a plane orthogonal to Social

Determinants of Health (SDoH), intimately linked to the uniqueness of each site of care, operating on different axes: the phenotypes of care. These phenotypes directly profile the subtype of care a patient is actually receiving and complement our understanding of the SDoH, providing a tool for researchers using machine learning to move beyond the limitations of crude demographic labels to elicit new insights into care disparities.

This approach may be especially important for machine learning development. Often, SDoH or their proxies are naively integrated into models under the justification of fairness, but in reality, this may exacerbate biases. For instance, IMV turning prediction models trained on MIMIC might associate weight with ethnicity, systematically advising practitioners to turn heavier patients later, which is the opposite of what is intended, and this bias can also occur through ethnicity alone. The Care Phenotypes approach can thus support fairness evaluations for applied clinical algorithms.

Finally, our results indicate that no matter the amount of data or the sophistication of the machine learning models we employ, we will never resolve this issue unless we embed causal inference frameworks.[22] These frameworks should be designed in close collaboration with experienced clinical experts and should include a precise and attentive mapping of phenotypes of care derived from the unique circumstances in which the data was captured.

REFERENCES

1. WHO Commission on Social Determinants of Health and World Health Organization, 2008. *Closing the gap in a generation: health equity through action on the social determinants of health: Commission on Social Determinants of Health final report*. World Health Organization.
2. Krieger, N., Rowley, D.L., Herman, A.A., Avery, B. and Phillips, M.T., 1993. Racism, sexism, and social class: implications for studies of health, disease, and well-being. *American journal of preventive medicine*, 9(6), pp.82-122.
3. Charpignon, M.L., Byers, J., Cabral, S., Celi, L.A., Fernandes, C., Gallifant, J., Lough, M.E., Mlombwa, D., Moukheiber, L., Ong, B.A. and Panitchote, A., 2023. Critical bias in critical care devices. *Critical Care Clinics*, 39(4), pp.795-813.
4. McGowan, S.K., Sarigiannis, K.A., Fox, S.C., Gottlieb, M.A. and Chen, E., 2022. Racial disparities in ICU outcomes: a systematic review. *Critical care medicine*, 50(1), pp.1-20.
5. Breathett, K., Liu, W.G., Allen, L.A., Daugherty, S.L., Blair, I.V., Jones, J., Grunwald, G.K., Moss, M., Kiser, T.H., Burnham, E. and Vandivier, R.W., 2018. African Americans are less likely to receive care by a cardiologist during an intensive care unit admission for heart failure. *JACC: Heart Failure*, 6(5), pp.413-420.
6. Mayr, F.B., Yende, S., D'Angelo, G., Barnato, A.E., Kellum, J.A., Weissfeld, L., Yealy, D.M., Reade, M.C., Milbrandt, E.B. and Angus, D.C., 2010. Do hospitals provide lower quality of care to black patients for pneumonia?. *Critical care medicine*, 38(3), pp.759-765.
7. Frei, C.R., Mortensen, E.M., Copeland, L.A., Attridge, R.T., Pugh, M.J.V., Restrepo, M.I., Anzueto, A., Nakashima, B. and Fine, M.J., 2010. Disparities of care for African-Americans and Caucasians with community-acquired pneumonia: a retrospective cohort study. *BMC*

- health services research*, 10, pp.1-11.
8. Teotia, K., Jia, Y., Woite, N.L., Celi, L.A., Matos, J. and Struja, T., 2024. Variation in monitoring: Glucose measurement in the ICU as a case study to preempt spurious correlations. *Journal of Biomedical Informatics*, 153, p.104643.
 9. Chang, T., Nuppnau, M., He, Y., Kocher, K.E., Valley, T.S., Sjoding, M.W. and Wiens, J., 2024. Racial differences in laboratory testing as a potential mechanism for bias in AI: A matched cohort analysis in emergency department visits. *PLOS Global Public Health*, 4(10), p.e0003555.
 10. Modra, L., Higgins, A., Vithanage, R., Abeygunawardana, V., Bailey, M. and Bellomo, R., 2021. Sex differences in illness severity and mortality among adult intensive care patients: A systematic review and meta-analysis. *Journal of critical care*, 65, pp.116-123.
 11. Chang, T., Nuppnau, M., He, Y., Kocher, K.E., Valley, T.S., Sjoding, M.W. and Wiens, J., 2024. Racial differences in laboratory testing as a potential mechanism for bias in AI: A matched cohort analysis in emergency department visits. *PLOS Global Public Health*, 4(10), p.e0003555.
 12. Modra, L.J., Higgins, A.M., Pilcher, D.V., Bailey, M. and Bellomo, R., 2024. Sex differences in vital organ support provided to ICU patients. *Critical Care Medicine*, 52(1), pp.1-10.
 13. Zettersten, E., Jäderling, G., Bell, M. and Larsson, E., 2020. Sex and gender aspects on intensive care. A cohort study. *Journal of critical care*, 55, pp.22-27.
 14. Modra, L.J., Higgins, A.M., Pilcher, D.V., Bailey, M.J. and Bellomo, R., 2022. Sex differences in mortality of ICU patients according to diagnosis-related sex balance. *American journal of respiratory and critical care medicine*, 206(11), pp.1353-1360.
 15. McHenry, R.D., Moultrie, C.E., Quasim, T., Mackay, D.F. and Pell, J.P., 2023. Association between socioeconomic status and outcomes in critical care: a systematic review and meta-analysis. *Critical Care Medicine*, 51(3), pp.347-356.
 16. Khanijahani, A., Iezadi, S., Gholipour, K., Azami-Aghdash, S. and Naghibi, D., 2021. A systematic review of racial/ethnic and socioeconomic disparities in COVID-19. *International journal for equity in health*, 20, pp.1-30.
 17. Mullany, D.V., Pilcher, D.V. and Dobson, A.J., 2021. Associations between socioeconomic status, patient risk, and short-term intensive care outcomes. *Critical Care Medicine*, 49(9), pp.e849-e859.
 18. Gershonovitch, R., Yarom, N. and Findler, M., 2020. Preventing ventilator-associated pneumonia in intensive care unit by improved oral care: a review of randomized control trials. *SN comprehensive clinical medicine*, 2(6), pp.727-733.
 19. Charpignon, M.L., Byers, J., Cabral, S., Celi, L.A., Fernandes, C., Gallifant, J., Lough, M.E., Mlombwa, D., Moukheiber, L., Ong, B.A. and Panitchote, A., 2023. Critical bias in critical care devices. *Critical Care Clinics*, 39(4), pp.795-813.
 20. Johnson, A., Bulgarelli, L., Pollard, T., Gow, B., Moody, B., Horng, S., Celi, L. A., and Mark, R. (2024) 'MIMIC-IV' (version 3.1), PhysioNet. Available at: <https://doi.org/10.13026/kpb9-mt58>.
 21. Yarnell CJ, Johnson A, Dam T, Jonkman A, Liu K, Wunsch H, et al. Do thresholds for invasive ventilation in hypoxemic respiratory failure exist? A cohort study. *Am J Respir Crit Care Med*. 2023;207: 271–282.
 22. Bareinboim, E., Correa, J.D., Ibeling, D. and Icard, T., 2022. On Pearl's hierarchy and the foundations of causal inference. In *Probabilistic and causal inference: the works of judea pearl* (pp. 507-556).
 23. Winkelman, C. and Chiang, L.C., 2010. Manual turns in patients receiving mechanical ventilation. *Critical Care Nurse*, 30(4).
 24. Patra BG, Sharma MM, Vekaria V, Adekkanattu P, Patterson OV, Glicksberg B, Lepow LA, Ryu E, Biernacka JM, Furmanchuk A, George TJ, Hogan W, Wu Y, Yang X, Bian J, Weissman M, Wickramaratne P, Mann JJ, Olfson M, Champion TR, Weiner M, Pathak J. Extracting social determinants of health from electronic health records using natural language processing: a systematic review. *J Am Med Inform Assoc*. 2021 Nov 25;28(12):2716-

2727. doi: 10.1093/jamia/ocab170. PMID: 34613399; PMCID: PMC8633615.
25. Cook LA, Sachs J, Weiskopf NG. The quality of social determinants data in the electronic health record: a systematic review, *Journal of the American Medical Informatics Association*, Volume 29, Issue 1, January 2022, Pages 187–196, <https://doi.org/10.1093/jamia/ocab199>
26. Penev YP, Buchanan TR, Ruppert MM, Liu M, Shekouhi R, Guan Z, Balch J, Ozrazgat-Baslanti T, Shickel B, Loftus TJ, Bihorac A. Electronic Health Record Data Quality and Performance Assessments: Scoping Review. *JMIR Med Inform* 2024;12:e58130. doi: 10.2196/58130
27. Chin MK, •oàn L, Russo RG, Roberts T, Persaud S, Huang E, Fu L, Kui KY, Simona Kwon C, Yi SS, Methods for retrospectively improving race/ethnicity data quality: a scoping review, *Epidemiologic Reviews*, Volume 45, Issue 1, 2023, Pages 127–139, <https://doi.org/10.1093/epirev/mxad002>
28. Drows, F., Wallace, J., Benuzillo, J., Markewitz, B.A., & Samore, M.H. 2012. Protocol adherence in the intensive care unit. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22.(1) doi.org/10.1002/hfm.20280
29. Mohammed S, Matos J, Doutreligne M, Celi LA, Struja T. Racial Disparities in Invasive ICU Treatments Among Septic Patients: High Resolution Electronic Health Records Analysis from MIMIC-IV. *Yale J Biol Med*. 2023 Sep 29;96(3):293-312. doi: 10.59249/WDJI8829. PMID: 37780990; PMCID: PMC10524813.
30. Adherence to Guidelines in Adult Patients with Traumatic Brain Injury: A Living Systematic Review Maryse C. Cnossen, Annemieke C. Scholten, Hester F. Lingsma, Anneliese Synnot, Emma Tavender, Dashiell Gantner, Fiona Lecky, Ewout W. Steyerberg, and Suzanne Polinder *Journal of Neurotrauma* 2021 38:8, 1072-1085
31. Sojka PC, Maron MM, Dunsiger SI, Belgrave C, Hunt JI, Brannan EH, Wolff JC. Evaluation of Reliability Between Race and Ethnicity Data Obtained from Self-report Versus Electronic Health Record. *J Racial Ethn Health Disparities*. 2024 Jun 5;10.1007/s40615-024-02041-w. doi: 10.1007/s40615-024-02041-w. Epub ahead of print. PMID: 38839729; PMCID: PMC11634015.

-
- **looking first at care received to create**
- *. However, as ML approaches in clinical care are increasingly employed, it is clear that SDoH are not helpful for assessing if ML approaches are fair or if they increase inequality.*
- *In this paper, we argue that it is necessary to shift away from traditional demographic labels to more direct measures of quality of care received.*
- *We propose a novel approach, “Care Phenotypes” in order to assess the fairness of ML in relation to how Social Determinants of Health affect clinical care outcomes. The benefit of this method is that it directly measures the care that was received by patients, rather than relying on potentially inaccurate or complex demographic intersections.*
- *Care phenotypes speak to a growing body of concerns around how fairness in ML can be assessed and measured. It is important to clarify that Care phenotypes is not intended to be a direct intervention into clinical practice, rather it is a way to improve fairness evaluation for machine learning in healthcare. SDoC is an intervention that works in tandem with research on SDoH to improve fairness in clinical AI applications.*
- *This can be applied in fairness eval; causal inference work; disparity work.*