JACC: ADVANCES VOL. 4, NO. 5, 2025

ORIGINAL RESEARCH

Cardiac Output Estimation in the Intensive Care Unit

Eric Palanques-Tost, MS,[a,b] Roger Pallarès-López, MS,[a,b] Raimon Padrós-Valls, MS,[a,b] Steven Song, BS,[a,c] Erik Reinertsen, MD, PhD,[a,d] Timothy W. Churchill, MD,[a] Paige Stockwell, BS,[a] Eugene Pomerantsev, MD, PhD,[a] Joseph Garasic, MD,[a] Thoralf M. Sundt, MD,[e] Pinak Shah, MD,[f] Nicholas E. Houstis, MD, PhD,[a,g,*] Aaron D. Aguirre, MD, PhD[a,b,g,h,*]

ABSTRACT

BACKGROUND Cardiac output (CO) is a quintessential property of the cardiovascular system, one whose estimation is vital to patient care in critical illness. The most common techniques for assessing CO, thermodilution (TD) and the estimated Fick (eFick) approximation, force tradeoffs that motivate a need for new methods.

OBJECTIVES The purpose of this study was to novel CO estimators to fill key gaps in critical care medicine.

METHODS Machine learning was used to estimate CO from physiology measurements made during routine clinical care in the intensive care unit (ICU) or cardiac catheterization lab. Models were trained and validated using a curated set of 13,172 ground-truth measurements of TD-CO from 4,825 patients. Model performance was evaluated using regression metrics, trajectory analysis, classification accuracy, and ΔCO tracking.

RESULTS Three established eFick models all performed poorly in the ICU because their static estimates of oxygen consumption could not track the dynamics of critical illness. In the postcardiac surgery intensive care unit, the best eFick model erred in its CO predictions by 30% (mean absolute percentage error [MAPE]) with a coefficient of determination (R^2) of −1.5. The best model derived here, labeled CORE (Catheter Optimized caRdiac output Estimation), predicted CO with an MAPE of 14% ($P < 0.001$ vs eFick) and an R^2 of 0.58. These estimates could be calculated from measurements obtained with either a pulmonary artery catheter or a central venous catheter. The CORE model was also robust to the presence of moderate or severe tricuspid regurgitation, achieving an MAPE of 16% and R^2 of 0.65 relative to a ground-truth determined by the direct Fick technique with measured oxygen consumption.

CONCLUSIONS CO models that account for dynamic physiology in ICU patients were more accurate than widely used eFick models and more versatile than TD. The performance of these models combined with their adaptation to vascular access, broad applicability, ease of use, and ease of deployment should enable them to benefit patients across diverse ICU settings. (JACC Adv. 2025;4:101663) © 2025 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

From the [a]Cardiology Division, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; [b]Center for Systems Biology, Massachusetts General Hospital, Boston, Massachusetts, USA; [c]University of Chicago Pritzker School of Medicine, Chicago, Illinois, USA; [d]Research Laboratory of Electronics, Computer Science & Artificial Intelligence Laboratory, MIT, Cambridge, Massachusetts, USA; [e]Cardiac Surgery Division, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA; [f]Cardiology Division, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; [g]Healthcare Transformation Lab, Massachusetts General Hospital, Boston, Massachusetts, USA; and the [h]Wellman Center for Photomedicine, Massachusetts General Hospital, Boston, Massachusetts, USA. *These authors have contributed equally to this work.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received September 6, 2024; revised manuscript received January 30, 2025, accepted February 14, 2025.

ISSN 2772-963X https://doi.org/10.1016/j.jacadv.2025.101663

JACC: ADVANCES, VOL. 4, NO. 5, 2025
MAY 2025:101663

ABBREVIATIONS AND ACRONYMS

CCL = cardiac catheterization lab

CCL_{ext} = external cardiac catheterization lab

CICU = cardiac intensive care unit

CO = cardiac output

CORE = Catheter Optimized caRdiac output Estimation

CSICU = cardiac surgery intensive care unit

CVC = central venous catheter

eFick = estimated Fick

ICU = intensive care unit

MAPE = mean absolute percentage error

PAC = pulmonary artery catheter

R^2 = coefficient of determination

$ScvO_2$ = central venous oxygen saturation

$SmvO_2$ = mixed venous oxygen saturation

TD = thermodilution

TR = tricuspid regurgitation

$\dot{V}_{O_2}$ = oxygen consumption

Of all cardiovascular properties, perhaps none embodies circulatory performance better than cardiac output (CO).[1] Its measurement is thus fundamental to patient care, but there is no universal tool for the task. Every tool, whether premised on the Fick principle, indicator-dilution measurement, arterial pulse wave analysis, or other principles, must weigh a set of tradeoffs.[2,3] It must balance accuracy and precision against factors such as required vascular access, required expertise, applicability across patient subgroups (eg, shunt physiology or tricuspid regurgitation [TR]), and availability of equipment to name a few. Navigating such tradeoffs is the odyssey of every innovation that aims to assess CO. Furthermore, the ideal tool will vary by clinical context. The intensive care unit (ICU) is arguably the most demanding context of all, where CO can be highly dynamic, the degree of vascular access varies, patient physiology covers the gamut, measurements are made by staff with a range of technical expertise, and the results impact decisions with the highest of stakes.

The de facto standard for assessing CO in the cardiac intensive care unit (CICU) is the single-bolus thermodilution (TD) technique using a pulmonary artery catheter (PAC). For this reason, most new technologies benchmark their performance against TD, but the technique has some well-known limitations. Its availability may be limited, as PAC use is less common outside of CICUs. Its accuracy is often questioned in patients with significant TR, cardiac shunt physiology, or very low CO.[4] And its execution can be technically demanding, with subtleties such as confounding sources of heat transfer and a need for technical replicates, nuances that are less familiar outside the cardiac catheterization lab (CCL). The clinical gold standard for assessing CO is the "direct Fick" technique, wherein CO is computed from the Fick principle using resting measurements of a patient's oxygen consumption ($\dot{V}_{O_2}$) and other variables.[5,6] The tradeoffs of direct Fick are also well-known. It is accurate and applies to a wide swath of patient physiologies, but it is also more time consuming and requires a PAC together with specialized equipment for measuring $\dot{V}_{O_2}$, making it less accessible and less familiar in the ICU compared to TD. A technique that combines the universal applicability of direct Fick with high accuracy, compatibility with a range of vascular access, broad availability, and ease of use remains a platonic ideal.

One of the most widely used alternatives to TD, "estimated Fick" cardiac output (eFick), aims to capture the wide applicability of direct Fick but without the technical drawbacks. This approach entails the same calculation as the gold standard technique with 1 key exception – $\dot{V}_{O_2}$ is estimated rather than measured. When $\dot{V}_{O_2}$ is estimated well, eFick is accurate, easy to use, applies to a broad spectrum of patients, and is widely accessible in patients with a PAC. But when $\dot{V}_{O_2}$ is estimated poorly, eFick's advantage becomes its Achilles heel. Accurate $\dot{V}_{O_2}$ estimation is especially challenging in the ICU, the context of perhaps its most widespread use.[7-10] None of the commonly used $\dot{V}_{O_2}$ prediction equations were derived in ICU patients but rather in CCL patients.[11-13] And these CCL patients were largely comprised of children and young adults with congenital heart disease (**Table 1**). Furthermore, these prediction equations were derived 30 to 60 years ago with basic modeling techniques. Finally, these equations employ few if any $\dot{V}_{O_2}$ predictors that reflect a patient's dynamic physiology, making them ill-suited for the ICU. Despite these widely recognized limitations the use of eFick endures, primarily due to its simplicity and its broad applicability across conditions such as TR. The modern availability of high-resolution clinical data sets together with powerful modeling tools should enable novel CO estimators that retain the advantages of eFick while improving performance.

Here we sought to derive improved estimators of CO with key advantages over both eFick and TD, especially in the ICU (**Central Illustration**). We aimed to achieve accuracy, ease of use, and applicability across a wide spectrum of patients. We also sought to derive estimators with the versatility to be used in many kinds of ICU patients, in particular those in whom vascular access falls short of a PAC. To derive and validate these CO estimators, we used readily available clinical predictors drawn from large data sets curated from 2 distinct institutions. Improving CO estimation in the ICU would enable numerous opportunities to enhance patient care, from earlier detection of impending circulatory failure to gauging the efficacy of therapy.

METHODS

PATIENTS. The patients studied here were cared for in 4 settings, 2 distinct CCLs and 2 distinct CICUs. The 2 CCL data sets were derived from consecutive right heart catheterizations performed at Massachusetts General Hospital (MGH), between 2013 and 2022 (CCL data set), or at an external institution, Brigham and

TABLE 1 Patient Cohorts Used to Derive $\dot{V}_{O_2}$ or CO Models

Data Set	Dehmer	LaFarge	Bergstra	CCL	CSICU
Site	CCL	CCL	CCL	CCL	CSICU
Patients	108	879	250	1,026	2,311
Measurements	108	879	250	1,171	7,026
Dates	Before 1982	1961-1966	Before 1995	2013-2020	2016-2022
CO technique	TD	Fick calculation	Dye dilution	TD	TD
$\dot{V}_{O_2}$ technique	Fick calculation	Douglas bag	Fick calculation	Fick calculation	Fick calculation
Age, y	49 (–)	13 (–)	35 (23)	65 (14)	64 (12)
Age: min-max	21-73	3-40	1-83	20-96	16-91
Female: n (%)	39 (36)	363 (41)	108 (43)	372 (36)	669 (28)
BSA, m^2	-	-	1.62 (0.39)	1.93 (0.26)	1.96 (0.26)
$\dot{V}_{O_2}$, mL/min	-	-	231 (60)	217 (65)	207 (67)
$\dot{V}_{O_2}I$, mL/min/m^2	126 (26)	139 (26)	-	112 (28)	105 (30)

Comparison of data sets used to derive the CO prediction models in this study or the $\dot{V}_{O2}$ prediction models from 3 published studies, those of Dehmer et al,[12] LaFarge and Miettinen,[13] or Bergstra et al.[11] Note that 116 patients were present in both the CCL and CSICU data sets. The label "Fick calculation" refers to solving for CO or $\dot{V}_{O_2}$ using the Fick equation with measured values of all independent variables. The average of each continuous variable in this table was taken over all available measurements. Values are counts, mean (SD), or counts (%) where indicated.

BSA = body surface area; CCL = cardiac catheterization lab; CO = cardiac output; CSICU = cardiac surgery intensive care unit; TD = thermodilution; $\dot{V}_{O_2}$ = resting oxygen consumption; $\dot{V}_{O_2}I = \dot{V}_{O_2}$ indexed to BSA.

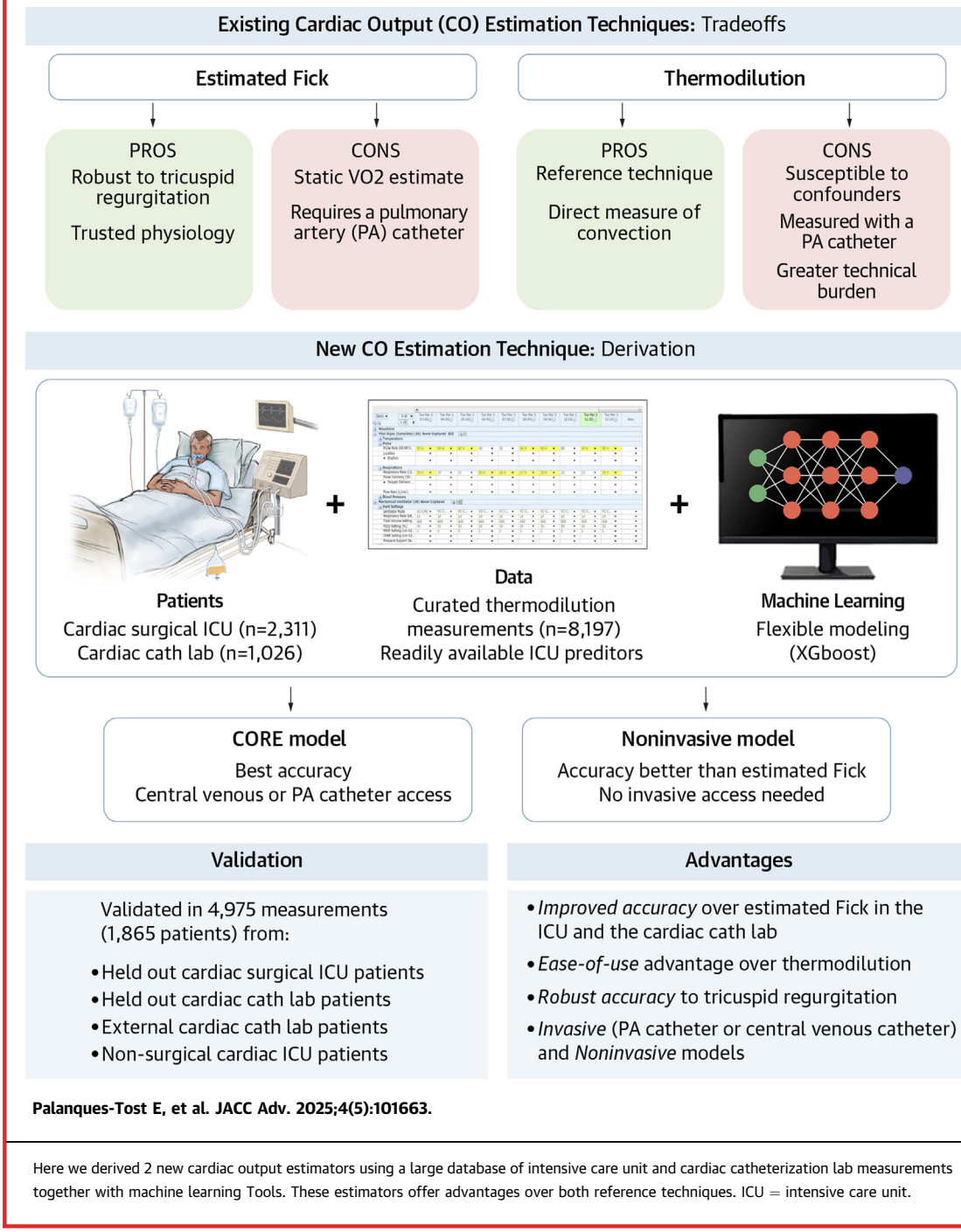
Women's Hospital, between 2009 and 2022 (external cardiac catheterization lab [CCL_{ext}] data set). The 2 ICU data sets were derived from patients cared for at MGH between 2010 and 2022, either in a postcardiac surgery intensive care unit (CSICU data set), or a distinct CICU (CICU data set), located on a separate floor, where a modern spectrum of nonsurgical cardiac critical illness is treated. A third nonoverlapping ICU data set was derived from patients in whom $\dot{V}_{O_2}$ had been measured directly with a metabolic cart. This data set was comprised of 102 nonventilated patients from both the CSICU and the CICU (direct Fick data set). This study was approved by the MGH Institutional Review Board (Protocol 2020P003053).

CARDIAC OUTPUT AND $\dot{V}_{O_2}$ ASSESSMENT. Three techniques for CO assessment were used for different tasks: model derivation, benchmarking, or sensitivity analysis. First, to derive new CO models, the ground-truth measurements of CO were made by TD with a PAC and a room-temperature injectate. These TD measurements were curated to enhance their reliability for model training. In particular, patients with known moderate or severe TR were excluded, as assessed by an echocardiogram performed within 2 weeks of the TD measurement; valvular regurgitation was assessed by integrating both quantitative and qualitative parameters in accordance with society guidelines.[14] Measurements were also excluded if the cardiac index was <1 or >5, with body surface area estimated by the DuBois formula. Second, to benchmark these new CO models, CO was estimated by 3 commonly used eFick formulas, those derived in LaFarge and Mittenten, Bergstra et al, and Dehmer et al.[11-13] Third, to assess the sensitivity of new CO models to the presence of TR, CO was determined by the direct Fick method in a subset of ICU patients (direct Fick data set described above). In these patients, CO was calculated using the Fick Equation with $\dot{V}_{O_2}$ measured by metabolic cart (Medgraphics) and averaged over a 3-minute interval, in duplicate. For several analyses $\dot{V}_{O_2}$ was also assessed by a second technique, namely calculation from the Fick equation using measured values of all independent variables including CO by TD. The calculation and clustering of CO or $\dot{V}_{O_2}$ trajectories is described in the Supplemental Methods.

CARDIAC OUTPUT PREDICTORS. Three classes of predictors were used for fitting CO models: clinical factors, vital signs, and laboratory measurements. Clinical factors included age, sex, height, body surface area, and active ventilator use (true or false). Vital signs included body temperature, heart rate, arterial blood pressure (systolic, diastolic, and mean), and arterial as well as mixed venous O_2 saturations. Laboratory measurements included hemoglobin; creatinine and lactate were also evaluated but not found to be informative as predictors (data not shown). Two configurations of these 13 predictors were evaluated, corresponding to those derivable from invasive vs noninvasive vascular access (Supplemental Table 4). Details of data set assembly are described in the Supplemental Methods.

SUPERVISED LEARNING OF CARDIAC OUTPUT MODELS. We trained 2 primary CO models, one with predictors whose measurement required invasive vascular access (pulmonary artery or central vein)

CENTRAL ILLUSTRATION **The Estimation of Cardiac Output Using Existing Techniques, Thermodilution and the Estimated Fick Approximation, Entails Tradeoffs Related to Accuracy and Utility**



Palanques-Tost E, et al. JACC Adv. 2025;4(5):101663.

Here we derived 2 new cardiac output estimators using a large database of intensive care unit and cardiac catheterization lab measurements together with machine learning Tools. These estimators offer advantages over both reference techniques. ICU = intensive care unit.

and the other with predictors obtainable noninvasively. We labeled the first, Catheter Optimized caRdiac output Estimation (CORE) model, and the latter, Noninvasive model. Both models were trained on a combined data set of patients from the MGH CCL and MGH CSICU (Supplemental Figures 1 and 6).

To train models we adopted a standard machine learning framework using reliable TD-CO measurements as ground-truth targets. We split patient data into training and test sets. The training set contained 80% of the data and was used for fitting and hyperparameter tuning where applicable. Algorithms with hyperparameters were tuned using Bayesian optimization with 3-fold cross-validation on the training set. Hyperparameter-optimized models were subsequently trained on the full 80% training-split of the data. The test set with the remaining 20% of the data was held out of the training process and used exclusively for model evaluation. To prevent data leak from influencing model performance, all measurements from a given patient were assigned to only one of the training or test sets. We compared several model fitting algorithms, including linear regression, support vector machine, neural network, random forest, and gradient boosting (XGBoost).

To evaluate models we first considered their performance on the holdout test set. Next, we assessed each model's ability to generalize beyond the distribution of the derivation data (MGH CCL and MGH CSICU) by evaluating its performance on 2 external validation sets, CCL_{ext} and CICU. We refer to them as external in so much as they were both drawn from distinct data distributions, in one case from patients at a distinct institution, the Brigham and Women's Hospital CCL (CCL_{ext}), and in the other from patients from a distinct ICU at MGH (CICU) where a distinct spectrum of nonsurgical cardiac critical illness is managed. No data from these external validation sets were used in any way for training models. Finally, we evaluated the CORE model's sensitivity to the presence of TR by comparing its CO prediction against direct Fick-CO using the ICU direct Fick data set described above. This data set was not used for training the CORE model.

STATISTICS. Continuous measurements are presented as the mean ± SD unless otherwise stated. To compare the performance of one CO model relative to another, mean absolute percentage errors (MAPE) between predictions and ground-truth CO were computed at test set points for both models. The paired Wilcoxon signed-rank test was then used to evaluate whether the new model's prediction errors were significantly different than the reference model's errors at corresponding points ($P < 0.05$, 2-sided). Statistical significance of area under the receiver operator curve (AUROC) differences was assessed via bootstrapping with 10,000 resamples of the test set. All data and statistical analyses were performed in Python.

TABLE 2 ICU Patient Characteristics

	CSICU (n = 2,969)	CICU (n = 451)
Outcomes		
In-hospital mortality (%)	8.0	6.9
Median length of stay (d)	2.0	4.2
Comorbidities and ICU therapies		
Mechanical ventilation (%)	87	66
Vasopressor use (%)	93	96
IABP use (%)	6.1	16
History of diabetes mellitus (%)	22	24
History of CABG (%)	22	18
History of CKD (%)	10	9.1
History of COPD (%)	6.9	6.2
History of CVA/TIA (%)	10	9.5
History of HF (%)	22	33
History of MI (%)	11	16
History of PCI (%)	7.5	17
History of PVD (%)	4.9	6.7
History of hypertension (%)	33	24
Anthropometrics		
Female (%)	28	22
Age (y)	65 ± 13	64 ± 13
BMI (kg/m^2)	28 ± 5.8	28 ± 6.3
BSA (m^2)	2.0 ± 0.25	2.0 ± 0.26
Height (m)	1.7 ± 0.10	1.7 ± 0.10
Weight (kg)	85 ± 20	83 ± 21

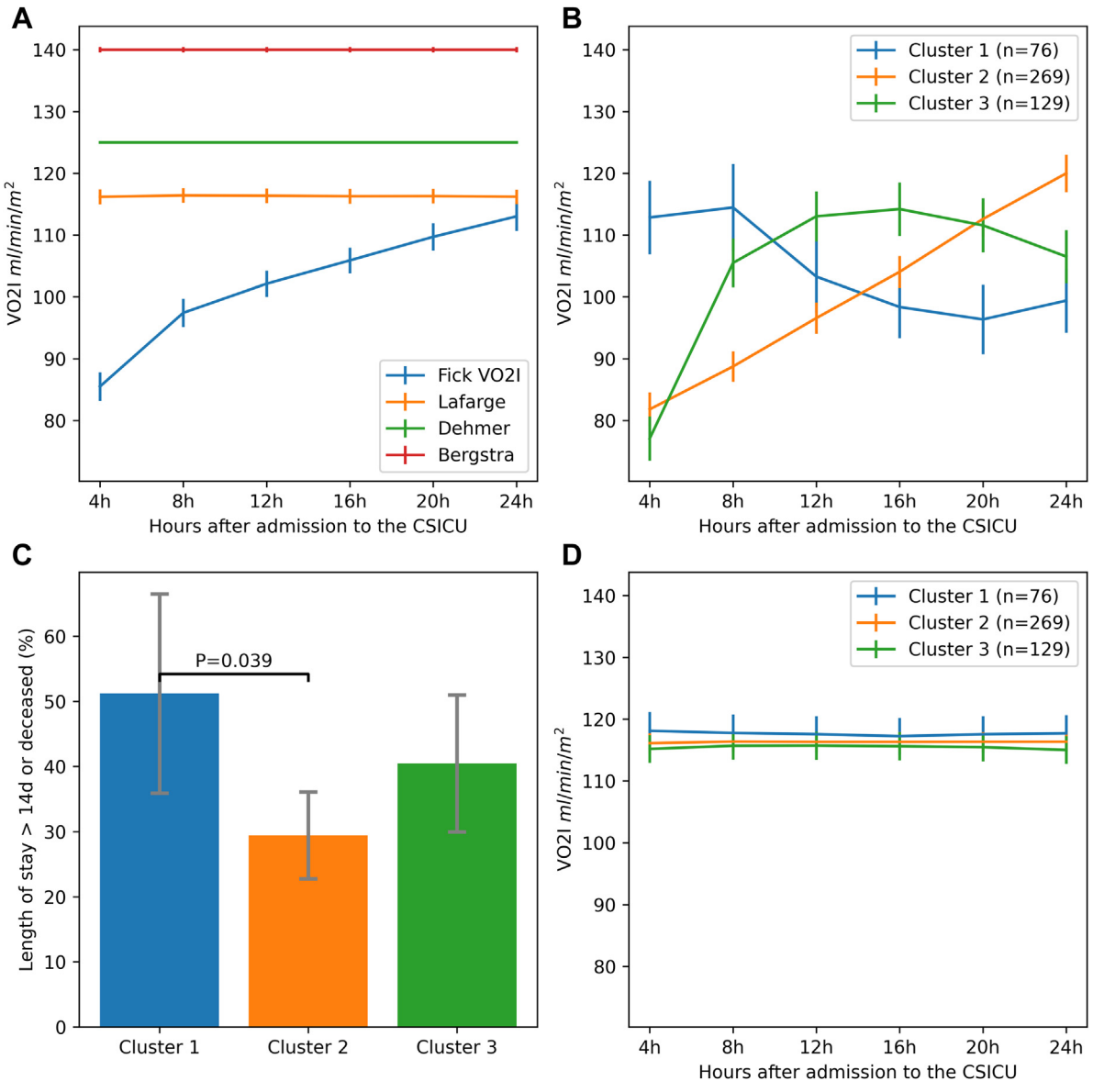
Values are % or mean ± SD unless otherwise indicated. Clinical characteristics of the ICU patients used to derive (CSICU) and validate (CSICU, CICU) the CO estimation models. Values are percent, median, or mean (SD) where indicated.

BMI = body mass index; BSA = body surface area; CABG = coronary artery bypass graft surgery; CICU = cardiac intensive care unit; CKD = chronic kidney disease; CO = cardiac output; COPD = chronic obstructive pulmonary disease; CSICU = cardiac surgery intensive care unit; CVA = cerebrovascular accident; HF = heart failure; IABP = intra-aortic balloon pump; ICU = intensive care unit; MI = myocardial infarction; PCI = percutaneous coronary intervention; PVD = peripheral vascular disease; TIA = transient ischemic attack.

RESULTS

To derive novel CO prediction models, we used a total of 8,197 ground-truth TD measurements from 3,221 unique patients, a data set that is an order of magnitude larger than those used to derive each of 3 commonly used eFick models (Tables 1 and 2).[11-13] The patients whose data were used for model derivation were cared for in 2 clinical contexts at our institution, the CSICU and the CCL, dating back no more than 12 years. The patients had a mean age of 65 with a range of 16 to 96, typical of adult cardiovascular care. To externally validate our CO prediction models we

FIGURE 1 **Oxygen Consumption Dynamics in the Cardiac Surgery Intensive Care Unit**



(A) Time course of oxygen consumption after admission to the cardiac surgery intensive care unit, averaged over 474 patient admissions, where $\dot{V}_{O_2}I$ was either determined from the Fick principle using cardiac output measured by thermodilution ("Fick $\dot{V}_{O_2}I$") or estimated using one of 3 prediction models: LaFarge, Dehmer, or Bergstra. (B) Postoperative trajectories of $\dot{V}_{O_2}I$, determined from measured thermodilution cardiac output in the same patients as a, but clustered into 3 groups using the K-Means algorithm. (C) Fraction of patients with prolonged hospitalization (>14 days) or death, stratified by $\dot{V}_{O_2}I$ cluster as calculated in B. (D) $\dot{V}_{O_2}I$ trajectories for each cluster of patients described in B, but estimated using the LaFarge equation for predicted oxygen consumption. Error bars reflect 95% CIs. CSICU = cardiac surgery intensive care unit; $\dot{V}_{O_2}I$ = indicates resting oxygen consumption indexed to body surface area.

used 2 data sets containing 2,776 TD measurements. These validation data sets spanned 2 distinct modern-day contexts, a nonsurgical CICU and a CCL from an external institution (**Table 2**, Supplemental Table 1).

We noted several characteristics of the patients in our data sets that could influence resting $\dot{V}_{O_2}$ and potentially force existing eFick models to extrapolate far beyond their derivation context. Compared to CCL patients from the eFick derivation cohorts, the CCL patients in this study were typically much older, the vast majority did not suffer from congenital heart disease, and their spectrum of cardiovascular disease

TABLE 3 Performance of Cardiac Output Prediction Models

		CSICU (N = 1,823)			CICU (N = 2,055)			CCL (N = 376)			CCL_{ext} (N = 721)		
Model	**Training data**	**MAPE**	**MAE**	**R^2**	**MAPE**	**MAE**	**R^2**	**MAPE**	**MAE**	**R^2**	**MAPE**	**MAE**	**R^2**
LaFarge	-	30	1.5	−1.5	26	1.3	−0.86	19	0.97	0.19	19	0.95	0.34
Dehmer	-	36	1.7	−2.1	31	1.5	−1.2	26	1.2	−0.04	19	0.90	0.42
Bergstra	-	47	2.2	−3.9	42	2.0	−2.7	36	1.7	−0.84	25	1.1	0.18
CORE	CSICU and CCL	**14**‡	**0.72**	**0.58**	**16**‡	**0.76**	**0.58**	**17**†	**0.84**	**0.51**	**16**‡	**0.80**	**0.49**

Model performance as judged by MAPE, MAE (in L/min), or R^2 was calculated for each of 3 eFick models (rows 1-3) and the CORE model fitted here. Each model was evaluated on 4 data sets (table header). TD-CO was used as ground truth. All models were evaluated on CSICU and CCL test data that had been set aside and not used for CORE model training. All models were evaluated on the entirety of the external validation data sets, CICU and CCL_{ext}, none of which had been used for model training. Table entries in bold indicate the best fitting model on a data set (column). Significant differences (Wilcoxon test) in the MAPEs of the CORE model compared to eFick-LaFarge on each data set (columns) are reported: $^{*}P < 0.05$, $^{†}P < 0.01$, $^{‡}P < 0.001$. TD-CO indicates cardiac output measured by thermodilution.

CCL = cardiac catheterization lab; CCL_{ext} = external cardiac catheterization lab; CICU = cardiac intensive care unit; CO = cardiac output; CSICU = cardiac surgery intensive care unit; MAE = mean absolute error; MAPE = mean absolute percentage error; N = the number of data records in the evaluation data set (test set or external validation set); R^2 = coefficient of determination; TD = thermodilution.

reflects modern day as opposed to the 1960s to 1980s. The ICU patients in this study presented an even starker contrast. Compared to our institution's CCL (and by extension the eFick derivation cohorts), they exhibited lower mean arterial pressure and hemoglobin levels, as well as higher heart rates, ventilator usage rates, and male sex ratio (Supplemental Table 1). These factors and more have the potential to influence resting $\dot{V}_{O_2}$ but are absent from existing eFick models of $\dot{V}_{O_2}$.

EFICK ESTIMATES OF $\dot{V}_{O_2}$ DYNAMICS AND CO. Given the pivotal role of $\dot{V}_{O_2}$ in the Fick calculation of CO we first evaluated the accuracy of eFick estimates of $\dot{V}_{O_2}$ in ICU patients. For each of 474 patients in the CSICU we computed ground-truth values of $\dot{V}_{O_2}$ from the Fick equation using measured TD-CO. Serial $\dot{V}_{O_2}$ values were assembled into a trajectory covering the first 24 hours of ICU admission and indexed to body surface area ($\dot{V}_{O_2}I$). The mean postoperative trajectory of ground-truth $\dot{V}_{O_2}I$ demonstrated marked deviation from the mean trajectory of $\dot{V}_{O_2}I$ calculated using eFick, both in absolute value and in trajectory shape, no matter which eFick model was used (**Figure 1A**). The inability of eFick $\dot{V}_{O_2}$ models to capture these observed dynamics would be expected to undermine their estimates of CO. Interestingly, we also discovered meaningful substructure within the ensemble of individual postoperative $\dot{V}_{O_2}I$ trajectories by using unsupervised K-means clustering (**Figure 1B**). We resolved patient trajectories into 3 subtypes of postoperative recovery that carried important prognostic implications. In particular, the fraction of patients experiencing prolonged hospitalization across these subtypes spanned a nearly 2-fold range, $P = 0.039$ (**Figure 1C**), though the mechanism of these outcome differences will require future study. Moreover, all substructure was lost when the $\dot{V}_{O_2}I$ s from each trajectory subtype were recalculated using eFick-LaFarge, the most dynamic of the $\dot{V}_{O_2}$ prediction models (**Figure 1D**).

We next quantified the accuracy of eFick models for CO prediction itself, evaluating the models across both our derivation and external validation data sets (**Table 3**, rows 1-3). Using a ground truth CO assessed by TD and performance metrics of MAPE, mean absolute error (MAE), or coefficient of determination (R^2), we found that the accuracy of any given eFick method was always substantially worse among ICU patients than CCL patients. Among eFick models, we found that the LaFarge model had the best performance in both the ICU and the CCL. In absolute terms, however, its accuracy in the ICU was concerning, with MAPE values of 30% in the CSICU and 26% in the CICU. Moreover, more than 40% of all CO estimates in the ICU using eFick-LaFarge had an absolute percentage error >20% (Supplemental Figure 2). Perhaps most problematic, the negative R^2 of the LaFarge model on both ICU data sets indicated a worse fit than simply assigning each patient's CO to be the mean CO across the data set. We also tried refitting the LaFarge model from scratch on our CSICU data, but the results were underwhelming—for example, MAPE only improved from 30% to 25% on the CSICU test set and the R^2 remained negative (Supplemental Table 2).

NEW CO PREDICTION MODELS. We sought to derive new CO prediction models with greater accuracy than eFick and good performance across both the ICU and the CCL. We took a more direct approach, predicting CO itself rather than $\dot{V}_{O_2}$, which proved just as effective as the 2-step paradigm of eFick (Supplemental Table 3). Our training data consisted of curated TD measurements of CO as ground-truth paired with 13 CO predictors (Methods, Supplemental Table 4) that are routinely measured, including a subset that require a PAC to collect. Notably, none of the predictor measurements should be undermined by conditions

TABLE 4 Performance of Cardiac Output Models Compatible With Distinct Vascular Access

	Model	CSICU			CICU			CCL			CCL_{ext}		
Vascular access	Metric	MAPE	MAE	R^2	MAPE	MAE	R^2	MAPE	MAE	R^2	MAPE	MAE	R^2
CVC		n = 119			n = 50			n = 376			n = 414		
	LaFarge	40	1.9	−1.6	36	1.7	−1.8	29	1.5	−5.1	24	1.3	−0.25
	CORE	19‡	0.87	0.56	21*	0.92	0.53	19‡	0.94	0.39	17‡	0.96	0.30
Noninvasive		n = 1,823			n = 2,055			n = 376			n = 721		
	Noninvasive model	18‡	0.88	0.40	20‡	0.97	0.34	20	1.0	0.24	21	1.1	0.15

Performance of CO models using data available from a CVC or noninvasive vascular access. To assess the performance of the CORE model with CVC data (row 2), the model was evaluated with $ScvO_2$ substituted for $SmvO_2$ in the model input; no training with $ScvO_2$ was performed. For comparison, the eFick-LaFarge model was also evaluated with $ScvO_2$ substituted for $SmvO_2$ (row 1). The noninvasive model was trained on the same data as the CORE model and compared to eFick-LaFarge evaluated with $SmvO_2$ data (Table 3). Significant differences (Wilcoxon test) in the MAPE of the CVC and noninvasive models compared to eFick-LaFarge on each test data set (columns) are indicated: *$P < 0.05$, †$P < 0.01$, ‡$P < 0.001$.

CCL = cardiac catheterization lab; CCL_{ext} = external cardiac catheterization lab; CO = cardiac output; CORE = Catheter Optimized caRdiac output Estimation; CSICU = cardiac surgery intensive care unit; CVC = central venous catheter; MAE = mean absolute error; MAPE = mean absolute percentage error; N = the number of data records on which the models were evaluated; $ScvO_2$ = central venous oxygen saturation; $SmvO_2$ = mixed venous oxygen saturation; R^2 = coefficient of determination.

that could compromise TD accuracy, such as TR or low CO. We trained a model on a combined set of CCL and CSICU patients using a range of techniques (Supplemental Table 5) and found that gradient boosting (XGBoost) performed the best. This model, which we labeled CORE (Methods), outperformed all 3 eFick methods, achieving a lower MAPE and MAE, as well as a higher R^2 often by a substantial margin (**Table 2**, Supplemental Tables 9, and 11, Supplemental Figure 5). The gains were most evident in the CSICU where the CORE model achieved a MAPE of 14% compared to 30% for eFick-LaFarge ($P < 0.001$, CSICU test set), a relative improvement in accuracy of 53%. In the CICU the CORE model performed similarly well, achieving a MAPE of 16% compared to 26% for eFick-LaFarge ($P < 0.001$). We also trained 2 context-specific CO prediction models on site-specific populations, namely on CSICU patient alone or CCL patients alone. The CORE model performed as well or better than the CSICU-specific model in the ICUs and the CCL-specific model in the CCLs (Supplemental Table 6).

To improve the versatility of CO estimation we considered models that could be used in patients without a PAC. We first sought a model that could be used with predictors obtainable from a central venous catheter (CVC), as might be found in a non-cardiac ICU. We reasoned that central venous oxygen saturation ($ScvO_2$) would be a key predictor[15] because feature importance analysis of the CORE model had revealed that mixed venous oxygen saturation ($SmvO_2$) was its most valuable predictor (Supplemental Table 7). However, our data set lacked sufficient $ScvO_2$ measurements to train such a model from scratch. Instead, we simply substituted $ScvO_2$ for $SmvO_2$ as an input to the CORE model. Model accuracy in the CSICU dropped slightly with this substitution, for example MAPE increased from 14% to 19% (**Table 4**). However, the CORE model with $ScvO_2$ still significantly outperformed eFick-Lafarge in both the CSICU and the CICU. We also considered the scenario with no invasive vascular access and fit a "noninvasive" CO model. Although its overall performance slipped compared to the CORE model, the Noninvasive model outperformed eFick-LaFarge in the ICU context (eg MAPE 18% vs 30% in the CSICU, $P < 0.001$) even though eFick made use of $SmvO_2$ from a PAC.

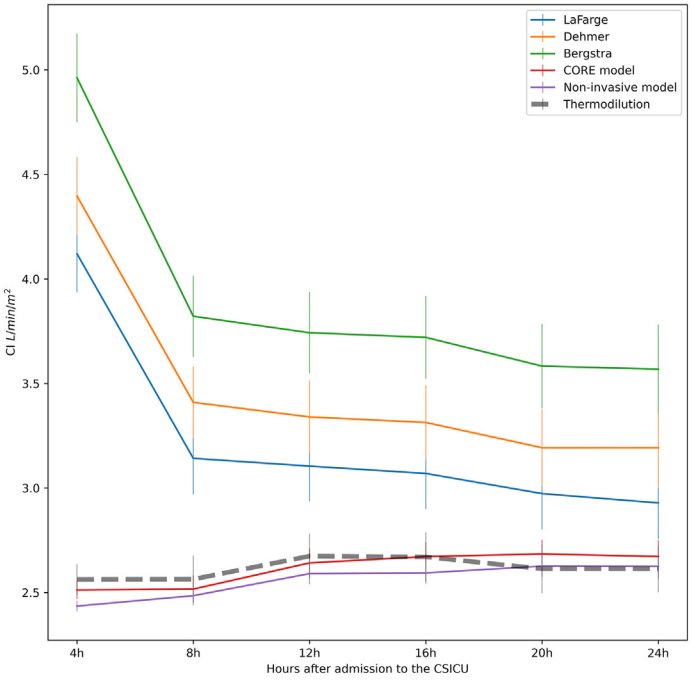
To test CORE's accuracy in a key patient subgroup that may confound the TD technique, we explored its sensitivity to the presence of TR. Our data set contained 108 measurements from 102 ICU patients in whom resting $\dot{V}_{O_2}$ was directly measured together with echocardiographic assessment of TR. We stratified these patients by the severity of their TR, calculated their ground-truth CO using the direct Fick technique, and then compared these values to CO predictions from the CORE model. Encouragingly, the performance of the CORE model in the ICU was unaffected by the presence of moderate or severe TR (N = 43), with a MAPE of 16% and an R^2 of 0.65 (Supplemental Table 8).

We went on to reanalyze postoperative trajectories in the CSICU but this time through the lens of CO dynamics. We computed the mean cardiac index over the first 24 hours after cardiac surgery as calculated by eFick methods, the CORE model, or the Noninvasive model and compared it to cardiac index measured by TD (**Figure 2**). Strikingly, cardiac index by eFick was not only error prone, but also its dynamics after cardiac surgery diverged from the TD-measured values. These errors may reflect the influence of low postoperative body temperature, anesthetic usage, and mechanical ventilation on $\dot{V}_{O_2}$, effects not captured by

the eFick models which could lead them to overestimate $\dot{V}_{O_2}$ and in turn, cardiac index. In contrast, the cardiac index values predicted by the CORE model or even the Noninvasive model tracked the TD measurements with much higher fidelity.

Finally, we evaluated whether the CORE and Noninvasive models derived here have the potential to improve clinical decision-making in the ICU. Compared to eFick, the new CO models proved to be better at diagnosing low cardiac index (**Figure 3** and Supplemental Figure 3). For example, in the CSICU population a classifier based on the CORE model achieved an AUROC of 0.88, better than eFick-LaFarge which achieved a value of 0.73 ($P < 0.001$). The CORE model was also better at tracking the direction of change in CO. We computed the fraction of instances where a CO change by TD (of a meaningful magnitude) also changed in the same direction by the CO estimate. This concordance rate between TD and the CORE model was 85% in the CSICU and 76% the CICU (**Figure 3**), but between TD and eFick-LaFarge the concordance rate was only 72% in the CSICU and 68% in the CICU (Supplemental Figure 4).

FIGURE 2 Cardiac Index Dynamics in the Cardiac Surgery ICU



Mean trajectories of cardiac index for the first 24 hours after cardiac surgery. Each trajectory reflects cardiac index averaged over 562 patient admissions in the cardiac surgery intensive care unit test set, as measured by thermodilution (black) or estimated by one of five models—3 eFick models (LaFarge, Dehmer, Bergstra) and 2 models derived here, the Catheter Optimized caRdiac output Estimation model and the noninvasive model. Error bars reflect variation in a model's mean predictions across patients (2 SEs). CORE = Catheter Optimized caRdiac output Estimation; CSICU = cardiac surgery intensive care unit.

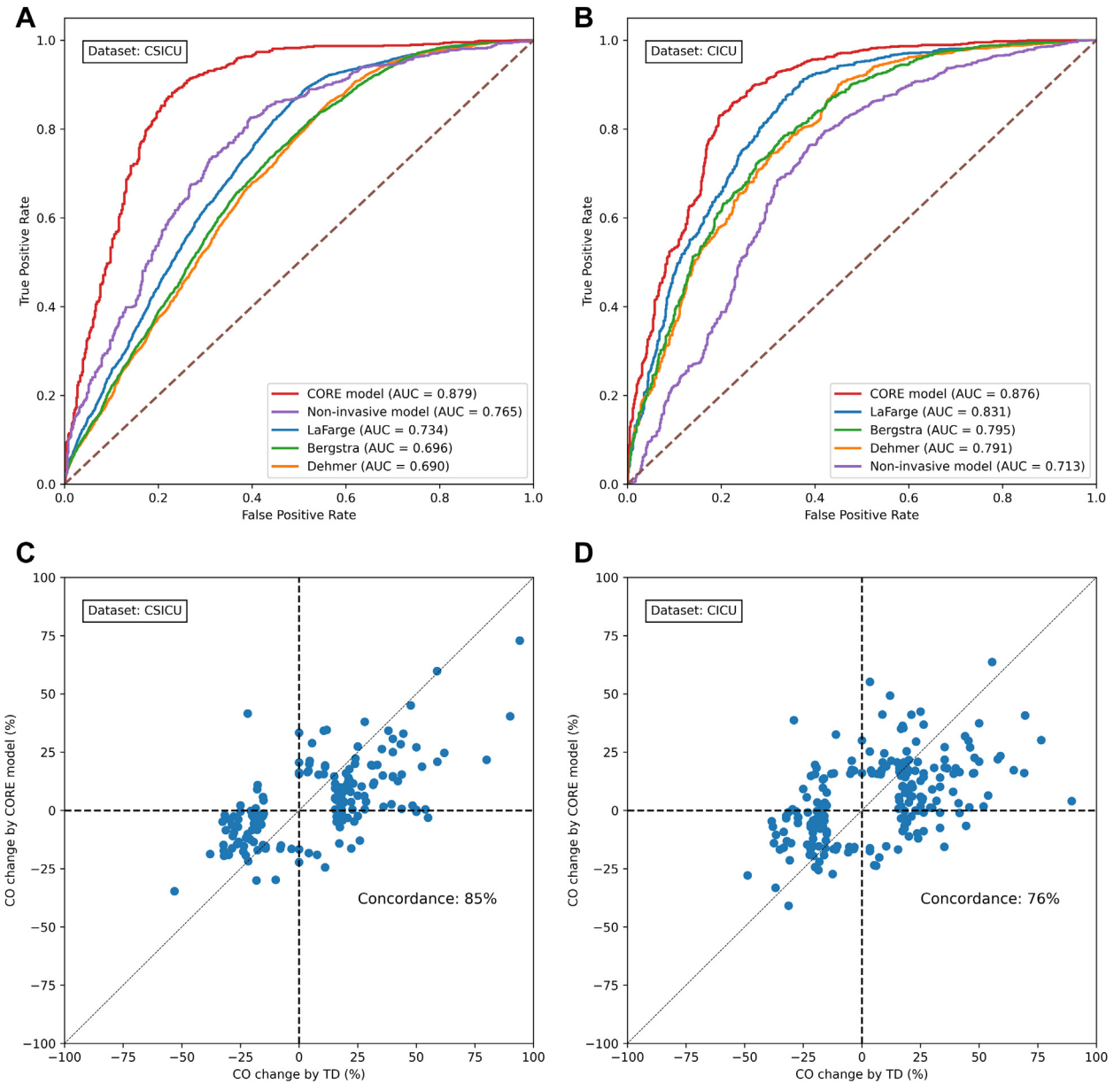
DISCUSSION

From calculating aortic valve area to managing shock, the assessment of a patient's CO is a cornerstone of cardiovascular care. In this study we developed several novel estimators of CO by applying modern machine learning tools to a data set of 4,825 patients and 13,172 ground-truth measurements of CO, together with commonly measured clinical predictors. Trust in these estimators stems in part from the large and diverse patient population on which they were trained and validated, including stable and unstable patients, surgical and nonsurgical patients, and patients from local and external institutions. Compared to widely used eFick models these estimators are markedly more accurate in the ICU, more effective at classifying low CO, and more faithfully track changes in TD-CO. Compared to TD they are easier to use, apply to a broader array of patient physiologies, and can be used with or without a PAC.

ESTIMATING CARDIAC OUTPUT. The drawback of existing eFick methods for estimating CO can be traced to their poor estimation of resting $\dot{V}_{O_2}$. Though this problem is well recognized in the literature,[16-22] we found that its magnitude was truly unmasked in the ICU. A key reason eFick methods fail is that their $\dot{V}_{O_2}$ models are based on very few predictors. Moreover, because these predictors are essentially static (except for heart rate, in eFick-LaFarge) they struggle to track the physiology associated with changes in $\dot{V}_{O_2}$. As a result, the eFick models are implicitly and rigidly adapted to the CCL patients they were trained on, capturing their mean $\dot{V}_{O_2}$ but unable to track $\dot{V}_{O_2}$ dynamics. By contrast, modern ICU patients are a far different population from the eFick-derived CCL patients (**Table 1**) and $\dot{V}_{O_2}$ can in fact be quite dynamic (**Figure 1**).[7-10,18] Consequently, eFick predictions of CO can fail in dramatic fashion (**Figure 2**).

To improve CO estimation, we used an expanded set of predictors together with flexible models to better fit complex relationships. In addition to the CORE and noninvasive models, we also developed CO models for distinct clinical contexts, namely an ICU-specific model or a CCL-specific model, thereby

FIGURE 3 **Classifying and Tracking Cardiac Output in the Intensive Care Unit**



(A) Receiver operator curves for the classification of cardiac index as low, <2.5 L/m^2, in the CSICU. Each curve reflects a classifier based on a distinct cardiac output prediction model. (B) Receiver operating curves for the same models as in A, except applied to cardiac index classification in the CICU. (C) Concordance plot for comparing changes in cardiac output as measured by thermodilution with changes estimated by the CORE model, evaluated on the CSICU test data. Cardiac output changes were calculated from consecutive measurements within 5 hours, and cardiac output changes < 15% in magnitude were excluded. Points in the upper right and lower left quadrants reflect concordant changes. (D) Concordance plot as in C but evaluated on the CICU data set. AUC = area under the receiver operator curve; CICU = cardiac intensive care unit; CO = cardiac output; CORE = Catheter Optimized caRdiac output Estimation; CSICU = cardiac surgery intensive care unit; TD = thermodilution.

capturing contextual factors implicitly by virtue of the patients the model was trained on. An ideal CO model would be based on sufficient predictors to genuinely fingerprint the physiology and thereby avoid the need for a collection of context-specific models. The CORE model described here, trained on a rich set of predictors and patients from both the CCL and ICU, is a step in that direction, as it was able to

match the performance of both context-specific models (Supplemental Table 6).

All methods for estimating CO have limitations and assumptions that must be considered when interpreting the results. In our CO prediction models, a perhaps surprising finding was that a CO model trained to estimate reliable TD values does not necessarily inherit the limitations of TD more broadly. For example, though TD measurements can theoretically be confounded by TR or low CO,[23,24] the predictors we used to train our CO models do not share these shortcomings. Indeed, part of our motivation was to fit a CO model that would be immune to the weaknesses of TD and thereby mirror the broad applicability of direct Fick. We were able to validate this by showing that in 102 of our ICU patients in whom direct Fick CO was available (ground truth established using measured $\dot{V}_{O_2}$), the CORE model performed similarly in patients with and without moderate or severe TR. Thus, the prediction approach at the heart of the CORE model merges several of the benefits of direct Fick with those of TD.

How accurate does a CO estimator need to be? Goals for absolute accuracy have long been recognized as a matter of judgment. We therefore explored 2 clinical use cases for CO estimators that shed light on their utility for decision-making. First, we built a classifier to diagnose low CO (cardiac index < 2.5) and found that the CORE model consistently outperformed eFick (**Figure 3**, Supplemental Figure 3, Supplemental Table 10). Second, we evaluated the ability of CORE to track changes in CO, assessing how often a change in CO predicted by the model matched the direction of change measured by TD. Again, the CORE model outperformed eFick, achieving a concordance rate with TD of 85% and 76% in the CSICU and CICU respectively.

WEIGHING TRADEOFFS. The most accurate of the CO models developed here, the CORE model, was trained with an $SmvO_2$ predictor measured by a PAC, raising the question of when it best complements or substitutes for a TD measurement. First, the CORE model can be used in patients without a PAC. Substituting $SmvO_2$ with $ScvO_2$ in the calculation entails only a modest drop in accuracy. This flexibility enables the model to be uniquely accessible across a wide swath of ICUs where central venous catheters predominate. Second, in a sizeable subset of patients with a PAC, the accuracy of TD may be a concern due to the presence of TR or very low CO. Though the magnitude and direction of TD error in these settings is debatable,[4-6,23-29] the theoretical risk of error is often sufficient for physicians to adopt a conservative approach and avoid TD altogether. Fortunately, the accuracy of the CORE model is resistant to these TD confounders on theoretical grounds, and empirically, we found it to be resistant to the most common of them, TR (Supplemental Table 8). Third, even in patients with PAC access, lack of technical familiarly with the TD measurement itself can be a concern, including with such factors as rapid saline injections, heat transfer confounders, and technical replicates. These concerns may be especially pronounced in ICUs where PAC use is less typical even though CVC use is common. The CORE model is easier to use in that regard as it requires nothing more than blood sampling from the PAC or CVC. Finally, for any individual patient the reliability of TD is at times difficult to judge. A clue that such a measurement may be problematic is when its value is at odds with the clinical picture, an all too familiar scenario. In such cases a complimentary assessment by a CO estimator such as the CORE model could prove valuable.

The estimated Fick technique remains a popular way to assess and track CO, despite warnings from published reports.[6,13,28,30-33] Relative to TD, eFick's appeal likely stems from 3 properties—its ease of use, its broad applicability across patients including those in whom TD measurements may be deemed suspect, and the plausibility of the eFick calculation based on its similarity to the gold standard method, direct Fick. Relative to eFick, the CO models developed here present favorable tradeoffs. The accuracy of the CORE model is far superior to eFick in the ICU and it both classifies and tracks CO better at the cost of a modestly more involved calculation and the need for additional predictors. Fortunately, the extra computation can easily be handled by any modern desktop or handheld computer (eg, smartphone) and the additional predictors are laboratory measurements and vital signs routinely collected during clinical care. Moreover, modern electronic medical record systems could readily automate such calculations.

A long sought goal of CO assessment is a tool that requires less invasive vascular access than TD.[34-36] Even if the tradeoff of such a tool were a modest loss of accuracy, the benefit of enabling CO estimation across many more clinical contexts may justify it. Highlighting such a need is the common clinical practice of attempting to gauge CO from an isolated measurement of $ScvO_2$ in patients with CVC access. To this end we showed that CO estimation can be performed with widely available clinical predictors obtained from a CVC or even noninvasively. Though the performance of these CO prediction models slips

compared to the CORE model with PAC predictors, they still have important advantages over eFick (**Table 4**, **Figure 2**). Indeed, CO estimation with noninvasive predictors was even more accurate in CICUs than CO estimation by eFick with $SmvO_2$ measured through a PAC. Future studies will be needed to determine whether the performance of the CORE model with ScvO2 input and the noninvasive models hold up in non-cardiac ICUs or even hospital floors.

The CO models developed here also compare favorably against several technologies developed as alternatives to TD. These TD competitors include technologies such as pulse-contour analysis, esophageal Doppler, transthoracic electrical bioimpedance, or carbon dioxide rebreathing, technologies whose tradeoffs are often judged by 2 classes of criteria.[3] The first include performance criteria such as precision, accuracy, and the capacity to influence clinical decision-making. A meta-analysis of the 4 techniques mentioned above quantified the ranges of their bias (0-0.77 L/min) and precision (1.07-1.22 L/min) relative to TD in patients recovering from cardiac surgery. By comparison, the CORE model's bias (0.06 L/min) and precision (0.95 L/min) in the CSICU were quite competitive (Supplemental Table 9). The second class of criteria encompasses several factors, including cost, safety, availability, ease of use, applicability across patient populations, and validation across clinical contexts. Relative to these criteria the CORE model is competitive as well. In terms of cost, safety, and availability it is arguably the nearest neighbor to TD, the standard of care. Furthermore, it is easier to use than TD and applies to a broader patient population. Finally, we have validated its performance in a far larger patient data set than many alternative technologies, one that encompasses a more heterogeneous spectrum of cardiovascular illness and especially critical care.

STUDY LIMITATIONS. The CO models developed here excel compared to eFick, but with R^2 metrics around 0.6, there remains variation in TD-CO that they do not yet capture. A component of this variation may be due to intrinsic noise in the TD measurement used as a training label, but we believe there is likely also physiology not accounted for by the set of 13 predictors we used. We chose predictors that are readily available, favoring ease of use and reduced model complexity, but there are many other predictors that could be incorporated (such as the myriad metrics of mechanical ventilation). There are also promising data sources yet to be exploited. The success of arterial waveform analysis in CO estimation, and the newly discovered wealth of information embedded in the electrocardiogram, point to at least 2 additional predictors that may further improve CO models.

The generalizability of these novel CO models may be compromised by the limitations of real-world data sets and retrospective analyses. For example, imputation was required here, particularly for measurements such as body temperature in CCL patients. Additionally, time alignment of measurements was achieved by defining simultaneity in terms of time windows. These time windows compromised the granularity of the dynamics captured by the models. Despite not fully extracting all available data and striking compromises entailed by real-world data, our models nevertheless demonstrated key performance advantages over standard practice. Importantly, model generalization was supported by the large diverse data sets, drawn from 2 institutions and 2 clinical settings and by their very similar performance on test sets and external validation data sets. Finally, the data sets used to fit the models are biased towards patients with cardiac illness. Model generalization will need to be further tested on diverse patients such as those treated in medical ICUs with septic shock or advanced liver disease.

CONCLUSIONS

All CO assessment methods face tradeoffs, especially in the ICU where dynamic patient physiology spans a wide spectrum, vascular access varies, expertise varies, and resources vary. The estimated Fick models were not designed for the ICU context, and though their limitations have long been recognized, no alternative estimation method has yet replaced them. As shown here, the wealth of data routinely collected on ICU patients combined with modern modeling tools enable superior CO estimators. These estimators have advantages over TD as well—they are easier to use, apply to a broader spectrum of patients, and are compatible with a wider range of vascular access, including absence of a PAC. Their deployment would require no additional devices, no new measurements outside of standard clinical practice, and no specialized training. Their adoption will ultimately hinge on further studies to confirm their accuracy and to prove their utility. Confidence in their accuracy will entail showing that their performance generalizes to other institutions, to common patient subsets, and in particular to medical ICUs. Confidence in their utility will be gained by showing that common electronic medical record systems can support their computation, that physicians perceive them as easy to use, and that they fill existing gaps such as patients

without a PAC. CO estimators that are more accurate than eFick, more versatile than TD, and tailored to the ICU, could become valuable tools in the management of critical illness.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

Dr Houstis was supported by the National Institutes of Health (R01HL173129). Dr Aguirre was supported by the Henry M Jackson Foundation for the Advancement of Military Medicine and by the National Institutes of Health (R01HL175344). Dr Aguirre also discloses grant funding for work unrelated to this study from Amgen Inc and Philips Healthcare. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Nicholas Houstis, Division of Cardiology, Massachusetts General Hosptial and Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114, USA. E-mail: nhoustis@mgh.harvard.edu. OR Dr Aaron Aguirre, 55 Fruit Street, Thier 212, Boston, Massachusetts 02114, USA. E-mail: aguirre.aaron@mgh.harvard.edu.

REFERENCES

1. Guyton AC. Regulation of cardiac output. *Anesthesiology*. 1968;29:314-326. https://doi.org/10.1097/00000542-196803000-00016

2. Kobe J, Mishra N, Arya VK, Al-Moustadi W, Nates W, Kumar B. Cardiac output monitoring: technology and choice. *Ann Card Anaesth*. 2019;22:6-17. https://doi.org/10.4103/aca.ACA_41_18

3. Peyton PJ, Chong SW. Minimally invasive measurement of cardiac output during surgery and critical care: a meta-analysis of accuracy and precision. *Anesthesiology*. 2010;113:1220-1235. https://doi.org/10.1097/ALN.0b013e3181ee3130

4. Reuter DA, Huang C, Edrich T, Shernan SK, Eltzschig HK. Cardiac output monitoring using indicator-dilution techniques: basics, limits, and perspectives. *Anesth Analg*. 2010;110:799-811. https://doi.org/10.1213/ANE.0b013e3181cc885a

5. Hoeper MM, Maier R, Tongers J, et al. Determination of cardiac output by the Fick method, thermodilution, and acetylene rebreathing in pulmonary hypertension. *Am J Respir Crit Care Med*. 1999;160:535-541. https://doi.org/10.1164/ajrccm.160.2.9811062

6. Khirfan G, Ahmed MK, Almaaitah S, et al. Comparison of different methods to estimate cardiac index in pulmonary arterial hypertension. *Circulation*. 2019;140:705-707. https://doi.org/10.1161/CIRCULATIONAHA.119.041614

7. Chiara O, Giomarelli PP, Biagioli B, Rosi R, Gattinoni L. Hypermetabolic response after hypothermic cardiopulmonary bypass. *Crit Care Med*. 1987;15:995-1000. https://doi.org/10.1097/00003246-198711000-00001

8. Kohanna FH, Cunningham JN, Catinella FP, Adams PX, Nathan IM, Pasternack BS. Cardiac output determination after cardiac operation. *J Thorac Cardiovasc Surg*. 1981;82:904-908. https://doi.org/10.1016/s0022-5223(19)39242-6

9. Li J, Schulze-Neick I, Lincoln C, et al. Oxygen consumption after cardiopulmonary bypass surgery in children: determinants and implications. *J Thorac Cardiovasc Surg*. 2000;119:525-533. https://doi.org/10.1016/s0022-5223(00)70132-2

10. Oudemans-van Straaten HM, Jansen PG, te Velthuis H, et al. Increased oxygen consumption after cardiac surgery is associated with the inflammatory response to endotoxemia. *Intensive Care Med*. 1996;22:294-300. https://doi.org/10.1007/BF01700449

11. Bergstra A, van Dijk RB, Hillege HL, Lie KI, Mook GA. Assumed oxygen consumption based on calculation from dye dilution cardiac output: an improved formula. *Eur Heart J*. 1995;16:698-703. https://doi.org/10.1093/oxfordjournals.eurheartj.a060976

12. Dehmer GJ, Firth BG, Hillis LD. Oxygen consumption in adult patients during cardiac catheterization. *Clin Cardiol*. 1982;5:436-440. https://doi.org/10.1002/clc.4960050803

13. LaFarge CG, Miettinen OS. The estimation of oxygen consumption. *Cardiovasc Res*. 1970;4:23-30. https://doi.org/10.1093/cvr/4.1.23

14. Zoghbi WA, Adams D, Bonow RO, et al. Recommendations for noninvasive evaluation of native valvular regurgitation: a report from the American society of echocardiography developed in collaboration with the society for cardiovascular magnetic resonance. *J Am Soc Echocardiogr*. 2017;30:303-371. https://doi.org/10.1016/j.echo.2017.01.007

15. Gutierrez G. Central and mixed venous O(2) saturation. *Turk J Anaesthesiol Reanim*. 2020;48:2-10. https://doi.org/10.5152/TJAR.2019.140

16. Kendrick AH, West J, Papouchado M, Rozkovec A. Direct Fick cardiac output: are assumed values of oxygen consumption acceptable? *Eur Heart J*. 1988;9:337-342. https://doi.org/10.1093/oxfordjournals.eurheartj.a062505

17. Li J. Accurate measurement of oxygen consumption in children undergoing cardiac catheterization. *Catheter Cardiovasc Interv*. 2013;81:125-132. https://doi.org/10.1002/ccd.24440

18. Li J, Bush A, Schulze-Neick I, Penny DJ, Redington AN, Shekerdemian LS. Measured versus estimated oxygen consumption in ventilated patients with congenital heart disease: the validity of predictive equations. *Crit Care Med*. 2003;31:1235-1240. https://doi.org/10.1097/01.CCM.0000060010.81321.45

19. Narang N, Gore MO, Snell PG, et al. Accuracy of estimating resting oxygen uptake and implications for hemodynamic assessment. *Am J Cardiol*. 2012;109:594-598. https://doi.org/10.1016/j.amjcard.2011.10.010

20. Narang N, Thibodeau JT, Levine BD, et al. Inaccuracy of estimated resting oxygen uptake in the clinical setting. *Circulation*. 2014;129:203-210. https://doi.org/10.1161/CIRCULATIONAHA.113.003334

21. Tonelli AR, Wang XF, Abbay A, Zhang Q, Ramos J, McCarthy K. Can we better estimate resting oxygen consumption by incorporating arterial blood gases and spirometric determinations? *Respir Care*. 2015;60:517-525. https://doi.org/10.4187/respcare.03555

22. Wolf A, Pollman MJ, Trindade PT, Fowler MB, Alderman EL. Use of assumed versus measured oxygen consumption for the determination of cardiac output using the Fick principle. *Catheter Cardiovasc Diagn*. 1998;43:372-380. https://doi.org/10.1002/(sici)1097-0304(199804)43:4<372::aid-ccd3>3.0.co;2-5

23. van Grondelle A, Ditchey RV, Groves BM, Wagner WW Jr, Reeves JT. Thermodilution method overestimates low cardiac output in humans. *Am J Physiol*. 1983;245:H690-H692. https://doi.org/10.1152/ajpheart.1983.245.4.H690

24. Cigarroa RG, Lange RA, Williams RH, Bedotto JB, Hillis LD. Underestimation of cardiac output by thermodilution in patients with tricuspid regurgitation. *Am J Med*. 1989;86:417-420. https://doi.org/10.1016/0002-9343(89)90339-2

25. Hamilton MA, Stevenson LW, Woo M, Child JS, Tillisch JH. Effect of tricuspid regurgitation on the reliability of the thermodilution cardiac output technique in congestive heart failure. *Am J Cardiol*. 1989;64:945-948. https://doi.org/10.1016/0002-9149(89)90851-5

26. Boerboom LE, Kinney TE, Olinger GN, Hoffmann RG. Validity of cardiac output measurement by the thermodilution method in the presence of acute tricuspid regurgitation. *J Thorac Cardiovasc Surg*. 1993;106:636-642.

27. Balik M, Pachl J, Hendl J. Effect of the degree of tricuspid regurgitation on cardiac output measurements by thermodilution. *Intensive Care Med*. 2002;28:1117-1121. https://doi.org/10.1007/s00134-002-1352-0

28. Fares WH, Blanchard SK, Stouffer GA, et al. Thermodilution and Fick cardiac outputs differ: impact on pulmonary hypertension evaluation.

Can Respir J. 2012;19:261–266. https://doi.org/10.1155/2012/261793

29. Hillis LD, Firth BG, Winniford MD. Analysis of factors affecting the variability of Fick versus indicator dilution measurements of cardiac output. *Am J Cardiol*. 1985;56:764–768. https://doi.org/10.1016/0002-9149(85)91132-4

30. Alkhodair A, Tsang MYC, Cairns JA, et al. Comparison of thermodilution and indirect Fick cardiac outputs in pulmonary hypertension. *Int J Cardiol*. 2018;258:228–231. https://doi.org/10.1016/j.ijcard.2018.01.076

31. Chase PJ, Davis PG, Wideman L, Starnes JW, Schulz MR, Bensimhon DR. Comparison of estimations versus measured oxygen consumption at rest in patients with heart failure and reduced ejection fraction who underwent right-sided heart catheterization. *Am J Cardiol*. 2015;116:1724–1730. https://doi.org/10.1016/j.amjcard.2015.08.051

32. Kresoja KP, Faragli A, Abawi D, et al. Thermodilution vs estimated Fick cardiac output measurement in an elderly cohort of patients: a single-centre experience. *PLoS One*. 2019;14:e0226561. https://doi.org/10.1371/journal.pone.0226561

33. Opotowsky AR, Hess E, Maron BA, et al. Thermodilution vs estimated Fick cardiac output measurement in clinical practice: an analysis of mortality from the veterans affairs clinical assessment, reporting, and tracking (VA CART) program and vanderbilt university. *JAMA Cardiol*. 2017;2:1090–1099. https://doi.org/10.1001/jamacardio.2017.2945

34. Bataille B, Bertuit M, Mora M, et al. Comparison of esCCO and transthoracic echocardiography for non-invasive measurement of cardiac output intensive care. *Br J Anaesth*. 2012;109:879–886. https://doi.org/10.1093/bja/aes298

35. Eyeington CT, Ancona P, Cioccari L, et al. Non-invasive estimation of cardiac index in healthy volunteers. *Anaesth Intensive Care*. 2018;46:290–296. https://doi.org/10.1177/0310057X1804600306

36. Greiwe G, Peters V, Hapfelmeier A, Romagnoli S, Kubik M, Saugel B. Cardiac output estimation by multi-beat analysis of the radial arterial blood pressure waveform versus intermittent pulmonary artery thermodilution: a method comparison study in patients treated in the intensive care unit after off-pump coronary artery bypass surgery. *J Clin Monit Comput*. 2020;34:643–648. https://doi.org/10.1007/s10877-019-00374-0

KEY WORDS **cardiac output, critical care, data science, machine learning, physiology**

APPENDIX **For supplemental methods, tables, and figures, please see the online version of this paper.**