Marinho *et al. Critical Care* (2026) 30:310
https://doi.org/10.1186/s13054-026-05938-6

Critical Care

CORRESPONDENCE Open Access

Evaluating possible treatment effect heterogeneity in a randomized trial of milrinone versus dobutamine in cardiogenic shock

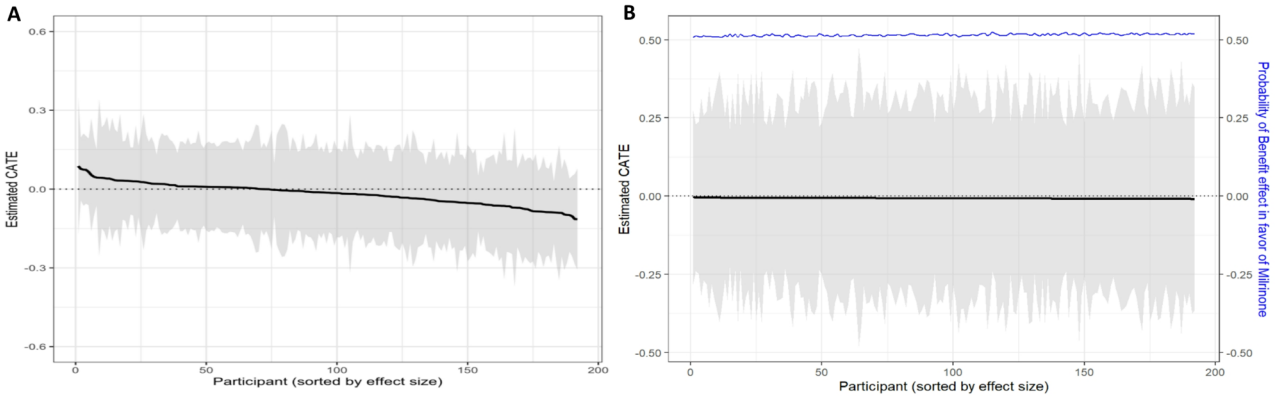
Lucas L. Marinho[1], Gabriele Lileikyte[2], Philippe Boileau[1,3], Ying Lu[1], Matthew Churpek[4], Alexandra Spicer[4], Michael Harhay[5], Zhe Chen[5], Adriana Luk[6], Fernando G. Zampieri[7], Caio A. M. Tavares[8,9], Alastair Proudfoot[10], Cristina Pop[1], Patrick J. F. Deschênes[1], Ryan Zarychanzki[11], Sylvain Lother[11], Ludhmila Abrahão Hajjar[12], Jacob C. Jentzer[13], Sean van Diepen[14], Pietro di Santo[13], Rebecca Mathew[15] and Patrick R. Lawler[1,6*]

Optimal inotrope selection in cardiogenic shock remains uncertain. Milrinone may be preferred in patients with pulmonary hypertension, right ventricular (RV) dysfunction, antecedent β-blocker therapy, tachycardia, or β1-agonist vasopressor receipt, whereas dobutamine may be preferred for rapid titration or in renal dysfunction.

In the *Milrinone as Compared with Dobutamine in the Treatment of Cardiogenic Shock (DOREMI)* randomized controlled trial (RCT) [1], outcomes did not differ between inotrope groups. Neutral average treatment effects may obscure clinically relevant heterogeneity of treatment effect (HTE) in the biologically and clinically heterogeneous syndrome of cardiogenic shock. Although prior single-variable subgroup analyses in DOREMI did not identify effect modification, these methods are constrained by prespecified hypotheses, uncertain cut-points, and limited power [2]. We hypothesized that effect-based HTE methods may uncover between-patient variation in individualized treatment effects (ITE) – the estimated treatment effect for each DOREMI participant.

The DOREMI RCT (NCT03207165) randomized 192 adults with cardiogenic shock to milrinone or dobutamine [1]. In this *post hoc* analysis, we used effect-based causal machine learning models to evaluate whether

*Correspondence:
Patrick R. Lawler
patrick.Lawler@mcgill.ca
[1]McGill University Health Centre, McGill University, 1001 boulevard Décarie, Montreal H4A 3J1, Canada
[2]Department of Clinical Sciences Helsingborg, Anaesthesia and Intensive Care, Lund University, Skåne University Hospital, Malmö, Sweden
[3]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada
[4]Department of Medicine, University of Wisconsin-Madison, Madison, USA
[5]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[6]University of Toronto, Toronto, Canada
[7]Department of Critical Care Medicine, University of Alberta, Edmonton, AB, Canada
[8]Academic Research Organization, Hospital Israelita Albert Einstein, São Paulo, Brazil
[9]Geriatric Cardiology Unit, Faculdade de Medicina, Instituto do Coracao (InCor), Hospital das Clínicas HCFMUSP, Universidade de São Paulo, São Paulo, Brazil
[10]Department of Perioperative Medicine, Barts Heart Centre, Barts Health NHS Trust, London, England; Critical Care & Perioperative Medicine Research Group, Queen Mary University, London, England
[11]Department of Internal Medicine, Max Rady College of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada
[12]Emergency and Intensive Care, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil
[13]Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN, USA
[14]Department of Critical Care Medicine, Divisions of Cardiology, Department of Medicine, University of Alberta, Edmonton, Canada
[15]CAPITAL Research Lab, Division of Cardiology, University of Ottawa Heart Institute, Ottawa, ON, Canada



Panel A: Individual conditional average treatment effects (CATEs) were estimated using a 10-fold nested cross-validation causal forest, with each patient's estimate derived exclusively from a held-out fold. Patients are ranked on the x-axis by predicted CATE, from greatest expected benefit with dobutamine (left) to greatest expected benefit with milrinone (right). The black line shows the point estimate for each individual; the gray band denotes pointwise 95% confidence intervals from model-based variance. The dashed horizontal line at y=0 indicates no treatment difference. The y-axis represents the individual ITE estimates, expressed as the absolute risk difference for the composite endpoint (milrinone – dobutamine; negative values favor milrinone) with 95% CIs. **Panel B:** Individual conditional average treatment effects (CATEs) were estimated using a Bayesian Additive Regression Trees (BART) model trained on the pooled cohort of patients treated with milrinone or dobutamine. Patients are ranked along the x-axis by predicted CATE, from greatest expected benefit with dobutamine (left) to greatest benefit with milrinone (right). The black line represents the posterior mean CATE for each individual, with gray shading indicating 95% credible intervals. The blue line shows the posterior probability that milrinone is more beneficial (Pr[CATE < 0]). The dashed horizontal line at y = 0 denotes no treatment difference. The left y-axis displays the estimated CATE (absolute risk difference for the composite endpoint), while the right y-axis shows the posterior probability of benefit favoring milrinone. *Variables considered in the models included age, sex, race, body mass index, left ventricular ejection fraction, heart rate, Vasoactive-Inotropic Score, concomitant vasopressor use, serum creatinine, time to randomization, atrial fibrillation, chronic obstructive pulmonary disease, right ventricular dysfunction, and Society for Cardiovascular Angiography and Interventions shock classification. Cardiac output and intracardiac hemodynamics could not be included owing to absent collection or prohibitive rates of missingness (>90%).

Fig. 1 Heterogeneity of Treatment Effect Evaluation*. **Panel A**: Effect-Based Causal Forest Model. **Panel B**: Effect-Based BART Model

milrinone (versus dobutamine) improved the in-hospital composite of all-cause mortality, cardiac arrest, mechanical circulatory support or transplantation, myocardial infarction, cerebrovascular accident, or new renal replacement therapy (RCT primary endpoint).

We selected baseline variables *a priori* based on clinical reasoning and usual care practices influencing inotrope selection (age, sex, race, body mass index, left ventricular ejection fraction, heart rate, Vasoactive-Inotropic Score, concomitant vasopressor use, serum creatinine, time to randomization, atrial fibrillation, chronic obstructive pulmonary disease, RV dysfunction, and Society for Cardiovascular Angiography and Interventions shock classification). Noninvasive estimates of cardiac output were not collected, and only 23/192 (12.0%) participants had a pulmonary-artery catheter at randomization, precluding inclusion of these parameters.

To assess HTE, we used machine-learning methods to estimate an ITE for each trial participant and evaluated variation across the cohort. ITEs were generated by predicting each participant's outcome using baseline variables, and, in addition, variables reflecting randomization to milrinone or dobutamine (Supplemental Methods). First, we applied causal forests – a prediction model that estimates ITE by maximizing a variable reflecting separation in predicted outcomes with milrinone compared to dobutamine [3]. Second, we applied Bayesian Additive Regression Trees (BART) [4], a tree-based predictive model that estimates ITEs using models with baseline and treatment variables [4].

All 192 DOREMI participants were included. Baseline characteristics and overall RCT outcomes have been reported [1]. Causal forests suggested limited variability in estimated ITEs. Although risk differences ranged from + 0.09 to – 0.11, confidence intervals (CIs) were wide and included unity for all patients, with most values clustered near zero (Fig. 1A). Likewise, no meaningful differences in treatment effect were observed with BART; ITEs estimates were centered around zero with wide credible intervals that included unity for all patients (Fig. 1B). The findings were similar for in-hospital mortality (76 events). To explore prior observations related to mean arterial pressure (MAP) [5], we performed a sensitivity analysis adding baseline MAP and β-blocker use, which showed no HTE. To check that no effect modifiers had been overlooked, we applied a machine learning statistical screening procedure for effect modifiers (Supplemental Methods) – this identified none (eTable1).

Although no HTE was identified, wide confidence and credible intervals suggested low confidence in estimated ITEs, which widely encompassed both possible superiority or inferiority of milrinone in all participants – likely owing to modest sample size. To understand whether sample size may have also limited prior single-variable subgroup analyses [1], we performed simulation studies to evaluate statistical power to detect HTE

using conventional logistic regression (Supplemental Methods). Simulations indicated that approximately 1,152 patients would be required to achieve 90% power to detect a single effect modifier associated with a 10% absolute risk difference (eTable2).

Overall, in this analysis of DOREMI, no meaningful between-patient variation in the effects of milrinone compared with dobutamine was observed. However, given wide confidence and credible intervals, as well as simulations contextualizing prior conventional subgroup analyses, HTE cannot be excluded, and definitive conclusions cannot be drawn.

These findings highlight a methodological challenge. As HTE analyses are increasingly applied in critical care, limited sample sizes may constrain their ability to detect clinically meaningful HTE. This was also suggested for single-variable subgroup analyses. More explicit consideration of sample size requirements and power for HTE detection is needed to support their reliable implementation.

Limitations included lack of available invasive and non-invasive hemodynamic measures, modest sample size, single cohort, and potentially the selection of candidate effect modifiers based on clinical practice. Parlow et al. [5] previously observed a trend towards possible differential outcomes by treatment based on post-randomization MAP (over 36 h), although not for baseline MAP. Adding baseline MAP to our models did not identify HTE either. Further studies may consider MAP and comprehensively record hemodynamics to facilitate HTE exploration. Larger studies are needed to determine whether personalized inotrope selection improves outcomes in cardiogenic shock.

Abbreviations

BART	Bayesian Additive Regression Trees
CATE	Conditional Average Treatment Effect
CI	Confidence Interval
DOREMI	Milrinone as Compared with Dobutamine in the Treatment of Cardiogenic Shock trial
HTE	Heterogeneous Treatment Effect
ITE	Individualized Treatment Effect
RV	Right Ventricle

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13054-026-05938-6.

Supplementary Material 1.

Acknowledgements
PRL is supported by a career award from the Fonds de Recherche du Quebec. MOH is supported by United States National Institutes of Health grant number R01-HL168202. The authors thank Dr. Virginie Beausejour Ladouceur for her support conceptualizing and reporting this work.

Author contributions
LM and YL performed the statistical analysis. LM and PRL were the major contributors to writing the manuscript. The study was designed by LM, PRL, SV, RM, GL, and PD. RM and PD performed the original trial and collected relevant data. All authors critically appraised the data. All authors read and approved the final manuscript.

Funding
None.

Data availability
The datasets used and/or analyzed during the current study are available from the authors on reasonable request.

Declarations

Ethics approval and consent to participate
The trial was approved by the Ottawa Health Science Network Research Ethics Board (approval number 20160975–01 H).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 19 January 2026 / Accepted: 28 February 2026
Published online: 21 May 2026

References

1. Mathew R, Di Santo P, Jung RG, Marbach JA, Hutson J, Simard T, Ramirez FD, Harnett DT, Merdad A, Almufleh A, et al. Milrinone as Compared with Dobutamine in the Treatment of Cardiogenic Shock. N Engl J Med. 2021;385(6):516–25.
2. Munroe ES, Spicer A, Castellvi-Font A, Zalucky A, Dianti J, Graham Linck E, Talisa V, Urner M, Angus DC, Baedorf-Kassis E, et al. Evidence-based personalised medicine in critical care: a framework for quantifying and applying individualised treatment effects in patients who are critically ill. Lancet Respir Med. 2025;13(6):556–68.
3. Wager S, Athey S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. J Am Stat Assoc. 2018;113(523):1228–42.
4. Chipman HA, George EI, McCulloch RE. Bart: bayesian additive regression trees. Annals Appl Stat. 2010;4:266–98.
5. Parlow S, Di Santo P, Mathew R, Jung RG, Simard T, Gillmore T, Mao B, Abdel-Razek O, Ramirez FD, Marbach JA, et al. The association between mean arterial pressure and outcomes in patients with cardiogenic shock: insights from the DOREMI trial. Eur Heart J Acute Cardiovasc Care. 2021;10(7):712–20.

Publisher's note