Evidence-based personalised medicine in critical care: a framework for quantifying and applying individualised treatment effects in patients who are critically ill

Elizabeth S Munroe, Alexandra Spicer*, Andrea Castellvi-Font*, Ann Zalucky*, Jose Dianti, Emma Graham Linck, Victor Talisa, Martin Urner, Derek C Angus, Elias Baedorf-Kassis, Bryan Blette, Lieuwe D Bos, Kevin G Buell, Jonathan D Casey, Carolyn S Calfee, Lorenzo Del Sorbo, Elisa Estenssoro, Niall D Ferguson, Rachel Giblon, Anders Granholm, Michael O Harhay, Anna Heath, Carol Hodgson, Timothy Houle, Cong Jiang, Lina Kramer, Patrick R Lawler, Aleksandra Leligdowicz, Fan Li, Kuan Liu, Amelia Maiga, David Maslove, Colin McArthur, Daniel F McAuley, Ary Serpa Neto, Charissa Oosthuysen, Anders Perner, Hallie C Prescott, Bram Rochwerg, Sarina Sahetya, Mariia Samoilenko, Mireille E Schnitzer, Kevin P Seitz, Faraaz Shah, Manu Shankar-Hari, Pratik Sinha, Arthur S Slutsky, Edward T Qian, Steve A Webb, Paul J Young, Fernando G Zampieri, Ryan Zarychanski, Eddy Fan, Matthew W Semler, Matthew Churpek†, Ewan C Goligher†, for the Platform of Randomized Adaptive Clinical Trials in Critical Illness (PRACTICAL) investigators and the Evidence-based Individualized Treatment Effects (EvITE) Group*

Clinicians aim to provide treatments that will result in the best outcome for each patient. Ideally, treatment decisions are based on evidence from randomised clinical trials. Randomised trials conventionally report an aggregated difference in outcomes between patients in each group, known as an average treatment effect. However, the actual effect of treatment on outcomes (treatment response) can vary considerably between individuals, and can differ substantially from the average treatment effect. This variation in response to treatment between patients—heterogeneity of treatment effect—is particularly important in critical care because common critical care syndromes (eg, sepsis and acute respiratory distress syndrome) are clinically and biologically heterogeneous. Statistical approaches have been developed to analyse heterogeneity of treatment effect and predict individualised treatment effects for each patient. In this Review, we outline a framework for deriving and validating individualised treatment effects and identify challenges to applying individualised treatment effect estimates to inform treatment decisions in clinical care.

Lancet Respir Med 2025

Published Online
April 15, 2025
https://doi.org/10.1016/S2213-2600(25)00054-2

*Co-first authors

†Co-senior authors

Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA (E S Munroe MD, H C Prescott MD); **Division of Pulmonary and Critical Care, Department of Medicine, University of Wisconsin–Madison, Madison, WI, USA** (A Spicer MS, M Churpek MD); **Department of Critical Care, Hospital del Mar, and Critical Illness Research Group (GREPAC), Hospital del Mar Research Institute (IMIM), Barcelona, Spain** (A Castellvi-Font MD); **Division of Respirology, Department of Medicine, University Health Network, Toronto, ON, Canada** (A Castellvi-Font, J Dianti MD, L Del Sorbo MD, N D Ferguson MD, C Oosthuysen PhD, E Fan MD, E C Goligher MD); **Toronto General Hospital Research Institute, Toronto, ON, Canada** (A Castellvi-Font, M Urner MD, Prof N D Ferguson, Prof E Fan, E C Goligher); **Department of Critical Care Medicine, Snyder Institute for Chronic Diseases, Cumming School of Medicine, University of Calgary and Alberta Health Services, Foothills Medical Center, Calgary, AB, Canada** (A Zalucky MD); **Division of Pulmonary, Critical Care, Allergy and Sleep Medicine, University of California San Francisco, San Francisco,**

Introduction

Fundamentally, medicine aims to provide treatments that will result in the best outcome for each patient. This core concept motivates personalised medicine. Personalisation requires an understanding of how treatments affect outcomes for individual patients. This requirement raises an epistemological problem: the true effect of treatment on outcomes for individual patients is unmeasurable because an individual cannot be observed under both scenarios of treatment received and treatment not received (counterfactual conditions).[1] Instead, randomised controlled trials (RCTs) typically combine the treatment effects for all patients into an average treatment effect (ATE), which represents the aggregated

Key messages

- The average treatment effect reported in a clinical trial does not necessarily represent the treatment effect for individual patients in the trial. Treatment effects in individual patients can be larger, smaller, or in the opposite direction of the average treatment effect. This variation in response to treatment between patients is called heterogeneity of treatment effect (HTE).
- Several approaches exist for analysing HTE in randomised trials. The most basic approach, conventional subgroup analysis, assesses whether a single patient characteristic alters the effect of treatment on an outcome; this approach has important limitations. Data-derived subgroups (eg, subphenotypes) can incorporate multiple patient characteristics to assess treatment effect, although clinically important HTE could still exist within each group. Risk-based and effect-based models can incorporate multiple patient characteristics and examine HTE across a spectrum of predicted risk or predicted treatment effect.
- By accounting for complex interactions between multiple variables that might modify treatment effect, effect-based models can provide predictions of the individualised treatment effect (ITE) for each patient. Such models are at risk of statistical overfitting and spurious detection of HTE. These risks can be mitigated by robust model derivation and validation, as outlined in this Review.
- Clinical trials can use different approaches to prospectively address potential HTE. Novel trial designs might establish eligibility or stratify randomisation and analysis on the basis of prespecified hypothesised determinants of HTE, use information accrued about HTE during a trial to adapt eligibility criteria and randomisation in real-time, or prespecify plans to estimate ITEs after trial completion if substantial HTE is anticipated.
- There are several challenges in applying predicted ITEs to inform treatment decisions in clinical care, including the lack of a consensus framework for prospectively validating predicted ITE, barriers to data availability and timing, and challenges with regulation and implementation of ITE models in clinical care.

CA, USA (A Zalucky, Prof C S Calfee MD); **Interdepartmental Division of Critical Care Medicine** (J Dianti, M Urner, Prof N D Ferguson, P R Lawler MD, Prof A S Slutsky MD, Prof E Fan, E C Goligher) **and Department of Anesthesiology & Pain Medicine** (M Urner), **University of Toronto, Toronto, ON, Canada; School of Medicine and Public Health, University of Wisconsin–Madison, Madison, WI, USA** (E Graham Linck MS, M Churpek); **Center for Reasearch, Investigation, and Systems Modeling of Acute Illness, Department of Critical Care Medicine, University of Pittsburgh, PA, USA** (V Talisa PhD, Prof D C Angus MD); **Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine, University of Pittsburgh, Pittsburgh, PA, USA** (F Shah MD); **Department of Medicine, Harvard Medical School, Boston, MA, USA** (E Baedorf-Kassis MD); **Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA** (B Blette PhD); **Department of Intensive Care and Laboratory of Experimental Intensive Care and Anesthesiology, Amsterdam, Netherlands** (L D Bos MD, L Kramer MS); **Division of Pulmonary and Critical Care, Department of Medicine, University of Chicago, Chicago, IL, USA** (K G Buell MBBS); **Division of Allergy, Pulmonary, and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA** (J D Casey MD, K P Seitz MD, E T Qian MD, M W Semler MD); **Hospital Interzonal San Martin de La Plata, Buenos Aires, Argentina** (Prof E Estenssoro); **Division of Biostatistics** (R Giblon MS, A Heath PhD) **and Institute of Health Policy, Management, and Evaluation** (K Liu PhD), **Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada** (R Giblon MS, A Heath PhD); **Department of Intensive Care, Copenhagen University Hospital—Rigshospitalet, Copenhagen, Denmark** (A Granholm PhD, Prof A Perner PhD); **Department of Biostatistics, Epidemiology and Informatics, Perelman**

	Abbreviation	Definition	Clinical example
Average treatment effect	ATE	In a clinical trial, the aggregated between-group difference in outcomes[2]	In an RCT evaluating fluid-liberal *vs* fluid-restrictive resuscitation strategies in sepsis, the ATE would represent the average difference in outcomes (eg, mortality) between patients randomised to each strategy [2]
Heterogeneity of treatment effect	HTE	The magnitude of variation of treatment effects across a population;[2] HTE is present when patients respond differently to treatment based on their individual characteristics	In two recent RCTs comparing fluid-liberal *vs* fluid-restrictive strategies in sepsis, the ATE showed no difference in mortality between thr two strategies;[3,4] however, if there is HTE, a fluid-restrictive strategy could benefit some patients while potentially harming others
True individual treatment effect	..	The unobservable difference in potential outcomes with or without treatment for an individual patient;[2] true individual treatment effect is unmeasurable because we cannot observe the outcome for the patient under both counterfactual conditions (ie, treated and untreated)	The true individual treatment effect would represent the difference in outcome (eg, mortality) that a specific patient with sepsis would have if they received a fluid-liberal *vs* fluid-restrictive resuscitation strategy
Conditional average treatment effect	CATE	Treatment effect for a patient or groups of patients conditional on their pre-treatment characteristics;[5,6] CATEs can include treatment effects conditioned on a single characteristic (ie, conventional subgroup analysis) or they can refer to treatment effects conditioned on multiple characteristics	In trials of fluid-liberal *vs* fluid-restrictive resuscitation strategies, the CATE would be the average treatment effect of a fluid strategy (liberal *vs* restrictive) for a subpopulation of patients with a particular set of baseline characteristics
Subgroup average treatment effect	..	The average between-group difference in an outcome within a given subgroup;[1] this is a type of CATE conditioned on a single baseline characteristic	In a subgroup analysis of a recent RCT comparing fluid-liberal and fluid-restrictive strategies,[4] patients requiring respiratory support had a larger subgroup average treatment effect with fluid-restriction (eg, larger benefit) than patients not on respiratory support
Individualised treatment effect	ITE	The CATE meant to represent—as closely as possible—the treatment effect for an individual patient conditioned on multiple baseline characteristics;[6] it is often estimated using multivariable models	In trials of fluid-liberal *vs* fluid-restrictive resuscitation strategies, the ITE is the expected treatment effect of a fluid strategy (liberal *vs* restrictive) for a given patient conditioned on their baseline characteristics

RCT=randomised controlled trial.

Table 1: **Definitions of terms**

between-group difference in an outcome (table 1).[1–6] These ATEs can differ from treatment effects in individual patients, which can be larger, smaller, or opposite in direction to the ATE.[7–9] This variability in response to treatment is referred to as heterogeneity of treatment effect (HTE).[2]

HTE is a particularly important consideration in critical care, a field in which treatments are often targeted on the basis of broadly defined clinical syndromes, such as sepsis and acute respiratory distress syndrome (ARDS), which encompass clinically and biologically heterogeneous populations.[10] This heterogeneity might partly explain the limited success in identifying effective treatments in critical care trials.[11] For example, two recent RCTs comparing fluid-liberal and fluid-restrictive resuscitation strategies in patients with sepsis found no difference in outcomes on average.[3,4] Similarly, multiple trials have reported no significant difference in mortality between lower versus higher positive end-expiratory pressure (PEEP) strategies in patients with ARDS.[12] However, the absence of a significant ATE for a trial population does not necessarily mean that choices about fluid resuscitation or ventilator settings do not matter for individual patients; it is possible that some patients might have benefited from the tested treatment strategies while others were harmed by them, creating an overall null effect. For example, clinical intuition suggests that an older patient with chronic heart failure who develops sepsis and respiratory failure from pneumonia could have more benefit from a restrictive fluid strategy than a young patient with sepsis due to diarrhoeal illness. A recent study found that response to protocolised resuscitation varied among patients, with patient characteristics such as albumin contributing to HTE.[13] Similarly, in ARDS, the benefit of a higher PEEP strategy can vary substantially according to a patient's potential for lung recruitment (the capacity to open atelectatic portions of the lung).[14]

If meaningful HTE is present, relying on ATE to make clinical decisions could lead to suboptimal treatment, depriving some patients of potentially beneficial treatments and exposing others to unnecessary harm. Due to this limitation, HTE has often been framed as a problem that restricts the interpretability of RCTs. Alternatively, HTE can be viewed as an opportunity to personalise treatment (figure 1); information about HTE could guide clinical

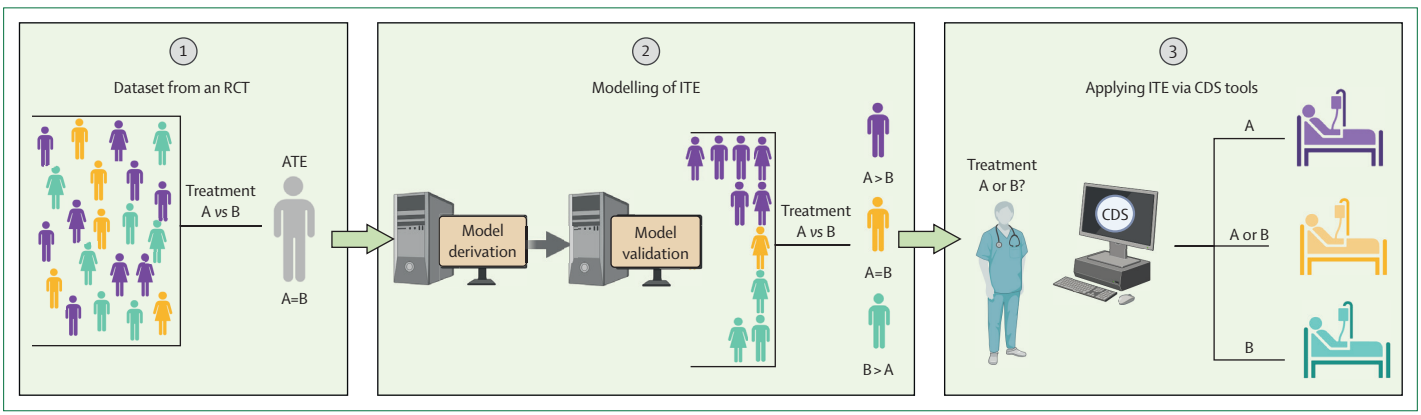

Figure 1: **Potential clinical analysis design for enhanced decision making in critical care**
This figure illustrates how understanding HTE could help to inform clinical decision making at the bedside. Panel 1 shows the ATE from an RCT comparing treatment A versus option B; this panel represents scenarios where two treatments (A and B) are compared or where a new treatment (treatment A) is compared with standard care or placebo (option B), with A and B resulting in the same outcome on average. Panel 2 outlines steps for modelling ITE from a clinical trial. In this step, which is the focus of this Review, model derivation and validation are used to identify differential effects of treatment in patients with different sets of characteristics. In this panel, treatment A works better (shown by > in the figure) than treatment B in patients with one set of characteristics (purple individuals) and vice versa for patients with a different set of characteristics (green individuals), and the treatments work similarly for patients with other characteristics (yellow individuals). A goal might be for HTE analyses to culminate in panel 3, the development of CDS tools that can be applied to enhance decision making at the bedside. However, how to go from panel 2 to panel 3 (ie, how to clinically confirm ITE-based decision tools and apply them in practice) remains a challenge in critical care. ATE=average treatment effect. HTE=heterogeneity of treatment effect. ITE=individualised treatment effect. CDS=clinical decision support. RCT=randomised controlled trial.

decisions about fluid management, ventilator settings, and other therapeutics in individual patients.

Statistical approaches for analysing HTE in RCTs involve estimation of the effect of treatment on outcomes for patients conditioned on one or more of their baseline characteristics, referred to as the conditional average treatment effect (CATE).[5] When the CATE is meant to represent—as closely as possible—the expected effect on individuals, it is called the individualised treatment effect (ITE). Multivariable models or machine learning methods can be used to predict ITEs on the basis of baseline characteristics.[6,15] For example, the ITE of a particular fluid resuscitation or PEEP strategy could be estimated for an individual patient on the basis of their demographics, comorbidities, illness severity, lung compliance, etc. In this Review, we use the term ITE to refer to an approximation of an individual effect; this ITE is distinct from a true individual treatment effect, which cannot be measured (table 1).[15] Illustrative examples of ITE in critical care trials are summarised in figure 2.[16–20]

In this Review, we describe emerging methods for the identification of HTE and prediction of ITE, review novel clinical trial designs that aim to address known or unknown sources of HTE to obtain more individualised estimates of treatment effects, discuss the challenges of applying ITE predictions to inform clinical care, and outline future directions to fully leverage the reality of HTE to advance towards care that is both evidence-based and personalised.

When does HTE matter?

Experts disagree as to when HTE matters. Some argue that estimating treatment effects for individuals is useful only when there is substantial HTE and meaningful potential downsides to treatment in terms of harm or cost.[15] For example, when an inexpensive treatment (eg, paracetamol for a headache) benefits some patients without causing relevant harm in others, the ATE could be sufficient to guide clinical decision making. Estimating ITEs could require substantial information, and the high costs of conducting a large trial to estimate ITEs might not be worthwhile if a treatment has no real downside of expense or risk. Those who approach health care at the system or population level—such as policy makers and regulators—might be more likely to adopt this perspective.

An alternative view contends that estimation of ITEs is always preferable to ATEs. In trials without a significant ATE, some patients could still derive benefit from treatment.[21] Moreover, the definitions of meaningful benefit, tolerable risk, and acceptable cost are contingent on patient values, clinical context, and system resources. Therefore, some experts argue that truly personalised decision making requires knowledge of the treatment effect for individuals. From this standpoint, clinical trials should always include HTE analyses designed to generate sufficiently precise estimates of ITEs to inform clinical decision making. This perspective might be shared by those involved in clinical decisions at the bedside, including patients, caregivers, and clinicians.

Analysing HTE in trials with a null ATE

The Predictive Approaches to Treatment Effect Heterogeneity (PATH) statement was developed by an expert panel to provide general guidance for analysing HTE in clinical trials.[15] As outlined in this statement, assessment of HTE in trials with positive results is important for understanding the distribution of effects and to identify patients most likely to benefit from treatment. The PATH statement advises against

School of Medicine, University of Pennsylvania, Philadelphia, PA, USA (M O Harhay PhD); **Department of Intensive Care, Austin Hospital, Melbourne, VIC, Australia** (A Serpa Neto PhD); **Department of Critical Care Medicine, Hospital Israelita Albert Einstein, Sao Paolo, Brazil** (A Serpa Neto); **Australian and New Zealand Intensive Care Research Centre, Monash University, Melbourne, VIC, Australia** (Prof C Hodgson PhD, A Serpa Neto, Prof S A Webb MD, P J Young MD); **Department of Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA** (T Houle PhD); **Faculté of Pharmacie, Université de Montréal, Montreal, QC, Canada** (C Jiang PhD, M Samoilenko PhD, M E Schnitzer PhD); **Division of Cardiology, Department of Medicine, McGill University Health Centre, Montreal, QC, Canada** (P R Lawler); **Division of Critial Care Medicine, Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada** (A Leligdowicz MD); **Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA** (F Li PhD); **Division of Acute Care Surgery, Department of Surgery, Vanderbilt University Medical**

Center, Nashville, TN, USA (A Maiga MD); **Department of Critical Care Medicine, Queen's University, Kingston, ON, Canada** (D Maslove MD); **Department of Critical Care Medicine, Te Toka Tumai Auckland City Hospital, Auckland, New Zealand** (C McArthur MBChB); **Wellcome–Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland** (Prof D F McAuley MD); **VA Center for Clinical Management Research, Ann Arbor, MI, USA** (H C Prescott); **Department of Medicine, McMaster University, Hamilton, ON, Canada** (B Rochwerg MD); **Division of Pulmonary and Critical Care Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD, USA** (S Sahetya MD); **Department of Social and Preventive Medicine, School of Public Health, Université de Montréal, Montreal, QC, Canada** (M E Schnitzer); **Centre for Inflammation Research, Institute for Regeneration and Repair, The University of Edinburgh, Edinburgh, UK** (Prof M Shankar-Hari PhD); **Department of Anesthesiology, Washington University School of Medicine in St Louis, St Louis, MO, USA** (P Sinha MBChB); **Intensive Care Unit, Wellington Hospital, Wellington, New Zealand** (P J Young); **Medical Research Institute of New Zealand, Wellington, New Zealand** (P J Young); **Department of Critical Care, University of Melbourne, Melbourne, VIC, Australia** (P J Young); **Department of Critical Care Medicine, Faculty of Medicine and Dentistry, University of Alberta and Alberta Health Services, Edmonton, AB, Canada** (F G Zampieri MD); **Department of Internal Medicine, University of Manitoba, Winnipeg, MB, Canada** (Prof R Zarychanski MD)

Correspondence to:
Ewan C Goligher, Toronto General Hospital Research Institute, Toronto, ON M5G 2N2, Canada
ewan.goligher@utoronto.ca

See Online for appendix

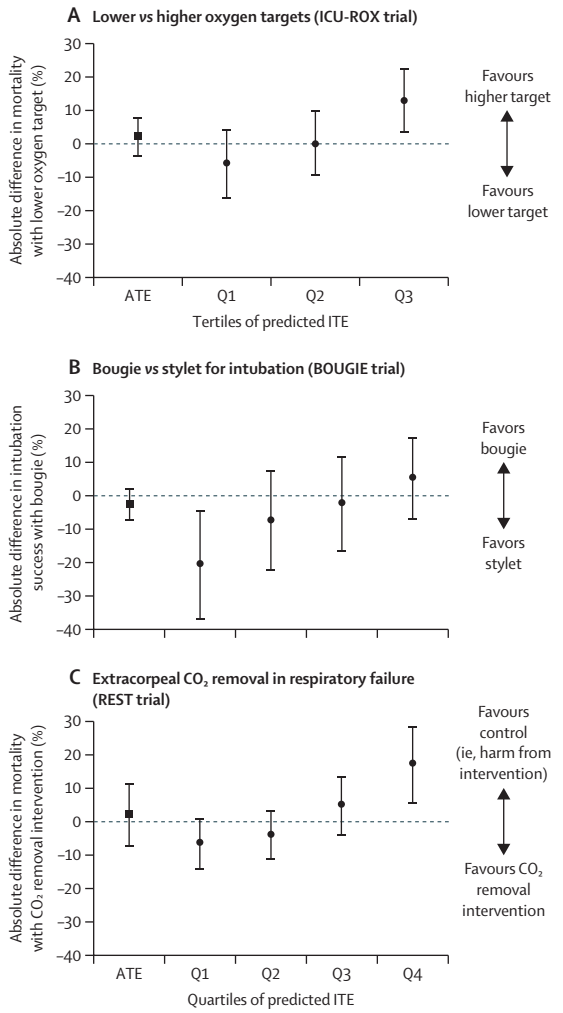

Figure 2: **Examples of individualised treatment effects in critical care trials**
This figure summarises findings from recent HTE analyses of critical care trials, and was created by compiling published results from the referenced manuscripts and presenting them in a standardised format to allow comparison (appendix p 16). The ATE from the original trial and results from the post-hoc HTE analyses are presented as absolute risk differences, for which 0 (dashed line) means there was no difference in outcomes between interventions. The arrows explain the clinical implication of absolute differences above and below 0 (eg, in panel 1, absolute differences above 0 favour higher targets and those below 0 favour lower targets). The black square represents the ATE and 95% confidence interval from the original trial. The black circles represent observed absolute difference in outcomes by quantile of predicted ITE in the HTE analyses, with 95% confidence interval in A and B and 95% credible interval in C, which was a Bayesian analysis. (A) Results from the ICU-ROX trial[16] comparing lower versus higher oxygen targets in mechanically ventilated critically ill patients and a post-hoc HTE analysis. Primary outcome: mortality. (B) Results from the BOUGIE trial[17] comparing use of a bougie versus stylet among patients who are critically ill undergoing endotracheal intubation and a post-hoc HTE analysis.[18] Primary outcome: successful intubation on first attempt. (C) Results from the REST trial[19] of protective ventilation with extracorporeal CO_2 removal in patients who are critically ill with respiratory failure and a post-hoc HTE analysis.[20] Primary outcome: mortality. ATE=average treatment effect. HTE=heterogeneity of treatment effect. ITE=individualised treatment effect.

performing HTE analyses in trials with null ATEs, given the risk of identifying spurious, clinically insignificant HTE.[15] However, the statement recognises a key exception to this rule that is commonly present in critical care trials: situations in which treatment-related harm affects the primary outcome (eg, mortality). In such cases, treatment-related mortality in some patients could offset a potential mortality benefit in others, resulting in a null ATE.[15] Many HTE analyses of critical care trials with null ATEs have found evidence of this scenario (figure 2).[21–24] Therefore, HTE analysis is probably warranted in many critical care trials, even if the ATE is null. Importantly, to maximise interpretability, HTE analyses should be prespecified at the outset of a clinical trial on the basis of a priori suspicion for HTE, and not performed selectively in trials finding no significant ATE.

Conceptual approaches to analysis of HTE

HTE analyses can have many goals, depending on the clinical or research question. Three primary goals are: to establish whether HTE exists in a population for a given treatment (ie, do different patients respond differently to this treatment?); to predict the direction and magnitude of treatment effect for individual patients (ie, given this patient's characteristics, how will they respond to this treatment?); and to generate mechanistic hypotheses based on observed HTE (ie, which patient characteristics are associated with a differential response to treatment?).

Several different methods can be used to analyse HTE, with varying ability to achieve these core goals (figure 3, table 2).[16,18,20,21,23–31] Each method uses a different solution to the challenge of grouping patients based on multiple characteristics. Importantly, these methods measure association, not causation—they do not show that observed effect modifiers (eg, age) are causally responsible for differential treatment effects. Prespecification of the approach, included variables, and rationale for expected differences in treatment effect is recommended to ensure reproducible results.[25,32]

Conventional subgroup analysis

Conventional one-variable-at-a-time subgroup analyses compare treatment effects between subgroups of patients defined by a single baseline characteristic that is hypothesised to modify treatment response (eg, age >65 years).[25] This approach is inefficient and could substantially underestimate HTE (table 2).[32,33] Additionally, patients often fit into multiple subgroups simultaneously, which makes it challenging to apply the results. For example, if harm is observed in older patients and benefit is observed in female patients, how should an older female patient be treated? When reported, conventional subgroup analyses should be appraised for rigour and quality using validated tools such as the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN).[25]

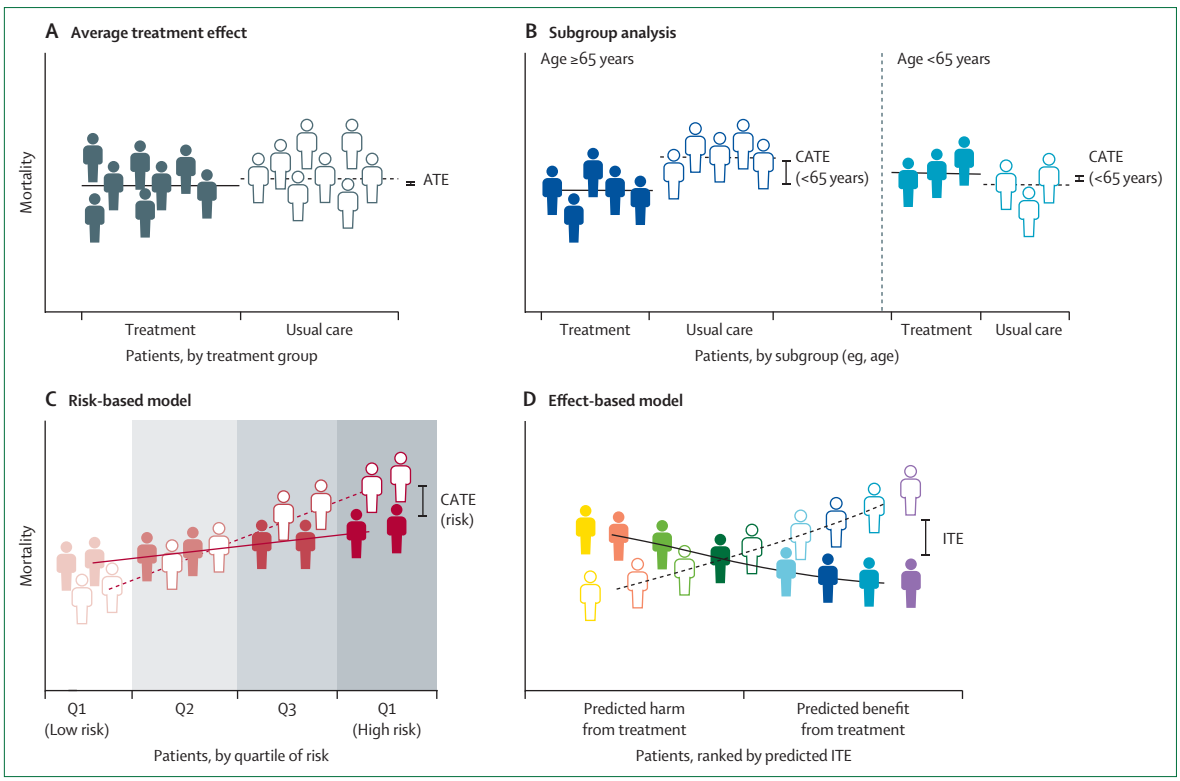

Figure 3: **Overview of approaches for analysing heterogeneity of treatment effect**
This figure presents a simplified example trial where substantial HTE is present. In this example, N=16 patients were assigned to treatment A (solid individuals; n=8) versus standard care (outlined individuals; n=8). The solid lines lines represent outcomes for those receiving treatment, and dashed lines represent outcomes for those receiving usual care. (A) ATE. This panel displays sample results for the overall trial. The ATE suggests small possible benefit. (B) Subgroup analysis. This panel represents a traditional subgroup analysis ,generating CATE on the basis of a single characteristic. (C) Risk-based models of HTE. This panel displays a risk-based model, in which patients in the trial are grouped by quartiles of baseline risk of the outcome (eg, using the Acute Physiology and Chronic Health Evaluation IV score for mortality), and the effect of treatment A is compared within each quartile, generating CATE by level of risk. (D) Effect-based models of HTE. Effect-based models use machine learning methods that use a patient's baseline characteristics, treatment group assignment, and interactions between these variables to predict the CATE for each patient on the basis of multiple characteristics, which is also known as the individualised treatment effect. This approach allows us to predict how individual patients might respond to treatment A based on multiple baseline variables. ATE=average treatment effect. CATE=conditional average treatment effect. HTE=heterogeneity of treatment effect. ITE=individualised treatment effect.

Data-derived subgroup analysis

Data-derived subgroups overcome key limitations of the conventional approach to subgroup analyses by grouping patients using multiple, rather than single, baseline characteristics. Statistical techniques—such as unsupervised clustering—are used to partition patients into groups that appear distinct on the basis of combinations of baseline characteristics, including research biomarkers.[34] Such groupings are hypothesis-generating and require replication. Once groupings are confirmed, these groups—or clusters—can be treated as prespecified subgroups in HTE analyses or used for stratification during enrolment in clinical trials.[35,36] In critical care, this approach has been used to identify subphenotypes of ARDS and sepsis that have been associated with HTE in post-hoc subgroup analyses of multiple critical care trials.[26–38]

Although promising, this approach has several limitations. First, this data-derived subgroups method estimates treatment effects for groups of patients rather than individuals, limiting the degree of personalisation. Clinically significant HTE might still exist within each subgroup. Additionally, the unsupervised clustering algorithms used to define patient clusters are derived without consideration of treatment assignment or outcome, so there is no inherent reason to anticipate that subgroups derived in this manner would show differential treatment effects. Although this reality might limit the ability to detect HTE, it might also be an advantage: identifying clusters of patients who are biologically similar, independent of treatment response, could offer insight into novel shared pathways for future research.

Risk-based analysis

Risk-based HTE analyses use prognostic models to predict a patient's baseline risk of an outcome and then estimate treatment effects for patients conditioned on baseline risk. Predicted risk can be established using either previously validated off-the-shelf models, such as Acute Physiology and Chronic Health

	Description	Limitations*	Putative effect modifier	Model building step	Model testing step	Salient examples in the literature
Subgroup analysis: single variable	Compares subgroup average treatment effects between subgroups of patients defined by a single baseline clinical or physiological characteristic	Captures a limited range of treatment effects; carries risk of false negative results (due to low power within smaller subgroups) and false positive results (due to multiple testing); patients could fit into multiple subgroups	Single baseline characteristic	None	Subgroup analysis	Most trials include subgroup analyses; the ICEMAN criteria[25] are helpful to guide interpretation of results
Subgroup analysis: data-derived	Uses unsupervised clustering techniques (eg, latent class analysis) to partition patients into groups (ie, clusters) that are maximally distinct based on baseline characteristics, biomarkers, or both; subgroup average treatment effects are then compared across these clusters of patients	Unsupervised clustering techniques are blinded to treatment and outcome and do not directly measure HTE—therefore, this approach does not necessarily capture the full range of variation in treatment effect; clustering methods could identify groupings based on noise rather than true signal, so results should ideally be replicated to confirm findings; clusters are often categorical	Subgrouping variable defined on the basis of multiple baseline characteristics (including both clinical and biomarkers)	Cluster identification, which is often performed in separate study populations or using previously identified clusters (eg, inflammatory subphenotypes)	Subgroup analysis by cluster	Calfee et al,[26] Sinha et al,[27] Famous et al,[28] Calfee et al,[29] Sinha et al[30]
Risk-based modelling	Compares treatment effect across patients conditional on predicted baseline risk of an outcome	Does not necessarily capture the full range of variation in treatment effect, as factors could modify treatment effect without modifying baseline risk of the outcome; more difficult to implement with a continuous outcome	Risk variable computed from multiple baseline characteristics	Development of a new risk model; this step is skipped if using an off-the-shelf risk model (eg, APACHE)	Compare treatment effect by stratifying patients by baseline risk of an outcome	Goligher et al,[21] Ely et al[31]
Effect-based modelling	Uses statistical and machine learning approaches to predict each patient's ITE using baseline characteristics, treatment assignment, and interaction terms between patient characteristics and treatment	Extensive validation required to avoid overfitting and spurious detection of HTE; data hungry (requires large sample size for adequate power); relies on patterns within the dataset to predict ITE, with minimal input based on existing literature or clinical insight	Predicted ITE computed on the basis of multiple baseline characteristics	Model derivation; this step can have high data cost because treatment and outcome data are needed to build the model	Model validation: compare ITEs predicted by the model to patient outcomes in the trial	Buell et al,[16] Seitz et al,[18] Goligher et al,[20] Blette et al,[23] Zampieri et al,[24]

APACHE=Acute Physiology and Chronic Health Evaluation. HTE=heterogeneity of treatment effect. ICEMAN=Instrument to assess the Credibility of Effect Modification Analyses. ITE=individualised treatment effect. *All analyses of HTE are limited by sample size and cannot address HTE driven by variables not measured in the study; all approaches identify associations between patient characteristics and treatment effect but do not establish causation.

Table 2: **Overview of statistical approaches for analysing heterogeneity of treatment effect**

Evaluation (APACHE) scores,[39] or models developed de novo within data from an RCT. The latter approach is susceptible to model overfitting on risk, particularly in RCTs with small sample sizes, and should be designed following existing guidance for prediction models.[40,41]

Potential advantages of risk-based analyses are that the prognostic models incorporate multiple characteristics to generate a single putative effect modifier (predicted risk) and allow treatment effects to be compared across the continuum of baseline risk, rather than a small number of discrete categories (eg, subgroups by low *vs* high APACHE score).[42] Risk-based analyses can be useful because the distribution of risk in clinical trials is generally skewed toward patients at lower risk of poor outcomes, who often have less potential to benefit from treatment than patients at higher risk.[42] Therefore, risk-based analysis can identify subpopulations at higher risk of poor outcomes who could derive more benefit from treatment.[8,43]

However, there might be no inherent relationship between a patient's baseline risk of an outcome and how beneficial a given treatment will be for that patient. Two patients with different characteristics (eg, age or severity of illness) and a different probability of responding to therapy could share the same baseline risk. For example, a previously healthy patient with refractory septic shock could have similarly high predicted risk of mortality as an elderly patient with less severe sepsis, but be more likely to respond to a sepsis-specific therapy, such as an immune modulator. The relationship between risk and treatment effect is complex and can vary depending on the clinical scenario. Indeed, previous risk-based HTE analyses in critical care have identified larger treatment effects in patients at higher risk of poor outcomes,[31,44] larger treatment effects in patients at lower risk of poor outcomes,[21] and no difference in treatment effects across baseline risk.[45] Detailed discussion of risk-based HTE analysis is outside the scope of this Review

and is thoroughly covered in previous reviews and the PATH statement.[8,15]

Effect-based analysis

Effect-based analyses model treatment effect for individual patients using statistical approaches, such as machine learning methods (appendix pp 10–11).[46,47] Effect-based analyses model the association between multiple baseline characteristics and the difference in outcomes between treatment groups to generate ITE predictions (table 2).[48] Because some effect-based modelling strategies can account for interactions between multiple variables that influence treatment effect (eg, the influence of age on treatment effect could depend on frailty, and vice versa), these models can theoretically provide more accurate estimates of individual treatment effects than subgroup or risk-based analyses. However, this strength also puts effect-based models at risk of statistical overfitting and identification of spurious HTE. Thus, rigorous statistical validation is required. Due to their complexity, effect-based models require large sample sizes to achieve adequate statistical power. As effect-based modelling is comparatively novel and might be unfamiliar to many clinicians, we provide a brief overview of this approach, in the context of a recently published example from the literature,[16] in the Quantifying ITE using effect-based models section. Relevant methods are surveyed in more detail in the appendix (pp 8–9, 14–15).

Quantifying ITE using effect-based models

To understand how effect-based models are used to quantify ITE, we consider the example of oxygen targets in mechanically ventilated patients. Two large trials comparing liberal versus conservative oxygen targets found no significant difference in mortality—the Pragmatic Investigation of Optimal Oxygen Targets (PILOT) trial and the Intensive Care Unit Randomised Trial Comparing Two Approaches to Oxygen Therapy (ICU-ROX).[22,49] A recent post-hoc analysis of these trials used effect-based modelling to quantify HTE and predict ITEs in the trial populations (figure 2A).[16] An overview of this approach is outlined in the appendix (pp 14–15).

The effect-based model was derived by fitting multiple types of effect model to the baseline characteristics of patients in the derivation cohort (the PILOT trial population). These effect models were compared using specific performance metrics (appendix p 12). The model that best predicted (on the basis of model performance metrics) ITE in the derivation cohort was then used to generate a predicted ITE for each patient in the validation cohort (ICU-ROX trial population). The validity of the model was assessed by comparing observed treatment effects between subgroups defined by predicted ITE. This study found that the subgroup of patients with a predicted ITE consistent with benefit from a lower oxygen target had a significant mortality benefit when receiving lower targets, whereas the subgroup of patients with a predicted ITE consistent with harm from lower oxygen targets had a significant increase in mortality when receiving lower targets, confirming the validity of the ITE estimates.

Once validated, effect-based models can be interrogated to understand which baseline characteristics had the greatest influence on ITE estimates (appendix p 13). For example, we might want to understand which characteristics are associated with a favourable response to a low oxygen target strategy. Although not showing causality, such information from effect-based models can help to identify factors associated with differential treatment responses, which in turn can be used to design experimental models to explore mechanisms of disease and treatment effect.

Designing clinical trials in the context of HTE

HTE has historically been viewed as a problem for clinical trial design, in which HTE is the noise that attenuates the signal of ATE. However, when the goal is to personalise evidence-based medicine, HTE provides valuable signal about treatment effect for individuals—necessary information that is neglected by a narrow focus on ATE. Conceptually, clinical trials can be designed to leverage HTE via at least two broad approaches: prospective detection of, and adaptation to, HTE during the trial or analyses of HTE at the conclusion of the trial.

Predictive biomarkers: addressing HTE in trial design

A predictive biomarker is any clinical, physiological, or biological characteristic that is hypothesised to predict differential treatment effect. For example, in addition to traditional biomarkers (eg, IL-6 concentration), a patient's baseline risk of an outcome or predicted ITE can be a predictive biomarker. Predictive biomarkers can have varying credentials—that is, there can be varying degrees of evidence that treatment effect will vary across the range of values for that predictive biomarker.[50] The appropriate trial design for addressing HTE will depend on the credentials of the proposed predictive biomarker.

Predictive enrichment

Designing trials to incorporate HTE is often discussed in relation to predictive enrichment. Predictive enrichment is the strategy of targeting enrolment towards patients expected to have a more favourable response to an intervention.[51] When predictive biomarker credentials are deemed to be very strong on the basis of pre-existing mechanistic or clinical trial data, trials could be designed to simply exclude patients who do not have the predictive biomarker.[50] This situation, however, is probably rare and this approach should be adopted with caution.

When predictive biomarker credentials are weaker (ie, it is less clear that the biomarker will predict differential treatment response), trials should enrol patients from the entire population. However, randomisation and analysis can be stratified based on the predictive biomarker, which enables efficient prospective estimates of ATE for

subpopulations defined by the predictive biomarker. To enhance trial efficiency, adaptive designs can discontinue enrolment separately in each biomarker-defined stratum when benefit or futility are confirmed.[52–54]

Several ongoing trials in critical care provide instructive examples of this adaptive approach. The Precision Medicine Adaptive Network Platform Trial in Hypoxemic Acute Respiratory Failure (known as PANTHER) is using biological phenotypes in ARDS as predictive biomarkers for novel therapeutics.[36] The Driving Pressure Limited Ventilation in Hypoxaemic Respiratory Failure (DRIVE) RCT on the Platform of Randomized Adaptive Clinical Trials in Critical Illness (PRACTICAL) platform is using respiratory system elastance as a predictive biomarker for ventilator management (NCT05440851).[36] The TRAITS trial is using a combination of biological and physiological abnormalities (termed treatable traits) to establish eligibility of patients who are critically ill for randomisation to therapeutics that are thought to be mechanistically linked to those traits.[55]

Identification of relevant predictive biomarkers during trials

Often, relevant predictive biomarkers are not known at the outset of a clinical trial. When there is a range of plausible potential predictive biomarkers, adaptive designs can be used to adapt enrolment or randomisation using information about relevant predictive biomarkers accumulated during the trial.[52–54] A seminal example of this approach comes from the Investigation of Serial studies to Predict Your Therapeutic Response with Imaging and Molecular Analyses 2 (I-SPY 2) trial, an ongoing adaptive platform phase 2 clinical trial of breast cancer therapies.[56] Patients with breast cancer are grouped into ten molecular subtypes based on hormone receptor status and gene score and randomly assigned to up to five different experimental therapies. As information about treatment response to each therapy in each subtype accumulates during the trial, response-adaptive randomisation is used to increasingly allocate patients to beneficial therapies within each subtype, maximising the probability of reaching a conclusion of benefit for a given therapy. If powered appropriately, this approach can also identify relevant predictive biomarkers during a trial. It is important to note that response-adaptive randomisation is only one approach to adaptive trial design and is subject to potential limitations. Other trial features can be adapted, including sample size, available interventions, eligibility criteria, or outcomes.[53] As with all clinical trials, the optimal adaptive design depends on the research question.

In critical care, the ongoing Anti-Thrombotic Therapy to Ameliorate Clinical Complications in Community-Acquired Pneumonia (or ATTACC-CAP; NCT05848713) trial uses a two-stage design to accommodate possible HTE. In the initial stage, treatment effects are estimated on the basis of combinations of several candidate effect modifiers. In the second stage, enrolment is restricted to patients with clinical characteristics predicting a higher probability of clinically meaningful treatment benefit from the initial analysis, thereby enriching the overall trial population for potential responders.

In theory, when predictive biomarkers are entirely unknown, one could design a trial using interim risk-based or effect-based models to identify relevant subpopulations who have differential treatment effects, and this information could be used to adapt the trial design in real-time. Given the large information requirements and analytical complexity of these analyses, such a trial would be challenging to execute. Nevertheless, this approach represents a compelling goal for future trial innovation.

Large-scale trials to predict ITEs

All of the aforementioned approaches to trial design involve designing trials to estimate ATEs within subgroups or subpopulations. An alternative approach could be to reorient the goal of clinical trials away from measuring ATEs altogether and instead aim to use the information accrued in RCTs to generate valid models for predicting ITE. Using simple designs, broad inclusion criteria, and electronic health record-based or registry-based data collection, pragmatic trials can enrol large, representative populations and obtain granular data on baseline characteristics and outcomes—all of which can facilitate effect-based modelling to predict ITEs.[57] However, the derivation of such effect-based models, even when prespecified, requires rigorous validation. Furthermore, such large-scale trials are typically restricted to comparing treatments that patients are already receiving in clinical care (eg, intravenous fluids or intubation strategies[17]), rather than evaluating new drugs or devices. Initiatives focused on lowering trial costs, enrolling larger sample sizes, and streamlining the collection and harmonisation of comprehensive datasets between trials will help to facilitate HTE evaluations.

Challenges in applying HTE in practice

The ability to predict ITEs offers new and compelling means to implement evidence-based personalised medicine. Nevertheless, there are many challenges to applying HTE analyses at the bedside.

The evidentiary threshold required to justify the use of predicted ITE to guide treatment in clinical practice is controversial. One of the most well known examples of applying ITE to inform clinical decision making is the Atherosclerotic Cardiovascular Disease (ASCVD) Risk Assessment Tool, which is now widely disseminated despite no prospective randomised assessment of its benefit.[58] This approach was feasible because the ASCVD tool was built on well validated cardiovascular risk factors. In critical care, where factors associated with treatment benefit are less clearly established, ITE models are only hypothesis-generating and need to be

prospectively validated before adoption into practice. This process of prospective validation, or clinical confirmation, is distinct from the statistical model validation discussed above. Clinical confirmation should aim to verify that the predicted ITE accurately identifies who benefits from treatment and to evaluate whether personalising care on the basis of predicted ITE improves outcomes compared with standard approaches. There is no consensus framework for prospectively establishing when ITE prediction models are adequately validated and confirmed ready for use in clinical practice.

As critical care has so far had little experience with ITE models, rigorous clinical confirmation with prospective trials is probably needed. However, traditional RCTs could face challenges in evaluating such ITE-based decision tools given the large sample sizes needed to detect the expected benefits and the inability to assess the acceptability of the decision tool in routine clinical care. Achieving requisite sample sizes for confirmation trials might be feasible for common interventions, such as oxygen targets, but is more challenging for trials of novel therapeutics. As we gain more experience with methods for analysing HTE, we can improve our understanding of required sample sizes for ITE models and use HTE to improve the precision of our trials, even within sample size limitations. In the meantime, in addition to addressing sample size limitations by advocating for improved trial infrastructure and harmonisation of data between trials, incorporation of additional data could help to overcome these barriers. For example, prospective target trial emulation using observational data could be a useful tool for HTE analyses, providing an opportunity to benchmark findings from RCTs and allowing generalisation of findings to populations not investigated in the original trial.[59] Similarly, for novel therapeutics, phase 4 and post-marketing surveillance studies could represent potential opportunities to both discover HTE and validate HTE found in earlier-phase trials, as long as granular patient-level data are available.

Implementation of predicted ITE also presents an important challenge. To be translated to the bedside, predicted ITE must be presented in such a manner that clinicians can act on the information to improve management, which will likely require real-time clinical decision support tools (figure 1). The success of such decision tools hinges on the validity of the ITE-based model, real-time availability of information required for the model, and clinician willingness to use the tool. The latter consideration is important, as clinicians have historically had reservations about using complex black-box models to make treatment decisions.[60]

There could also be regulatory considerations for prospective validation trials, as clinical decision support tools based on complex ITE models might be viewed as medical devices in some jurisdictions.[61]

There are additional practical considerations for applying ITE-based decision tools in clinical care. Importantly, the quality and timing of real-time data might differ from data used to develop ITE models. Complex models often require multiple data inputs that might or might not be routinely available in clinical practice (eg, real-time research biomarkers), which could restrict their application. Conversely, clinicians might use factors in clinical decision making that are not included in ITE models. For example, current methods for predicting ITE consider only baseline, single time-point variables, yet critical illness is dynamic and these variables often rapidly evolve during a patient's illness. Considering the planned application of ITE models from the outset could help to build models that translate well to clinical practice.

Limitations of ITE

The treatment effects estimated in risk-based and effect-based analyses are not always clinically meaningful. The appropriateness of ITE-guided treatment depends on multiple factors, including the direction and magnitude of treatment effect, availability of treatment, associated harms and costs, and the patient's goals and values. For high-risk or high-cost treatments (eg, extracorporeal CO_2 removal), HTE could have an important influence on practice. For example, in fields other than critical care, findings from HTE analyses have helped to target surgical interventions to appropriate candidates.[62] In contrast, for inexpensive, low-risk treatments (eg, paracetamol for a headache), detailed investigation of HTE is unlikely to change practice and pursuing the large-scale trials required to predict and validate ITE could waste valuable resources. As such, performing HTE analyses might not always be appropriate, especially in small trials or when evidence for HTE from clinical or translational studies is scarce.

Furthermore, even when HTE is suspected, the optimal approach to analysing HTE has not been established. In general, we challenge the convention that subgroup analyses should be held as the standard for analysing HTE in clinical trials. Risk-based and effect-based approaches are methodologically superior to conventional subgroup analysis, as they allow evaluation of treatment effect across multiple characteristics simultaneously. Therefore, when HTE is suspected (eg, when studying broad clinical syndromes such as sepsis or ARDS), risk-based or effect-based analyses, or both, should be prospectively incorporated into trial analysis plans in addition to—or even in lieu of—conventional subgroups.[63] However, as a field, our experience with HTE analyses and ITE is limited, and at this time it is not possible to identify a clear framework for when to use risk-based versus effect-based approaches. As we gain more experience with these approaches, we must continue to evaluate and compare these methods with the aim of developing

***Panel*: Proposed future directions for research to advance analyses of heterogeneity of treatment effect and individualised treatment effect in critical care**

- Bring together diverse perspectives from patients, clinicians, policy makers, and regulators to work towards consensus on the goals of heterogreneity of treatment effect (HTE) analyses.
- Develop statistical approaches that address limitations of existing risk-based and effect-based models, particularly sample size limitations and risk of overfitting.
- When HTE is suspected, incorporate prespecified risk-based or effect-based models, or both, in RCT analysis plans.
- Continue validating risk-based and effect-based methods to develop a framework to establish when these methods should be used as the primary method for analysing HTE within clinical trials.
- Use existing trial design features to understand HTE in critical care trials, particularly in areas where large HTE is expected.
- Continue to support initiatives focused on lowering trial costs to enrol larger sample sizes, and streamlining the collection of richer datasets. The ability to accrue comprehensive data efficiently and to harmonise these data between trials will facilitate HTE evaluations.
- Develop frameworks for confirming findings of HTE analyses and testing decision rules based on predicted ITE in clinical practice.
- Consider the clinical implications of HTE analyses from the outset of a trial, balancing the potential benefits of predicting ITEs with the resources required for risk-based and effect-based analyses and the limitations of applying results from these models to under-represented patient populations.

Search strategy and selection criteria

This Review is a summary statement from an international Round Table—organised by the Platform of Randomized Adaptive Clinical Trials in Critical Illness (PRACTICAL) platform trial and Evidence-based Individualized Treatment Effects (EvITE) Group—on heterogeneity of treatment effect (HTE) in critical care. To review current methodologies for HTE analyses, identify challenges, and develop a framework for approaching HTE in critical care trials, over 30 clinician-scientists and statisticians were invited to attend a dedicated Round Table meeting in Toronto, Canada, on Nov 27, 2023. Invitees were selected on the basis of published expertise and previous engagement in HTE analysis and methods within critical care. The meeting was sponsored by the PRACTICAL) and the EvITE Group. To mentor future clinician-scientists, eight early-career researchers were also invited to attend. The Round Table meeting was anchored around prespecified meeting objectives and organised into four sessions. Each session included topic-focused presentations reviewing current literature, identified by the speakers, followed by structured discussion. After the meeting, working groups were convened to conduct additional literature reviews and summarise the evidence and group discussions for each section. Methodological details are provided in the appendix (pp 3–5) along with full author contributions (appendix pp 6–7). Definitions, methodology, and challenges proposed within the Round Table meeting were iteratively refined via discussion within the larger EvITE working group and with additional selected experts in the field. Data for this Review were identified by searches of PubMed and references from relevant articles (last search conducted May, 2024) using the search terms "heterogeneity of treatment effect", "subgroup analysis", "conditional average treatment effect", "individualised treatment effect", "risk-based analysis", and "effect-based analysis." Articles published in English between 1955 and 2024 were considered. For simplicity, in this Review we focus the discussion on the outcome of short-term mortality and assume that patients uniformly prefer to avoid death over other outcomes.

guidance for their incorporation into trial analysis plans.[64]

The application of ITE models in clinical care also risks exacerbating health-care disparities, if not approached carefully. First, applying results from adaptive trials could lead to overly narrow clinical guidelines or drug approvals, or both, if benefits are observed only in small subgroups. Additionally, if a population of patients is under-represented in a clinical trial,[65] then a model of ITE derived from that clinical trial might not validly inform care for that population.[66–68] How well this challenge is addressed by de-biasing techniques[69,70] or approaches that combine observational and RCT data in an effort to improve generalisability remains uncertain.[71,72] Given this risk of bias, studies should report detailed information on the characteristics of the patient population in which ITE models were developed, and confirmation studies should pay particular attention to how tools and interventions perform across differing patient populations.[73] Additionally, since patients from low-income and middle-income countries are often not included in clinical trials conducted in high-income countries, models built in such trials could have limited applicability in vast regions of the world, increasing pre-existing disparities in the delivery of care.[74,75] In all cases, HTE analyses should be undertaken with careful consideration of the potential clinical effect, required resources, and generalisability of the results.

Conclusions and future directions

The ATEs reported in randomised trials represent the traditional reference standard for evidence-based medicine but often do not represent the effect of treatment on outcomes for individual patients. By tackling the challenges (panel) involved in using RCTs to derive, validate, and inform implementation of ITE, the critical care community can facilitate progress towards the goal of evidence-based personalised medicine.

Contributors

ESM, AS, AC-F, AZ, JD, EGL, VT, MU, EF, MWS, MC, and ECG contributed to the conception and design of the work, literature search, and figures, writing of the original draft, and review and editing of the manuscript. DCA, EB-K, BB, LDB, KGB, JDC, CSC, LDS, EE, NDF, RG, AG, MOH, AH, CH, TH, CJ, LK, PRL, AL, FL, KL, AM, DM, CM, DFM, ASN, CO, AP, HCP, BR, MS-H, SS, MS, MES, KPS, FS, PS, ASS, ETQ, SAW, PJY, FGZ, and RZ contributed to the conception of the work, and review and editing of the manuscript. All authors approved the final manuscript.

Declaration of interests

ESM reports grant funding from the US National Institutes of Health (NIH; grant numbers T32HL007749, F32HL172463, and L30HL170379). AS reports grants through her institution from NIH (National Heart, Lung, and Blood Institute [NHLBI] grant number R01HL157262-02 and National Institute of General Medical Sciences [NIGMS] grant number R01GM123193) and US Department of Defense (DOD; grant number PRMRP W81XWH-21-1-0009). EGL received grant funding from NIH (grant numbers T155T15LM007359 and NHLBI R01HL157262). MU received funding from the Canadian Institutes of Health Research (CIHR) and the Interdepartmental Division of Critical Care Medicine at the University of Toronto (ON, Canada). DCA received grant funding from NIH (grant number NIGMS R01GM141081); and personal fees from Abionyx, AM-Pharma, Bayer, Inotrem, and Spectral. VT, DCA, and

FS receive grant funding from NIH (grant number R01GM141081). LDB reports being on the advisory board of Sobi NL, Impentri, Novartis, AstraZeneca, and CSL Behring; received grant funding (via institution) from Volition and Santhera; received grant funding from Longfonds, Innovative Medicine Initiative, Amsterdam University Medical Center, Health Holland via Longfond, and ZonMw; and received personal fees (via institution) from Scailyte. KGB received grant funding from NIH–NHLBI (grant number T32HL007605). CSC is a council member of the International Sepsis Forum; reports grants through her institution from NIH, Roche Genentech, and Quantum Leap Healthcare; received personal fees from Vasomune, Gen1e Life Sciences, NGM Bio, Cellenkos, Calcimedica, Arrowhead, EnliTISA, and Novartis; has a patent on metagenomic sequencing for sepsis diagnosis issued through the Regents of the University of California (CA, USA) and Chan Zuckerberg BioHub; and was supported by an NHLBI grant (R35HL140026). LDS reports grant funding through his institution from the Canadian Mount Sinai Hospital University Health Network Academic Medical Organization Innovation Fund and CIHR; and has applied for a patent for biomarkers for detecting of outcome or risk of patients with a respiratory illness. AG received grant funding through his institution from Sygeforsikringen "danmark", the Novo Nordisk Foundation, and the Independent Research Fund Denmark for other projects. AH reports grant funding as Canada Research Chair in Statistical Design and from the Natural Sciences and Engineering Research Council of Canada. MOH received grant funding from NIH (grant numbers U01HL168419 and R01HL168202); and personal fees from Elsevier, the American Thoracic Society, Unlearn.AI, and Guidepoint. CJ reports grant funding from the CIHR; and a postdoctoral fellowship awarded by the StatLab at the Centre de Recherches Mathématique. PRL received financial support from the Fonds de Recherche de Quebec. FL reports grant funding from NIH–NHLBI (grant number R01HL168202) and the Patient-Centered Outcomes Research Institute (PCORI; ME-2020C3-21072, ME-2020C1-19220, and ME-2022C2-27676); and has received personal fees from PCORI as a consultant. AM reports grant funding from NIH (grant number 1K23GM150110) and the American College of Surgeons. DFM reports personal fees for consultancy for acute respiratory distress syndrome from GlaxoSmithKline, Boehringer Ingelheim, Novartis, Bayer, Aptarion, Direct Biologics, Eli Lilly, Sobi, and Aviceda; and personal fees for sitting on a data and safety monitoring board (DSMB) from Vir Biotechnology and Faron Pharmaceuticals; his institution has received funds from grants from the National Institute for Health and Care Research (NIHR), Wellcome Trust, UK Medical Research Council, Innovate UK, Randox, Novavax, and Northern Ireland Health and Social Care R&D Division, all unrelated to this work; has a patent (issued to institution) for a treatment for acute respiratory distress syndrome; and is NIHR Scientific Director for Research Programmes. AP received grant funding through his institution from the Novo Nordisk Foundation, Sygeforsikringen, and the Independent Research Fund Denmark. HCP reports being on the DSMB of the AID-ICU trial and SMARA trial; is co-chair of the Surviving Sepsis Campaign Guidelines; and received grants from the NIH–NHLBI, Agency for Healthcare Research and Quality, US Veterans Affairs, Blue Cross Blue Shield of Michigan, and US Centers for Disease Control and Prevention. ASS reports personal fees for speaking, or consultancy from Draeger Medical, Predicate, SafeBVM, SaNOtize, Sensory Cloud, and Stimit; sits on the Board of Thornhill Medical and De Motu Cordis; chairs the DSMBs of the PREVENT VILI (NIH-sponsored), RheSolve (Galvanize Therapeutics), and BREATHE-3 (Apreo Health) trials; and reports shares in Thornhill Medical, De Motu Cordis, Signal 1, SafeBVM, and Sensory Cloud. SS received grant funding through her institution from the NIH–NHLBI and DOD; and personal fees from Getinge. MS reports financial support from her tier 2 Canada Research Chair in Causal Inference and Machine Learning. KPS received grant funding from the NIH–NHLBI (grant number T32HL087738). FGZ received personal fees from Baxter Hospitalar. RZ received grant funding from the CIHR, Research Manitoba, and Octapharma. EF is supported by grants from the CIHR; and reports personal fees for speaking or consultancy from Aerogen, ALung Technologies, Baxter, Boehringer Ingelheim, Getinge, Inspira, Mallinckrodt, and Vasomune. MC received grant funding through his institution from the NIH–NHLBI (grant number R01HL157262-02), NIH–National Institute of Diabetes and Digestive Kidney Diseases (grant number R01DK126933A-01), NIH–NIGMS (grant numbers R35-13362546 and R01GM123193), and the DOD (grant number PRMRP W81XWH-21-1-0009); has a patent issued for a risk stratification algorithm (eCART) and receives royalties through his institution. ECG receives salary support and operating grant funds from the National Sanitarium Association; received grants from the CIHR; reports personal fees for speaking or consultancy from Getinge, Vyaire, Lungpacer, Stimit, Bioage, Heecap, Baxter, Zoll, and Drager; is on the advisory board of Getinge, Lungpacer, and Heecap; is on the DSMB of HiOX-SR; and received grant funding through his institution from the CIHR and National Sanitarium Association. All other authors declare no competing interests.

Acknowledgments

The authors acknowledge the following organisations, which provided organisational support and funding towards meeting costs for the Round Table meeting: the Platform of Randomized Adaptive Clinical Trials in Critical Illness, the University Health Network Foundation, and the Evidence-based Individualized Treatment Effects Group. These organisations had no role in the design or writing of this Review or in the decision to submit this manuscript for publication. The content of this Review is solely the responsibility of the authors and does not necessarily represent the official views of the US National Institutes of Health, US Department of Veterans Affairs, US Government, or the Patient-Centered Outcomes Research Institute.

References

1 Angus DC, Chang CH. Heterogeneity of treatment effect: estimating how the effects of interventions vary across individuals. *JAMA* 2021; **326:** 2312–13.

2 Kravitz RL, Duan N, Braslow J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q* 2004; **82:** 661–87.

3 The National Heart, Lung, and Blood Institute Prevention and Early Treatment of Acute Lung Injury Clinical Trials Network. Early restrictive or liberal fluid management for sepsis-induced hypotension. *N Engl J Med* 2023; **388:** 499–510.

4 Meyhoff TS, Hjortrup PB, Wetterslev J, et al. Restriction of intravenous fluid in ICU patients with septic shock. *N Engl J Med* 2022;**386:** 2459–70.

5 Goligher EC, Heath A. Bringing clinical trials to the bedside: conditional average treatment effects in the Balanced Solution versus Saline in Intensive Care Study (BaSICS) trial. *Ann Am Thorac Soc* 2023; **20:** 794–96.

6 Vegetabile BG. On the distinction between "conditional average treatment effects" (CATE) and "individual treatment effects" (ITE) under ignorability assumptions. *arXiv* 2021; published online Aug 10. https://doi.org/10.48550/arXiv.2108.04939.

7 Rothwell PM, Mehta Z, Howard SC, Gutnikov SA, Warlow CP. From subgroups to individuals: general principles and the example of carotid endarterectomy. *Lancet* 2005; **365:** 256–65.

8 Iwashyna TJ, Burke JF, Sussman JB, Prescott HC, Hayward RA, Angus DC. Implications of heterogeneity of treatment effect for reporting and analysis of randomized trials in critical care. *Am J Respir Crit Care Med* 2015; **192:** 1045–51.

9 Ioannidis JP, Lau J. The impact of high-risk patients on the results of clinical trials. *J Clin Epidemiol* 1997; **50:** 1089–98.

10 Maslove DM, Tang B, Shankar-Hari M, et al. Redefining critical illness. *Nat Med* 2022; **28:** 1141–48.

11 Harhay MO, Wagner J, Ratcliffe SJ, et al. Outcomes and statistical power in adult critical care randomized trials. *Am J Respir Crit Care Med* 2014; **189:** 1469–78.

12 The National Heart, Lung, and Blood Institute ARDS Clinical Trials Network. Higher versus lower positive end-expiratory pressures in patients with the acute respiratory distress syndrome. *N Engl J Med* 2004; **351:** 327–36.

13 Shah FA, Talisa VB, Chang CH, et al. Heterogeneity in the effect of early goal-directed therapy for septic shock: a secondary analysis of two multicenter international trials. *Crit Care Med* 2025; **53:** e4–14.

14 Goligher EC, Kavanagh BP, Rubenfeld GD, et al. Oxygenation response to positive end-expiratory pressure predicts mortality in acute respiratory distress syndrome. A secondary analysis of the LOVS and ExPress trials. *Am J Respir Crit Care Med* 2014; **190:** 70–76.

15 Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement. *Ann Intern Med* 2020; **172:** 35–45.

16 Buell KG, Spicer AB, Casey JD, et al. Individualized treatment effects of oxygen targets in mechanically ventilated critically ill adults. *JAMA* 2024; **331:** 1195–204.

17 Driver BE, Semler MW, Self WH, et al. Effect of use of a bougie vs endotracheal tube with stylet on successful intubation on the first attempt among critically ill patients undergoing tracheal intubation: a randomized clinical trial. *JAMA* 2021; **326:** 2488–97.

18 Seitz KP, Spicer AB, Casey JD, et al. Individualized treatment effects of bougie versus stylet for tracheal intubation in critical illness. *Am J Respir Crit Care Med* 2023; **207:** 1602–11.

19 McNamee JJ, Gillies MA, Barrett NA, et al. Effect of lower tidal volume ventilation facilitated by extracorporeal carbon dioxide removal vs standard care ventilation on 90-day mortality in patients with acute hypoxemic respiratory failure: the REST randomized clinical trial. *JAMA* 2021; **326:** 1013–23.

20 Goligher EC, McNamee JJ, Dianti J, et al. Heterogeneous treatment effects of extracorporeal CO_2 removal in acute hypoxemic respiratory failure. *Am J Respir Crit Care Med* 2023; **208:** 739–42.

21 Goligher EC, Lawler PR, Jensen TP, et al. Heterogeneous treatment effects of therapeutic-dose heparin in patients hospitalized for COVID-19. *JAMA* 2023; **329:** 1066–77.

22 The ICU-ROX Investigators and the Australian and New Zealand Intensive Care Society Clinical Trials Group. Conservative oxygen therapy during mechanical ventilation in the ICU. *N Engl J Med* 2020; **382:** 989–98.

23 Blette BS, Granholm A, Li F, et al. Causal Bayesian machine learning to assess treatment effect heterogeneity by dexamethasone dose for patients with COVID-19 and severe hypoxemia. *Sci Rep* 2023; **13:** 6570.

24 Zampieri FG, Damiani LP, Bagshaw SM, et al. Conditional treatment effect analysis of two infusion rates for fluid challenges in critically ill patients: a secondary analysis of Balanced Solution Versus Saline in Intensive Care Study (BaSICS) trial. *Ann Am Thorac Soc* 2023; **20:** 872–79.

25 Schandelmaier S, Briel M, Varadhan R, et al. Development of the Instrument to Assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ* 2020; **192:** e901–06.

26 Calfee CS, Delucchi K, Parsons PE, et al. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med* 2014; **2:** 611–20.

27 Sinha P, Kerchberger VE, Willmore A, et al. Identifying molecular phenotypes in sepsis: an analysis of two prospective observational cohorts and secondary analysis of two randomised controlled trials. *Lancet Respir Med* 2023; **11:** 965–74.

28 Famous KR, Delucchi K, Ware LB, et al. Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. *Am J Respir Crit Care Med* 2017; **195:** 331–38.

29 Calfee CS, Delucchi KL, Sinha P, et al. Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. *Lancet Respir Med* 2018; **6:** 691–98.

30 Sinha P, Furfaro D, Cummings MJ, et al. Latent class analysis reveals COVID-19-related acute respiratory distress syndrome subgroups with differential responses to corticosteroids. *Am J Respir Crit Care Med* 2021; **204:** 1274–85.

31 Ely EW, Laterre PF, Angus DC, et al. Drotrecogin alfa (activated) administration across clinically important subgroups of patients with severe sepsis. *Crit Care Med* 2003; **31:** 12–19.

32 Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005; **365:** 176–86.

33 Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ* 2015; **351:** h5651.

34 Grant RW, McCloskey J, Hatfield M, et al. Use of latent class analysis and k-means clustering to identify complex patient profiles. *JAMA Netw Open* 2020; **3:** e2029068.

35 Sinha P, Calfee CS. Phenotypes in acute respiratory distress syndrome: moving towards precision medicine. *Curr Opin Crit Care* 2019; **25:** 12–20.

36 PRACTICAL, PANTHER, TRAITS, INCEPT, and REMAP-CAP investigators. The rise of adaptive platform trials in critical care. *Am J Respir Crit Care Med* 2024; **209:** 491–96.

37 Lawler PR, Fan E. Heterogeneity and phenotypic stratification in acute respiratory distress syndrome. *Lancet Respir Med* 2018; **6:** 651–53.

38 Shankar-Hari M, Santhakumaran S, Prevost AT, et al. Defining phenotypes and treatment effect heterogeneity to inform acute respiratory distress syndrome and sepsis trials: secondary analyses of three RCTs. *Effic Mech Eval* 2021; **8:** 1–104.

39 Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006; **34:** 1297–310.

40 Harhay MO, Au DH, Dell SD, et al. Methodologic guidance and expectations for the development and reporting of prediction models and causal inference studies. *Ann Am Thorac Soc* 2020; **17:** 679–82.

41 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC Med* 2015; **13:** 1.

42 Kent DM, Hayward RA. Limitations of applying summary results of clinical trials to individual patients: the need for risk stratification. *JAMA* 2007; **298:** 1209–12.

43 Kerr EA, Hayward RA. Patient-centered performance management: enhancing value for patients and health care systems. *JAMA* 2013; **310:** 137–38.

44 McKown AC, Huerta LE, Rice TW, Semler MW. Heterogeneity of treatment effect by baseline risk in a trial of balanced crystalloids versus saline. *Am J Respir Crit Care Med* 2018; **198:** 810–13.

45 Santhakumaran S, Gordon A, Prevost AT, O'Kane C, McAuley DF, Shankar-Hari M. Heterogeneity of treatment effect by baseline risk of mortality in critically ill patients: re-analysis of three recent sepsis and ARDS randomised controlled trials. *Crit Care* 2019; **23:** 156.

46 Wang G, Heagerty PJ, Dahabreh IJ. Using effect scores to characterize heterogeneity of treatment effects. *JAMA* 2024; **331:** 1225–26.

47 Chen X, Harhay MO, Tong G, Li F. A Bayesian machine learning approach for estimating heterogeneous survivor causal effects: applications to a critical care trial. *Ann Appl Stat* 2024; **18:** 350–74.

48 Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc* 2018; **113:** 1228–42.

49 Semler MW, Casey JD, Lloyd BD, et al. Oxygen-saturation targets for critically ill adults receiving mechanical ventilation. *N Engl J Med* 2022; **387:** 1759–69.

50 Freidlin B, Korn EL. Biomarker enrichment strategies: matching trial design to biomarker credentials. *Nat Rev Clin Oncol* 2014; **11:** 81–90.

51 Stanski NL, Wong HR. Prognostic and predictive enrichment in sepsis. *Nat Rev Nephrol* 2020; **16:** 20–31.

52 Thall PF. Adaptive enrichment designs in clinical trials. *Annu Rev Stat Appl* 2021; **8:** 393–411.

53 Thorlund K, Haggstrom J, Park JJ, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians. *BMJ* 2018; **360:** k698.

54 Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med* 2017; **377:** 62–70.

55 The University of Edinburgh. TRAITS: enabling precision medicine for intensive care. 2023. https://traits-trial.ed.ac.uk (accessed July 22, 2024).

56 Wang H, Yee D. I-SPY 2: a neoadjuvant adaptive clinical trial designed to improve outcomes in high-risk breast cancer. *Curr Breast Cancer Rep* 2019; **11:** 303–10.

57 DeMasi SC, Self WH, Semler MW. Oxygen targets and the future of critical care clinical research. *Am J Respir Crit Care* Med 2023; **208:** 746–48.

58 Lloyd-Jones DM, Braun LT, Ndumele CE, et al. Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: a special report from the American Heart Association and American College of Cardiology. *Circulation* 2019; **139:** e1162–77.

59 Admon AJ, Donnelly JP, Casey JD, et al. Emulating a novel clinical trial using existing observational data. Predicting results of the PreVent study. *Ann Am Thorac Soc* 2019; **16:** 998–1007.

60 Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med* 2020; **3:** 17.

61 International Medical Device Regulators Forum. Software as a Medical Device (SaMD): key definitions. Dec 9, 2013. International Medical Device Regulators Forum, 2013.

62 Rothwell PM, Warlow CP, on behalf of the European Carotid Surgery Trialists' Collaborative Group. Prediction of benefit from carotid endarterectomy in individual patients: a risk-modelling study. *Lancet* 1999; **353:** 2105–10.

63 Goligher EC. The rise of adaptive platform trials in critical care. *Am J Respir Crit Care Med* 2024; **209:** 491–96.

64 Selby JV, Maas CCHM, Fireman BH, Kent DM. Potential clinical impact of predictive modeling of heterogeneous treatment effects: scoping review of the impact of the PATH Statement. *medRxiv* 2024: published online Feb 21. https://doi.org/10.1101/2024.05.06.24306774 (preprint).

65 Ma MA, Gutiérrez DE, Frausto JM, Al-Delaimy WK. Minority representation in clinical trials in the United States: trends over the past 25 years. *Mayo Clin Proc* 2021; **96:** 264–66.

66 Li J, Bzdok D, Chen J, et al. Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity. *Sci Adv* 2022; **8:** eabj1812.

67 Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat Med* 2021; **27:** 2176–82.

68 Wang R, Chaudhari P, Davatzikos C. Bias in machine learning models can be significantly mitigated by careful training: evidence from neuroimaging studies. *Proc Natl Acad Sci USA* 2023; **120:** e2211613120.

69 Yang J, Soltan AAS, Eyre DW, Yang Y, Clifton DA. An adversarial training framework for mitigating algorithmic biases in clinical machine learning. *NPJ Digit Med* 2023; **6:** 55.

70 Eid FE, Elmarakeby HA, Chan YA, et al. Systematic auditing is essential to debiasing machine learning in biology. *Commun Biol* 2021; **4:** 183.

71 Colnet B, Mayer I, Chen G, et al. Causal inference methods for combining randomized trials and observational studies: a review. *Stat Sci* 2024; **39:** 165–91.

72 Bica I, Alaa AM, Lambert C, van der Schaar M. From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges. *Clin Pharmacol Ther* 2021; **109:** 87–100.

73 Finlayson SG, Subbaswamy A, Singh K, et al. The clinician and dataset shift in artificial intelligence. *N Engl J Med* 2021; **385:** 283–86.

74 Riviello ED, Kiviri W, Twagirumugabe T, et al. Hospital incidence and outcomes of the acute respiratory distress syndrome using the Kigali modification of the Berlin definition. *Am J Respir Crit Care Med* 2016; **193:** 52–59.

75 Andrews B, Semler MW, Muchemwa L, et al. Effect of an early resuscitation protocol on in-hospital mortality among adults with sepsis and hypotension: a randomized clinical trial. *JAMA* 2017; **318:** 1233–40.