Check for updates

Bringing Clinical Trials to the Bedside: Conditional Average Treatment Effects in the Balanced Solution versus Saline in Intensive Care Study (BaSICS) Trial

Ewan C. Goligher, M.D., Ph.D.[1,2,3,4] and Anna Heath, M.Math., Ph.D.[5,6,7]

[1]Interdepartmental Division of Critical Care Medicine, [4]Department of Physiology, and [6]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; [2]Division of Respirology, Department of Medicine, University Health Network, Toronto, Ontario, Canada; [3]Toronto General Hospital Research Institute, Toronto, Ontario, Canada; [5]Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada; and [7]Department of Statistical Science, University College London, London, United Kingdom

ORCID ID: 0000-0002-0990-6701 (E.C.G.).

In this issue of *AnnalsATS*, Zampieri and colleagues (pp. 872–879) present the results of an unplanned secondary analysis of the BaSICS randomized trial (1) aimed at exploring heterogeneous treatment effects (HTEs) in the trial (2). In BaSICS, critically ill patients were randomized to faster versus slower infusion rates of intravenous boluses of crystalloid fluids (and to balanced crystalloid vs. saline solution in a factorial design; the latter randomization was not considered in the present study). The goal of this *post hoc* secondary analysis of BaSICS was to identify potential factors predicting benefit from either the slower or faster infusion rate to inform individualized clinical decision making. The authors report that they could identify subpopulations with a high probability of benefit (>90%) from the faster infusion rate or with a high probability of benefit from the slower infusion rate. This work is important, and not merely for its

hypothesis-generating clinical implications about fluid management in critical illness. To appreciate its importance, consider the two following vignettes.

Vignette 1: At a major international scientific congress, the results of the latest major randomized clinical trial (RCT) in critical care medicine testing a highly promising intervention are announced. The audience sighs with disappointment. After enrolling hundreds of patients, the investigators found "no significant difference" in mortality between intervention and control. "Another negative trial," you think to yourself with a measure of unsurprised disappointment. After all, most trials in critical care medicine seem to turn up "negative" results (3). You find yourself wondering, is there really any point in doing clinical trials if we rarely succeed in finding interventions that work for patients? Is the massive investment of time and resources and money in trials worth it if we're not informing or improving practice?

Vignette 2: You stand at the bedside of an acutely ill patient, overseeing resuscitation efforts. The patient has bilateral opacities on chest x-ray and is increasingly hypoxemic, and you decide to apply higher positive end-expiratory pressure (PEEP) on the ventilator. The patient's peripheral oxygen saturation improves, but driving pressure is now elevated at 18 cm H_2O, despite limiting tidal volume to the standard protective target of 6 ml/kg predicted body weight. Aware of recent evidence from meta-analysis suggesting that, on average, higher PEEP is associated with lower mortality in acute respiratory distress syndrome (4), you are inclined to maintain the higher PEEP level, but, at the same time, you have doubts. If higher PEEP and slightly improved oxygenation come at

the cost of higher driving pressures, doesn't that mean you're overdistending the lung? Are the results of that meta-analysis really generalizable to this patient? This patient might have met eligibility criteria for those trials (our usual criterion for generalizability), but is this patient the typical or representative case in those trials?

These vignettes highlight a basic and widely appreciated problem with applying information in trials. RCTs are undertaken to measure the unmeasurable: the counterfactual clinical outcome of interest under intervention and under nonintervention in *the same patient*. But RCTs do not quite yield this information; rather, they measure the difference in outcomes between exposure to treatment and exposure to control, *on average*, across two *populations* of patients. The reported effect size is the average treatment effect (ATE) across these populations. But as every intensive care unit (ICU) practitioner knows, there is no "average patient." Critically ill patients are highly heterogeneous, and many patient, disease, or management characteristics might modify the effect of treatment. At best, we can hope that if we systematically treat our population of patients in a manner similar to that delivered in a trial, we are likely to achieve a similar treatment effect, on average, in the population (provided the "population" you are treating is reasonably similar to the population in the trial). Perhaps more important, there may be individuals who benefited from or were harmed by intervention, even if the ATE is nil. Hence, the ATE is not sufficiently informative for those tasked with the care of individual patients.

All this highlights a fundamental tension between the epistemological concerns of the rarefied world of clinical trials and the practical concerns of treating clinicians, their
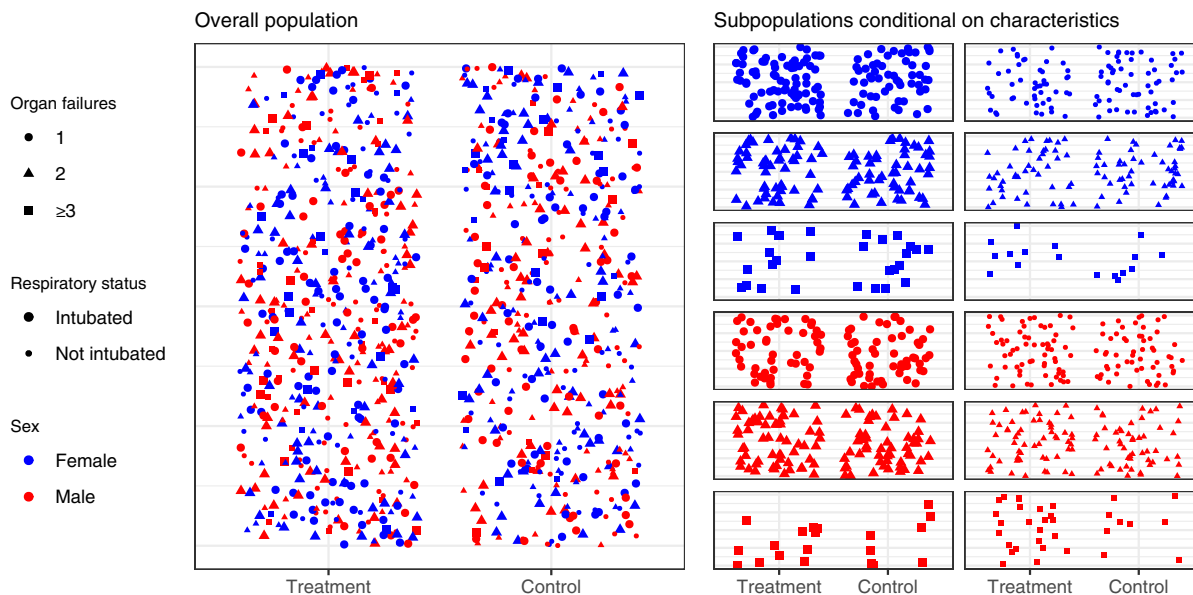
Figure 1. Measuring treatment effect in a clinical trial. The treatment effect (difference in outcome rate between treatment and control) for the overall population represents the average treatment effect across all patients (left panel). Treatment effect can also be estimated within subpopulations defined by multiple patient characteristics (e.g., sex, number of organ failures, use of invasive ventilation) (right panel); this represents the conditional average treatment effect—the average treatment effect conditional on subpopulation characteristics. Although there will still be additional unmeasured heterogeneous treatment effects within subpopulations, the conditional average treatment effect more closely resembles the treatment effect in those individual patients.

patients, and patients' families. Do we treat populations, or do we treat individuals? If RCTs provide ATEs in populations, how applicable are their results (whether "positive" or "negative") to the unique patient whom I find under my care at this moment?

Estimating more individualized treatment effects from clinical trials is therefore of real interest and importance. Although we cannot quite measure truly "individual" treatment effects (because any given individual can only be randomized to treatment or control at a given time; the counterfactual is not observable), we can quantify differences among treated and untreated patients in subpopulations defined by shared characteristics (5). The treatment effect observed in the subpopulation is still an ATE, but it is an ATE conditional on having those characteristics (Figure 1). Hence, it is referred to as the "conditional ATE" (cATE). Unlike traditional subgroup analysis (where each analysis only considers one "dimension"), newer methods combine multiple "dimensions" of HTE to generate more sensitive and informative models for predicting variation in treatment effect.

The work by Zampieri and colleagues provides a highly instructive example of such innovation to explore HTEs in trials.

They derived a logistic regression model combining multiple potential determinants of treatment effect. This model then predicted a cATE for each patient according to their individual clinical characteristics. "Treatment recommendation subgroups" were defined by grouping patients on the basis of whether their cATE entailed a high probability of benefit or harm with intervention or neither. The observed absolute risk difference was shown to differ significantly between these subgroups and to correspond qualitatively (directionally) to the predicted cATE, supporting the validity of the model for predicting treatment effect within this trial.

Several points should be noted. First, the derivation and validation procedure for the model of treatment effect would ideally be repeated many times to generate intervals for prediction error, but in this study, the procedure was only performed once. There is a persistent risk of overfitting to data within the analysis by Zampieri and colleagues, and the specific results with respect to fluid infusion rate can only be regarded as hypothesis generating.

Second, to provide a valid basis for individualized treatment decisions in clinical practice, HTE analysis should be fully prespecified in the prospectively planned

trial analysis using features readily available to clinicians in clinical practice. Tools to predict cATE need to be available and integrated within routine clinical care processes.

Third, because these methods are complex and unfamiliar to many practicing clinicians, it may be difficult to apply this information to guide clinical decision making. The methods for conducting and reporting of HTE analysis need to be standardized, and tools to support bedside implementation (such as clinical decision support tools) will likely be required.

Importantly, although the analysis can identify patients most likely to benefit from treatment, it can also increase sample size requirements. Given the difficulties of adequately recruiting patients for ICU trials, it may be difficult to precisely measure cATE in trials.

In the future, treatment decisions in the ICU might be guided by using RCT-derived statistical models to compute the cATE of an intervention for an individual patient. Such personalized information derived from clinical trials may represent a richer and more effective use of data. ∎

References

1. Zampieri FG, Machado FR, Biondi RS, Freitas FGR, Veiga VC, Figueiredo RC, et al.; BaSICS investigators and the BRICNet members. Effect of slower vs faster intravenous fluid bolus rates on mortality in critically ill patients: the BaSICS randomized clinical trial. *JAMA* 2021;326:830–838.
2. Zampieri FG, Damiani LP, Bagshaw SM, Semler MW, Churpek M, Azevedo LCP, et al.; BRICNet. 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: 872–879.
3. Harhay MO, Wagner J, Ratcliffe SJ, Bronheim RS, Gopal A, Green S, et al. Outcomes and statistical power in adult critical care randomized trials. *Am J Respir Crit Care Med* 2014;189:1469–1478.
4. Dianti J, Tisminetzky M, Ferreyro BL, Englesakis M, Del Sorbo L, Sud S, et al. Association of positive end-expiratory pressure and lung recruitment selection strategies with mortality in acute respiratory distress syndrome: a systematic review and network meta-analysis. *Am J Respir Crit Care Med* 2022;205:1300–1310.
5. Angus DC, Chang CH. Heterogeneity of treatment effect: estimating how the effects of interventions vary across individuals. *JAMA* 2021;326: 2312–2313.