

VIEWPOINT

When Should Clinicians Act on Non-Statistically Significant Results From Clinical Trials?

**Paul J. Young, PhD,
MBChB**

Medical Research Institute of New Zealand, Wellington, New Zealand; and Wellington Hospital Intensive Care Unit, Wellington Hospital, Wellington, New Zealand.

**Christopher P.
Nickson, MBChB**

Alfred Hospital Intensive Care Unit, Alfred Hospital, Melbourne, Victoria, Australia; and Monash University School of Public Health and Preventive Medicine, Melbourne, Victoria, Australia.

**Anders Perner, MD,
PhD**

Department of Intensive Care, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark.



Author Audio
Interview

Corresponding

Author: Paul Young, PhD, MBChB, Intensive Care Unit, Wellington Hospital, Private Bag 7902, Wellington South, New Zealand (paul.young@ccdhub.org.nz).

jama.com

Understanding whether the results of a randomized clinical trial (RCT) are clinically actionable is challenging. Reporting standards adopted by *JAMA* and other leading journals lead to relative uniformity of presentation of RCT findings that help simplify critical appraisal.¹ Such uniform reporting also means that the conclusion of the trial may be dichotomized as “positive” or “no difference” based on the statistical significance of the primary outcome. Dichotomization based on the statistical significance of the primary outcome variable reflects the correct, albeit narrow, interpretation of the experiment that the RCT represents. It also reflects decisions made by the investigators in the design of the study and highlights findings in relation to prespecified assumptions. However, there are situations in which a broader appreciation of the results may suggest that non-statistically significant results in the primary outcome of a clinical trial could influence and perhaps change practice. This includes consideration of the outcome in terms of effect size and accompanying CIs, placing the findings from the trial in the context of the totality of the existing relevant evidence.

The Importance of Considering Other Evidence

The Fluid Loading in Abdominal Surgery: Saline vs Hydroxyethyl Starch (FLASH) trial² evaluated protocolized fluid administration with hydroxyethyl starch (HES) in 826 patients at risk of postoperative kidney injury who were undergoing major abdominal surgery, and is an example of an RCT that may change practice despite a non-statistically significant primary outcome finding. In the FLASH trial, the primary composite outcome of death or major postoperative complication occurred in 139 of 389 patients (36%) in the HES group and 125 of 386 patients (32%) in the saline group. Although the *P* value for this difference was .33, the mortality and acute kidney injury rates were numerically higher among patients who received HES. Given the preexisting evidence that HES increases acute kidney injury and mortality in critically ill adults, the data from the FLASH trial suggest that HES is potentially harmful to patients undergoing major surgery and make ongoing use of HES difficult to justify.³

Comparisons of Standard Treatments

Decisions to adopt or de-adopt interventions at the policy level depend not only on the evidence around their effects on clinical outcomes, but also on costs of care. There is value in broad consideration of evidence about clinical and economic consequences. Information about these aspects may exist prior to the RCT, and information from the RCT can further inform these issues. For example, the ANDROMEDA-SHOCK RCT compared a resuscitation strategy targeting normalization of peripheral perfusion with a strategy targeting serum lactate levels in 434 patients

with septic shock.⁴ The primary outcome, mortality at 28 days, occurred in 74 patients (34.9%) in the peripheral perfusion group and 92 patients (43.4%) in the lactate group (hazard ratio, 0.75 [95% CI, 0.55-1.02]; *P* = .06). Despite the non-statistically significant primary outcome, the reported CI suggests that benefit with the peripheral perfusion strategy is more likely than harm. For clinicians who prefer using a peripheral perfusion strategy rather than a strategy using target serum lactate levels prior to this trial, these data provide some support for continuing to use this strategy. Even for other clinicians, adopting a therapeutic strategy based on clinical examination, while acknowledging that further research is warranted to resolve ongoing uncertainty, may be reasonable.

Arguably, in the absence of substantial differences in overall health care costs between these 2 strategies, the statistical significance of the primary outcome may be of less importance because when established treatments are compared, even subtle changes in the probability that one intervention is more effective than another can potentially inform clinical decision-making. However, because the ANDROMEDA-SHOCK RCT only provided information on clinical outcomes, clinicians can only get a sense of what the overall health care costs with the different strategies might be by making educated guesses based on the reported outcomes and their own estimates of the direct treatment costs. To provide genuine assurance that changing practice based on the findings of this trial and similar trials that compare standard treatments will not increase health care costs, formal cost-effectiveness analyses could be conducted.⁵

De-Adoption of Invasive, Expensive Therapies

The case for practice change based on uncertain evidence is perhaps strongest when it means that therapies for which there is strong evidence that they are relatively more invasive, labor-intensive, or more expensive than alternatives can be de-adopted. Yet, in most cases, strong evidence of this type is usually lacking. Because some therapies can be more invasive and expensive initially, but ultimately less expensive overall, a nuanced interpretation of uncertain evidence in relation to clinical effects needs to be accompanied by a nuanced interpretation in relation to overall health care costs.

In the Coronary Angiography after Cardiac Arrest (COACT) trial,⁶ 552 patients with out-of-hospital cardiac arrest with no signs of ST-elevation myocardial infarction were assigned to immediate or delayed angiography. At 90 days, 176 patients (64.5%) in the immediate angiography group and 6265 patients (67.2%) in the delayed angiography group were alive. The differences in these mortality findings (the primary outcome) were not statistically significant, and the hazard ratio of 0.90 (95% CI, 0.67-1.20) is compatible with both relevant harm and benefit of

immediate angiography. However, cardiologists who had previously undertaken immediate coronary angiograms in this patient group might be reassured by the point estimate for the mortality treatment effect favoring delayed angiography and by the absence of demonstrable harm with this approach in any of the secondary outcomes in the trial.

Moreover, because decisions about whether to undertake immediate angiography often need to be made outside of normal work hours and may require mobilization of a specific team and hospital resources in general, these data suggest delayed angiography is likely to reduce overall health care costs. Delaying angiography may have benefits for other patients who require emergency care because outside-hours hospital resources are often limited. The data are not definitive, but a mortality estimate favoring delayed angiography decreases the probability that immediate coronary angiography reduces mortality and, importantly, there is a strong argument against demanding definitive data before de-adopting a labor-intensive intervention that has not been shown in an RCT to benefit patients. While acknowledging that there is uncertainty about overall costs, doing "less," particularly when resources are limited, simplifies care, which may have collateral benefits.

Comparisons of Standard Treatments

When data from an RCT, taken in the context of the overall evidence-base, suggest higher cost care may be harmful or that more a more simple established treatment approach may be beneficial, clinicians may act on these findings, even when the difference in the primary outcome is not statistically significant. Conversely, when non-statistically significant differences in the primary outcome from a comparative-effectiveness RCT favor an established therapy that is expensive or complex, deciding whether findings are actionable is more difficult. The burden of proof required to do "more" should probably be higher than to do "less," particularly if doing "more" is likely to increase overall health care costs. A recent post hoc analysis of a previously published RCT that compared the association between venovenous extracorporeal membrane oxygenation vs conventional mechanical ventilation and mortality among patients with severe acute respiratory distress syndrome illustrates the potential usefulness of bayesian analysis in aiding clinical decision-making in this situation, particularly when there is a spectrum of beliefs in the clinical community about the prior probability that the more expensive treatment is beneficial.⁷ Because higher up-front costs may ultimately be accompanied by downstream cost savings, a nuanced interpretation of likely costs that account for such uncertainties is needed.

Evaluation of New Therapies and Technologies

Because the addition of therapies adds complexity that may distract attention from other aspects of care and may increase the risk of errors and interactions, the threshold to change practice based on the findings of an RCT that investigates a new therapy should probably be higher than is applied when evaluating an RCT that compares 2 existing therapies. Methodological rigor, the biological plausibility of the prespecified treatment effect size, and statistical and clinical significance of the primary outcome are all crucial considerations when evaluating an RCT that investigates a new therapy. Another relevant consideration when deciding whether to implement the results of an RCT evaluating a new drug (or other intervention, such as a new device) is the possibility of previously unrecognized adverse events of the drug (or device) occurring with implementation of the trial findings into routine care.

The adoption of new technologies deserves special consideration. Various drivers and incentives contribute to the general conviction that a more technological approach is intrinsically better than a less technological one. In the case of new technologies (such as medical devices), even greater skepticism is required by clinicians given that the regulatory burden for technologies is minimal compared with pharmacological therapeutics and that technologies may be invasive. The use of intravascular microaxial left ventricular assist devices for management of cardiogenic shock following acute myocardial infarction has increased despite the absence of RCT evidence that these devices improve patient-centered outcomes. Concerningly, a 2020 observational study that included 1680 propensity-matched pairs found that the use of these devices, compared with the use of an intra-aortic balloon pump, was associated with increased in-hospital mortality (45% vs 34.1%) and major bleeding (31.3% vs 16.0%).⁸

Conclusions

Whether the findings of an RCT should inform clinical practice depends not only on whether the primary outcome achieved the prespecified criteria for declaring statistical significance, but also on a broader understanding of the overall likelihood that one treatment represents a better option for patients than the other. This understanding comes from consideration of trial methodology, other evidence, comparative cost-effectiveness, invasiveness, and labor-intensiveness and from recognition that the burden of proof required for practice change may vary depending on the nature of the therapy being investigated.

ARTICLE INFORMATION

Published Online: May 8, 2020.
doi:10.1001/jama.2020.3508

Conflict of Interest Disclosures: Dr Young reported receiving a research fellowship from Health Research Council of New Zealand during the conduct of the study. Dr Perner reported receiving grants from Novo Nordisk Foundation outside the submitted work. No other disclosures were reported.

REFERENCES

- Bauchner H, Golub RM, Fontanarosa PB. Reporting and interpretation of randomized clinical trials. *JAMA*. 2019;322(8):732-735.
- Futier E, Garot M, Godet T, et al; FLASH Trial Group. Effect of hydroxyethyl starch vs saline for volume replacement therapy on death or postoperative complications among high-risk patients undergoing major abdominal surgery. *JAMA*. 2020;323(3):225-236.
- Zampieri FG, Cavalcanti AB. Hydroxyethyl starch for fluid replacement therapy in high-risk surgical patients. *JAMA*. 2020;323(3):217-218.
- Hernández G, Ospina-Tascón GA, Damiani LP, et al. Effect of a resuscitation strategy targeting peripheral perfusion status vs serum lactate levels on 28-day mortality among patients with septic shock. *JAMA*. 2019;321(7):654-664.
- Sanders GD, Maciejewski ML, Basu A. Overview of cost-effectiveness analysis. *JAMA*. 2019;321(14):1400-1401. doi:10.1001/jama.2019.1265
- Lemkes JS, Janssens GN, van der Hoeven NW, et al. Coronary angiography after cardiac arrest without ST-segment elevation. *N Engl J Med*. 2019;380(15):1397-1407. doi:10.1056/NEJMoa1816897
- Goligher EC, Tomlinson G, Hajage D, et al. Extracorporeal membrane oxygenation for severe acute respiratory distress syndrome and posterior probability of mortality benefit in a post hoc bayesian analysis of a randomized clinical trial. *JAMA*. 2018;320(21):2251-2259.
- Dhruva SS, Ross JS, Mortazavi BJ, et al. Association of use of an intravascular microaxial left ventricular assist device vs intra-aortic balloon pump with in-hospital mortality and major bleeding among patients with acute myocardial infarction complicated by cardiogenic shock. *JAMA*. Published online February 10, 2020. doi:10.1001/jama.2020.0254