

VIEWPOINT

David L. DeMets, PhD
Department of
Biostatistics and
Medical Informatics,
University of
Wisconsin–Madison,
Madison.

Thomas Cook, PhD
Department of
Biostatistics and
Medical Informatics,
University of
Wisconsin–Madison,
Madison.

Challenges of Non-Intention-to-Treat Analyses

Over the past 5 decades, the randomized clinical trial has become the gold standard for evaluation of the risks and benefits of new interventions, including drugs, medical devices, and surgical procedures.¹

To justify the use of randomization, it is important to note that in a nonrandomized study comparing 2 interventions, a small P value (of which $P < .05$ is generally considered statistically significant) for a statistical comparison between groups can be due to 1 of 3 sources: chance, causation, or confounding. Because randomized assignments cannot be associated with participant characteristics, effective randomization eliminates the third possibility, enabling direct assessment of potential causal relationships provided that the study is designed, conducted, and analyzed properly.

Proper trial conduct and analysis include ensuring complete follow-up and unbiased ascertainment of the outcomes of interest among all randomized participants. The exclusion of participants or outcomes from the analysis, because outcomes either are missing or are deliberately omitted from the analysis, reintroduces the third possibility—that observed differences are due to confounding. The intention-to-treat (ITT) principle formalizes this idea and that, up to the desired level of statistical significance (eg, $P < .05$), observed differences between groups are due to the causal effect of assignment to the intervention. The P value or corresponding confidence interval already accounts for any chance baseline covariate imbalances.

In trials for which all randomized participants are fully adherent to their assigned intervention, there is agreement that use of ITT for the analysis accurately represents the effects. However, when not all participants are adherent to the assigned intervention, investigators may be tempted to try to correct for this by use of “as-treated” analyses—analyses that reassign participants to the intervention that they actually received; “per-protocol” analyses—those that exclude participants with adherence below a threshold; or “on-treatment” analyses—analyses that terminate follow-up at or shortly after discontinuation of the intervention. Because each of these approaches violates the ITT principle, each is subject to confounding, and there is no reason to believe that such approaches can correct the “problem.” The underlying problem is in the conduct of the trial, not in the analysis. It is impossible to reliably recover the causal effect of full adherence from a trial in which full adherence is not achieved.

Because the degree of potential confounding when outcomes are missing cannot be determined (the required data are by definition missing) the only statistically sound approach is sensitivity analyses, which involve assessing the robustness of the result to a range of plausible mechanisms responsible for the missing data.

May et al² have described problems arising from failure to follow the ITT principle and noted implications for study design and conduct. For example, ITT requires that,

to the extent possible, all participants should be followed up for the full duration of the study regardless of their adherence to assigned treatment; “off treatment” should not imply “off study.” In placebo-controlled trials in which it is likely that nonadherence to assigned treatment “dilutes” the effect of the treatment, study power may be reduced unless sample sizes are increased. However, in active-controlled trials, nonadherence may not always result in such dilution; for example, nonadherence in the control group may actually increase the observed difference between groups.

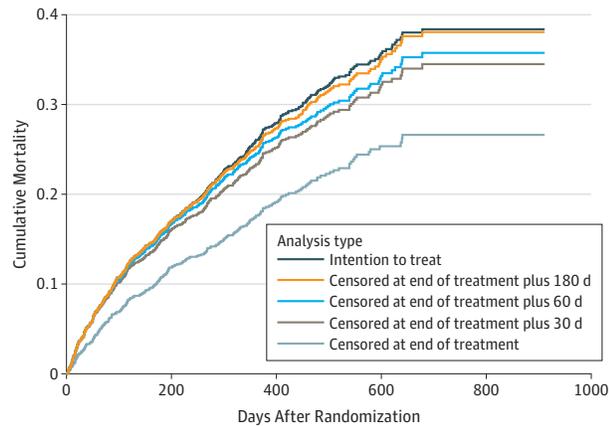
Stratifying the analysis by predefined baseline covariates is a standard analysis, but the stratification must not depend on outcomes or experiences of a participant that occur after randomization. A few examples illustrate the problem. One of the earliest randomized clinical trials in cardiology, the Coronary Drug Project, which compared 5 cholesterol-lowering drugs with a placebo group, followed the ITT principle for its primary analyses but illustrates how analysis based on adherence can yield results that are clearly not interpretable.³ Participants in the placebo group were divided into those who took at least 80% of their assigned medication and those who did not. Analysis of those 2 groups strongly indicated that the former had better mortality results than the latter, which is highly unlikely to be due to a causal effect of placebo. This and other similar analyses clearly demonstrated that adherence is associated with underlying risk and should be considered an outcome rather than a predictor.

The Anturane Reinfarction Trial⁴ was a randomized clinical trial that compared the drug sulfapyrazone with placebo in a population of patients with recent myocardial infarction. Eligibility for analyses, predefined in the protocol, required that participants adhere to their assigned intervention for at least 7 days; however, the application of this rule varied by how eligibility was assessed or who made the assessment.² In this case, classification of eligibility was done after the trial ended, when unblinding or partial unblinding was possible. Furthermore, exclusions based on postrandomization measurements reintroduced confounding and thereby subverted the randomization. In addition, not all events that occurred in the follow-up period were included. Including or not including all events and all patients in their randomized groups resulted in a large shift in the P values from significant (ie, $P < .05$) to nonsignificant.

Similarly, in the APPROVE trial,⁵ which was a trial of rofecoxib vs placebo for cancer prevention, the idea that “off treatment” implies “off study” was written into the protocol; follow-up ceased 14 days after participants withdrew from their assigned intervention, and cardiovascular outcomes occurring 14 days after participants stopped were unreported. Subsequently, a full year of additional follow-up was conducted and vital status was obtained for more than 95% of randomized participants. The original “on-treatment” analysis suggested an

Corresponding Author: David L. DeMets, PhD, Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison, 610 Walnut St, WARF Building, Room 201, Madison, WI 53726 (demets@biostat.wisc.edu).

Figure. Cumulative Mortality Assessed by Intention-to-Treat Analysis and by Variations of “On-Treatment” Analyses for the EVEREST Trial



initial 18-month “honeymoon” during which the mortality risk due to long-term use of rofecoxib was the same as placebo. The subsequent ITT reanalysis suggested that the mortality risk began shortly after randomization with no honeymoon.

While the on-treatment analysis in the APPROVE trial appeared to underestimate the harmful effects of rofecoxib during the initial 18-month treatment period, Yang et al⁶ described an on-treatment analysis from the SAVOR trial⁷ that apparently overestimated the harmful effect of saxagliptin. Patients randomized to active treatment experienced better glycemic control than those in the placebo group such that more placebo-treated patients discontinued their assigned treatment and received rescue treatment. As a result, the on-treatment analysis excluded more deaths from the placebo group than from the intervention group, leading to an overestimate of the relative risk.⁶

The EVEREST trial⁸ assessed the short- and long-term effects of tolvaptan relative to placebo among patients hospitalized for acute decompensated heart failure. All-cause mortality was assessed for both superiority and noninferiority, the latter to ensure that acute, short-term benefit was not offset by increased mortality. Given the widespread perception that an on-treatment analysis may be preferred for tests of noninferiority, such an analysis was considered but subsequently rejected.

The Figure shows cumulative mortality for the EVEREST trial assessed by ITT and several variations of on-treatment analyses in which death times were censored at the end of treatment and end of treat-

ment plus 30, 60, or 180 days for the placebo group. Assuming that the placebo has no effect on mortality, the ITT estimate should represent the true mortality risk, whereas each variation of end of treatment shows lower mortality. The most plausible explanation of this result is that censoring is informative—that is, associated with mortality—and the on-treatment estimates of cumulative mortality are biased. More important, on-treatment comparisons of mortality between groups are confounded by the associations among assigned treatment, treatment discontinuation, and mortality and cannot be expected to represent causal effects of treatment.

Noninferiority trials are designed to demonstrate that another intervention effect is within some margin of a standard intervention, either on an absolute or relative scale, and thus acceptable in practice as an alternative to the standard intervention.¹ One rationale often given for non-ITT analyses is that poor adherence to the randomized nonstandard intervention reduces the observed difference between groups, biasing the comparisons, and that in this setting, an on-treatment analysis provides a better estimate of effect than does the ITT analysis. However, this may not be true in active-controlled trials, which constitute the majority of noninferiority trials. Non-ITT analyses can reintroduce confounding and thereby introduce biases of unknown magnitude and direction.²

The best policy in both superiority and noninferiority trials is to conduct the trial with optimum adherence to the assigned intervention. Except in rare circumstances, trials evaluate intervention strategies with less than perfect adherence. With less than optimal adherence, no satisfactory alternative to ITT exists.

Conclusions

Randomization is essential for objectively establishing causal associations between new interventions and outcomes. However, using randomization, a causal link with outcomes can be established only between assigned interventions rather than received interventions, and only then if there is complete follow-up and the ITT principle is applied. Most simplistic alternatives to ITT analysis reintroduce the confounding that randomization was intended to eliminate. It is unrealistic to imagine that simple statistical maneuvers can transform the trial that was conducted into the trial that ideally could have been conducted. To the greatest extent possible, all outcomes should be ascertained for all randomized participants, the primary analysis should be conducted according to the ITT principle, and the results should be interpreted using sound informed scientific judgement—for which there is no statistical substitute.

ARTICLE INFORMATION

Published Online: December 17, 2018.
doi:10.1001/jama.2018.19192

Conflict of Interest Disclosures: Dr DeMets reported receipt of grants from the National Institutes of Health and personal fees from Frontier Science, D. L. DeMets Consulting LLC, Actelion, the Duke Clinical Research Institute, the Population Health Research Institute [Canada], AstraZeneca, GlaxoSmithKline, DalCor, LivaNova, Boston Scientific, Medtronic, the Patient-Centered Outcomes Research Institute, Amgen, Merck, and Sanofi. No other disclosures were reported.

REFERENCES

1. Friedman L, Furberg C, DeMets D, et al. *Fundamentals of Clinical Trials*. 5th ed. New York, NY: Springer Science + Business Media LLC; 2015.
2. May GS, DeMets DL, Friedman LM, et al. The randomized clinical trial: bias in analysis. *Circulation*. 1981;64(4):669-673.
3. Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project. *N Engl J Med*. 1980;303(18):1038-1041.
4. Anturane Reinfarction Trial Research Group. The Anturane Reinfarction Trial: reevaluation of outcome. *N Engl J Med*. 1982;306(16):1005-1008.
5. Baron JA, Sandler RS, Bresalier RS, et al. Cardiovascular events associated with rofecoxib: final analysis of the APPROVE trial. *Lancet*. 2008;372(9651):1756-1764.
6. Yang F, Wittes J, Pitt B. Beware of on-treatment safety analyses [published online November 16, 2018]. *Clin Trials*. doi:10.1177/1740774518812774
7. Scirica BM, Bhatt DL, Braunwald E, et al. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. *N Engl J Med*. 2013;369(14):1317-1326.
8. Konstam MA, Gheorghiadu M, Burnett JC Jr, et al. Effects of oral tolvaptan in patients hospitalized for worsening heart failure: the EVEREST outcome trial. *JAMA*. 2007;297(12):1319-1331.