NEJM Evidence

Published July 26, 2022
NEJM Evid 2022; 1 (8)
DOI: 10.1056/EVIDctw2200018

CLINICAL TRIALS WORKSHOP | INTERFACE SERIES

Improving Clinical Trial Interpretation with ACCEPT Analyses

Michelle N. Clements, Ph.D.,[1] Ian R. White, Ph.D.,[1] Andrew J. Copas, Ph.D.,[1] Victoria Cornelius, Ph.D.,[2] Suzie Cro, Ph.D.,[2] David T. Dunn, Ph.D.,[1] Matteo Quartagno, Ph.D.,[1] Rebecca M. Turner, Ph.D.,[1] Conor D. Tweed, M.D.,[1] and A. Sarah Walker, Ph.D.[1]

C. Corey Hardin, M.D., Ph.D.,
Interface Series Editor

Introduction

Effective decision-making from randomized controlled clinical trials relies on robust interpretation of the numerical results. However, the language we use to describe clinical trials can cause confusion both in trial design and in comparing results across trials. ACceptability Curve Estimation using Probability above Threshold (ACCEPT) aids comparison between trials (even when of different designs) by harmonizing reporting of results, acknowledging that different interpretations of the results may be valid in different situations, and moving the focus from comparison with a prespecified value to interpretation of the trial data. ACCEPT can be applied to historical trials or can be incorporated into statistical analysis plans for future analyses. An online tool enables ACCEPT on up to three trials simultaneously.

The classic superiority trial aims to generate robust evidence that a new treatment is better than placebo. Active controls are used to assess superiority of a new treatment when the use of placebo is unethical, such as when an effective treatment is available. Noninferiority trials, aiming to show that the new treatment is not appreciably worse than the control, are often used to evaluate new drugs and interventions with expected efficacy similar to that of standard therapy but secondary advantages, such as less toxicity, ease of implementation, benefits in particular subgroups only, or lower cost. Similarly, equivalence trials aim to show that a new treatment is unlikely to differ appreciably from control in either direction, and supersuperiority trials aim to show evidence of a new treatment being better than control by at least a specified value.

The specification of trial type (e.g., as superiority or noninferiority) is important to enable assessment of whether a trial has met its aims. However, the different terms are in themselves confusing. Additionally, seemingly paradoxical situations can arise when comparing across trials, such as trial type differing depending on which treatment is assigned as “control” and trials with similar numerical results reaching different conclusions.

All clinical trial types can be linked by the prespecified “unacceptable value” to which the 95% confidence interval (CI) limits of the estimate of the difference between treatments are compared. Specification of trial type is equivalent to prespecification of the unacceptable

The author affiliations are listed at the end of the article.

Dr. Clements can be reached at michelle.clements@ucl.ac.uk or at MRC Clinical Trials Unit at UCL, 90 High Holborn, 2nd Floor, London WC1V 6LJ, United Kingdom.

value: zero (or one for a relative effect measure) in superiority trials, the noninferiority margin (less than zero) in noninferiority trials, and greater than zero in supersuperiority trials.

Comparison with the prespecified unacceptable value is an appropriate part of trial interpretation but leads to a binary conclusion of "trial aim met" or "trial aim not met." Outside drug regulation, binary conclusions are widely viewed as problematic,[1] because evidence should not be reduced to a single threshold but should be considered in context with other factors, such as the point estimate and CI.[2] Interpretation of noninferiority trial results also suffers from added complexity around whether the preset noninferiority margin was justified or is relevant for settings outside the trial. Importantly, stakeholders such as clinicians, patients, and policy makers may have differing but equally valid unacceptable difference values depending on the relative importance placed on secondary factors, such as cost and toxicity.

We advocate the wider use of ACCEPT as secondary analyses in clinical trials. We illustrate this using two HIV trials, Europe — Africa Research Network for Evaluation of Second-Line Therapy (EARNEST)[3] and SECOND-LINE,[4] which had similar quantitative results but from which different conclusions were drawn.[5] We demonstrate how alternative presentation of results could aid better comparison and integration of their findings. We present the trials together, but imagine ACCEPT being presented in each trial results paper separately as secondary analyses. For ease throughout, we measure differences as treatment minus control for a favorable outcome, so that positive values indicate higher efficacy in the tested treatment.

EARNEST and SECOND-LINE investigated raltegravir as a second-line therapy for HIV in comparison with standard therapy of nucleoside reverse transcriptase inhibitors (NRTI). EARNEST was carried out in low- and middle-income countries. It was prespecified as a superiority trial because raltegravir was more expensive than NRTI; therefore, it was thought that clear benefit would have to be shown for implementation. The prespecified unacceptable value was consequently zero.

SECOND-LINE was carried out predominantly in high-income countries. It was prespecified as a noninferiority trial because raltegravir was considered to have a better toxicity profile than NRTI; implementation was therefore considered to be worthwhile with similar efficacy. The prespecified unacceptable value was the noninferiority margin of −12% on the risk difference scale.

The original analysis of EARNEST compared the lower limit of the 95% CI of the difference between treatments (−2.4%) with the unacceptable value of zero, drawing the conclusion of "superiority not shown," and implementation was not recommended. Analysis of SECOND-LINE compared the lower limit of the 95% CI (−4.7%) with the unacceptable value of −12%, drawing the conclusion of noninferiority, and implementation was recommended. The question then arises as to how two trials with numerically similar results can reach opposing conclusions regarding implementation.

Differing, valid opinions on the unacceptable differences values (0% in EARNEST and −12% in SECOND-LINE), driven in part by different emphases on secondary benefits, led to the selection of different trial types and the resulting seemingly opposing recommendations. Interpretation through ACCEPT, including both graphs and tables, would have helped to clarify this paradox, enabling more nuanced interpretation of the results.

ACCEPT uses the primary analysis from a trial to plot the probability of the true difference between treatments being above an "acceptability threshold" for a range of possible threshold values (Fig. 1). ACCEPT can be presented for all trial types and outcomes. ACCEPT has been used only sporadically in clinical trials[6-12] with no consistent naming. ACCEPT is similar to the cost-effectiveness acceptability curves widely used in health economics, in which the weight of evidence, rather than binary conclusions, is a more widely accepted paradigm.

ACCEPT output is best presented in a graph with associated tables. A graph shows a continuous range of acceptability thresholds in which greater uncertainty around point estimates (with larger associated CIs) is reflected in a shallower slope. Additional tables present selected acceptability thresholds or the probability that the true value is between selected thresholds. To enable comparison of ACCEPT between trials, tables should include acceptability values for the unacceptable difference (specified in the trial design), zero, a reasonable range of potential alternative unacceptable values, and acceptability thresholds for the 2.5th, 50th, and 97.5th percentile acceptability values.

ACCEPT can be implemented using Bayesian analysis, which provides direct estimation of the probability that one

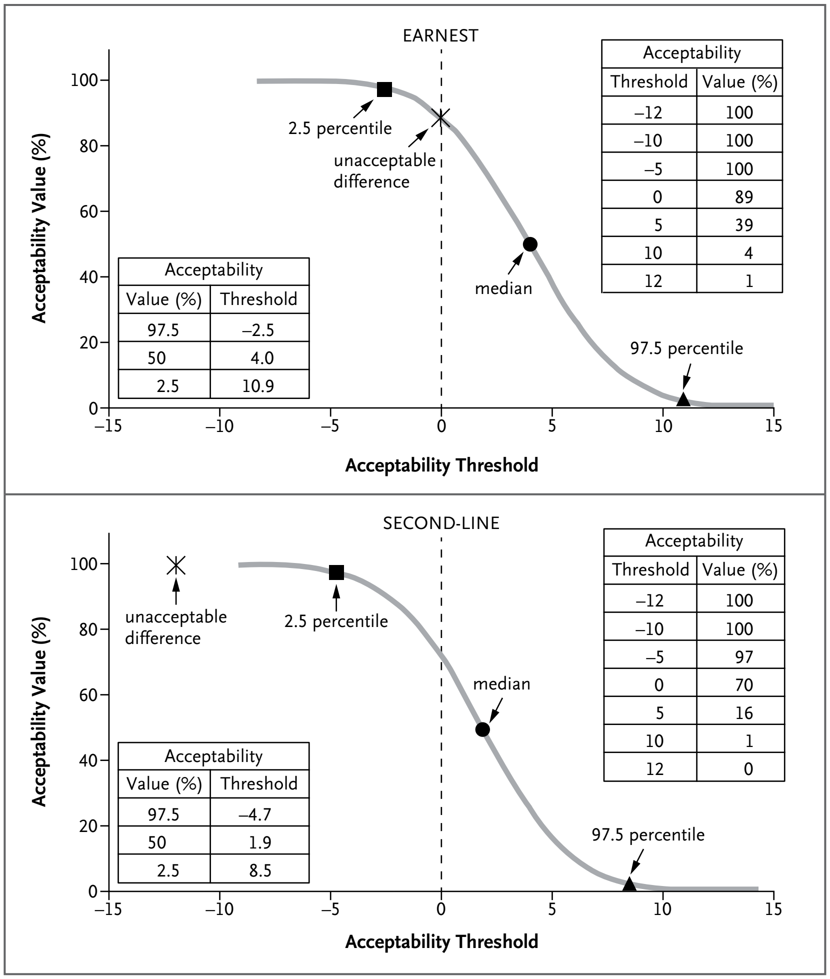

Figure 1. ACceptability Curve Estimation using Probability above Threshold (ACCEPT) Curves and Tables for EARNEST and SECOND-LINE Trials

An acceptability value is the probability that the true treatment difference is at least the acceptability threshold. Positive values indicate that raltegravir is better than nucleoside reverse transcriptase inhibitors (NRTI). For example, the probability that the true difference between treatments was at least 0 (i.e., that raltegravir is better than NRTI) was 89% in EARNEST and 70% in SECOND-LINE. Median estimates, 95% Bayesian credible intervals from models, and prespecified unacceptable differences are marked.

treatment is better or worse than another when the prior belief of the difference between treatments is added to the analysis. The degree of prior belief, termed priors, is based on existing data and/or expert opinion. Priors are specified as a distribution over the possible values that the difference can take, with noninformative priors essentially being a flat distribution and strongly informative priors being very concentrated around the area of highest belief. ACCEPT within a Bayesian framework is more consistent with the overall philosophy than within a frequentist framework. However, frequentist analysis using confidence curves[13,14] is expected to give very similar results to Bayesian analysis with uninformative priors. In the frequentist framework, the acceptability value is the one-sided P value for the treatment effect exceeding the acceptability threshold. An online tool enables ACCEPT for up to three trials simultaneously using summary information from frequentist or Bayesian analysis (https://egon.stats.ucl.ac.uk/projects/ACCEPT/). Further details of analyses, including statistical code, are available in the Supplementary Appendix.

Using ACCEPT for trial reporting, EARNEST results would still conclude that superiority was not shown but could also include a statement such as “ACCEPT suggested that there was an 89% probability that the true treatment difference was greater than zero (i.e., that raltegravir was better than NRTI) and 100% probability that the true treatment difference was above −5, equivalent to a 0% probability that raltegravir was worse than NRTI by at least 5 percentage points. There was a 39% probability raltegravir was better than NRTI by at least 5 percentage points."

Similarly, reporting of SECOND-LINE results with ACCEPT would still conclude that noninferiority was shown but could also add a statement asserting that "ACCEPT suggested that there was a 70% probability of the true treatment difference being greater than zero, a 97% probability of the true treatment difference being above −5 percentage points, and a 16% probability that raltegravir was better than NRTI by at least 5 percentage points."

Using ACCEPT, stakeholders requiring clear benefit of raltegravir for implementation could use acceptability thresholds of zero and above, concluding that the probability of raltegravir being better than NRTI was 89% in EARNEST and 70% in SECOND-LINE, but the probability of being more than 5 percentage points better was much lower, at 39% in EARNEST and 16% in SECOND-LINE. Other stakeholders more focused on other secondary benefits of raltegravir, such as lower toxicity, could use acceptability thresholds of zero and below, concluding at least 97% probability of the true treatment difference being more than −5 percentage points in either trial. This allows better comparison across trials than does the primary analysis alone.

Interpretation through ACCEPT has three main strengths. First, it enables comparison between trials and trial types by harmonizing reporting of results; the use of probabilities is straightforward and widely understood and reflects the uncertainty around the point estimate. Second, presentation of ACCEPT acknowledges that different acceptability thresholds may exist in different situations. ACCEPT allows clinicians, policy makers, and patients to make informed decisions on the basis of their setting and individual circumstances if they feel the original choice of unacceptable difference is not appropriate for their context. Third, ACCEPT moves the focus from comparison with the prespecified unacceptable value to interpretation of the trial data. This may be especially useful for trials prespecified as noninferiority to reduce focus on the selected unacceptable value and in situations in which restricted sample size reduces power, such as subgroup analysis and uncommon conditions. For subgroup analysis, ACCEPT can be presented separately for each subgroup using output from either models run separately for each subgroup or a single model where an interaction between subgroup and trial arm is fitted.

Use of ACCEPT does not remove all of the concerns that can arise with noninferiority trials, which are caused by the unacceptable difference being less than zero. Noninferiority trials cannot always provide assurance that the new treatment has a clinically relevant effect (greater than zero) relative to placebo, and therefore, it is important to carefully assess evidence about how much better the control treatment is than placebo when selecting the prespecified unacceptable difference/noninferiority margin to prevent bio-creep. Nonadherence in clinical trials may bias the estimate of treatment differences toward zero, especially if treatment crossover occurs, meaning that conclusions of noninferiority may be more likely with substantial nonadherence. Analysis of different trial populations (per protocol and intention to treat) or statistical adjustment must still be used to allow for this, but ACCEPT can help improve interpretation when comparing across different populations within a trial.

ACCEPT can be applied to historical trials or incorporated into statistical analysis plans for future analyses. ACCEPT has been advocated previously for use in clinical trials reporting, but its use has not become widespread, perhaps because of a lack of common language to discuss the analyses. Increased use of a variety of different trial designs means the time is right for unified design and interpretation through ACCEPT.

Disclosures

Drs. Clements, White, Copas, Dunn, Quartagno, Turner, Tweed, and Walker are supported by core support from the Medical Research Council (MRC) UK to the MRC Clinical Trials Unit at UCL (MC_UU_12023/22 and MC_UU_00004/09). Dr. Walker is a National Institute for Health Research (NIHR) Senior Investigator. Dr. Cro is supported by an NIHR advanced fellowship (NIHR300593).

Author disclosures and other supplementary materials are available at evidence.nejm.org.

We thank Andrew Nunn, Di Gibb, Hanif Esmail, Julia Bielicki, and Mike Sharland for thoughtful comments that greatly improved the manuscript.

Author Affiliations

[1] MRC Clinical Trials Unit at UCL, London

[2] Imperial Clinical Trials Unit, Imperial College London, London

References

1. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature 2019;567:305–307. DOI: 10.1038/d41586-019-00857-9.
2. Harrington D, D'Agostino Sr RB, Gatsonis C, et al. New guidelines for statistical reporting in the *Journal*. N Engl J Med 2019;381:285–286. DOI: 10.1056/NEJMe1906559.
3. Paton NI, Kityo C, Hoppe A, et al. Assessment of second- line antiretroviral regimens for HIV therapy in Africa. N Engl J Med 2014;371:234–247. DOI: 10.1056/NEJMoa1311274.
4. Boyd MA, Kumarasamy N, Moore CL, et al. Ritonavir-boosted lopinavir plus nucleoside or nucleotide reverse transcriptase inhibitors versus ritonavir-boosted lopinavir plus raltegravir for treatment of HIV-1 infection in adults with virological failure of a standard first-line ART regimen (SECOND-LINE): a randomised, open-label, non-inferiority study. Lancet 2013;381:2091–2099. DOI: 10.1016/S0140-6736(13)61164-2.
5. Dunn DT, Copas AJ, Brocklehurst P. Superiority and non-inferiority: two sides of the same coin? Trials 2018;19:499. DOI: 10.1186/s13063-018-2885-z.
6. Simon R. Bayesian design and analysis of active control clinical trials. Biometrics 1999;55:484–487. DOI: 10.1111/j.0006-341x.1999.00484.x.
7. Fluehler H, Grieve AP, Mandallaz D, Mau J, Moser HA. Bayesian approach to bioequivalence assessment: an example. J Pharm Sci 1983;72:1178–1181. DOI: 10.1002/jps.2600721018.
8. Laptook AR, Shankaran S, Tyson JE, et al.. Effect of therapeutic hypothermia initiated after 6 hours of age on death or disability among newborns with hypoxic-ischemic encephalopathy: a randomized clinical trial [published correction appears in JAMA 2018;319:1051]. JAMA 2017;318:1550–1560. DOI: 10.1001/jama.2017.14972.
9. Ryan EG, Harrison EM, Pearse RM, Gates S. Perioperative haemodynamic therapy for major gastrointestinal surgery: the effect of a Bayesian approach to interpreting the findings of a randomised controlled trial. BMJ Open 2019;9:e024256. DOI: 10.1136/bmjopen-2018-024256.
10. Ghosh P, Nathoo F, Gönen M, Tiwari RC. Assessing noninferiority in a three-arm trial using the Bayesian approach. Stat Med 2011;30:1795–1808. DOI: 10.1002/sim.4244.
11. Nunn AJ, Phillips PPJ, Meredith SK, et al. A trial of a shorter regimen for rifampin-resistant tuberculosis. N Engl J Med 2019;380:1201–1213. DOI: 10.1056/NEJMoa1811867.
12. Li H-K, Rombach I, Zambellas R, et al. Oral versus intravenous antibiotics for bone and joint infection. N Engl J Med 2019;380:425–436. DOI: 10.1056/NEJMoa1710926.
13. Shakespeare TP, Gebski VJ, Veness MJ, Simes J. Improving interpretation of clinical studies by use of confidence levels, clinical significance curves, and risk-benefit contours. Lancet 2001;357:1349–1353. DOI: 10.1016/S0140-6736(00)04522-0.
14. Bender R, Berg G, Zeeb H. Tutorial: using confidence curves in medical research. Biom J 2005;47:237–247. DOI: 10.1002/bimj.200410104.