Statistical questions on the electronic cigarettes trial

Drew Payne DO, Shengping Yang PhD

We are planning a clinical trial to evaluate whether electronic cigarettes (ECs) can be used to stop smoking. Ideally, ECs have a better effect than the currently used standard treatment of nicotine patches (NP). We are interested in comments from a statistical perspective on trial design and data analysis. Here are our questions:

1. What would our approach be in a superiority study?

The primary goal of a clinical study is to answer a research question(s) of interest to the investigator(s). Having a clear and well-thought research question allows the investigator(s) to effectively perform literature searches, set up research specific aims, and generate hypotheses. In addition, research questions also determine which trial design and data analysis methods are appropriate.

In the EC trial you are proposing, the aim is to compare EC treatment with the standard NP treatment to stop smoking. Before designing the study, it is important to do some literature searches. What is the efficacy of the NP treatment? Note that if the efficacy of NP treatment is low, and EC treatment is expected to have a higher efficacy, for example, based on pilot studies, then a superiority trial might be a good choice.

Once a research question has been asked, a hypothesis will be generated to provide a tentative and testable answer to it. For a superiority trial, the associated hypothesis is:

$$H_{null}: C = T$$
$$H_{alternative}: C \neq T$$

where C is the control and T is the new treatment.

Note that for clinical trials that have placebo groups, superiority trials are always used. Since a placebo is not supposed to have a true effect on the outcome, it makes sense that the new treatment is clearly superior to a placebo.

One clear advantage of using a superiority trial is that rejection of a null hypothesis automatically establishes assay sensitivity, which is defined as the ability of a clinical trial to distinguish an effective treatment from a less effective or ineffective treatment. This is, however, not the case for non-inferiority trials, although non-inferiority trials are becoming increasingly popular recently.

2. What would our approach be in a non-inferiority study?

A non-inferiority trial becomes a natural choice when the control group has high efficacy, and thus it is difficult to develop a new treatment that is superior to the control. At other times, a non-inferiority trial can be used when it is not ethical to include a placebo group. However, non-inferiority trials have inherent limitations due to the way the null and alternative hypotheses are generated.

The null and the alternative hypotheses of a non-inferiority trial are:

$$H_{null}: C - T \geq M$$
$$H_{alternative}: C - T < M$$

where M is non-inferiority margin.

Specifically, the null hypothesis is that the new treatment is inferior to the control by at least M, and the alternative hypothesis is that the new treatment is not inferior or is slightly inferior but by no more then M. As a consequence, poor execution of a non-inferiority trial, including protocol violation, patient dropout, and/or misclassification of the endpoint, can result in bias toward non-inferiority of the new treatment. To avoid

such obvious limitations, it is important to enforce very strict trial quality control throughout the study.

There are other limitations of a non-inferiority trial, which are also more or less associated with the way the hypothesis is generated, including:

a) Difficulty in setting up the non-inferiority margin.

b) No proof of assay sensitivity.

c) Possible violation of the constancy assumption (details on a non-inferiority trial can be found in the last October issue of the Southwest Respiratory and Critical Care Chronicles).

3. How would we approach the statistical power of the study?

The statistical power for testing the primary outcome is often set at 80% or 90%. The significance level is often set at 0.05. In other words, the null hypothesis will be rejected if the p value is less than 0.05.

It is possible that there are two primary outcomes in a study; to avoid inflated type I error caused by multiple testing, the significance level of the two associated tests can be conservatively set at 0.025 (0.05/2). Meanwhile the statistical power will be kept at the same 80% or 90%.

A number of statistical software can be used for calculating sample size/statistical power, e.g., PASS (NCSS, LLC. Kaysville, UT), EAST (Cytel, Cambridge, MA), SAS (SAS Institute, Inc., Cary, NC), and R (www.R-project.org). (Since sample size calculation is not the main focus of this article; it will not be discussed here.)

4. How would we approach the statistical power with consideration to secondary outcomes?

Besides the primary outcome, there can be many secondary outcomes that the investigator(s) might be interested in. In general, the study sample size/power is calculated based on the primary outcome only, and thus the power for comparing the secondary outcomes might be less or more than 80% (or 90%) depending on the nature of the secondary outcome measurements.

5. What data analysis would best fit this study?

In general, an "intent-to-treat" analysis is recommended for a superiority trial. All participants who are randomized will be included in the analysis and will be assigned to the original random treatment allocations, regardless of whether there are protocol violations. Although this analysis approach is conservative, it ensures that even when a study is poorly executed, e.g., substantial protocol violations, and/or participant dropout, it is unlikely that the new treatment can be (falsely) proven to have better efficacy than the control.

Data analysis is always specific to the distribution of the outcome measurements. In the proposed EC trial, the outcome is binary, i.e., quit smoking vs. continued smoking. Thus, a 2 by 2 contingency table can be generated, with the two columns as "quit" and "continued", and the two rows as "EC" and "NP" treatments. Depending on the actual count in each cell of the 2 by 2 table, either a chi-squared test or Fisher's exact test can be used to make the comparison. Rejecting the null hypothesis base on the p value (meaning the quit rate is higher for the EC treatment) means that the ECs are superior to the NP treatment; otherwise, we conclude that there is no evidence to demonstrate that ECs are more effective than the NP treatment.

An "intent-to-treat" approach can also be taken for analyzing a non-inferiority trial; however, such an approach does not have a conservative effect, instead it has the opposite effect. Poorer execution of a non-inferiority trial in general results in a similar efficacy measurement between the control and the new treatments, and thus consequently biases toward the alternative hypothesis, which is non-inferiority of the new treatment compared with the control. As an alternative, an analysis can be performed based on the "per-protocol" principle, in which only participants adherent to the trial protocol will be included in the analysis. In fact, it is always recommended that both analyses are performed for a non-inferiority trial. If the two results are similar, then they can be both presented; otherwise, efforts need to be taken to investigate what has caused the differences.

There are other issues to be considered in designing a randomized clinical trial, for example, blinding. Ideally, a double-blind trial is preferred; otherwise, it is difficult to avoid subjective biases.

Keywords: non-inferiority trial, non-inferiority margin, statistical power, outcome

REFERENCES

1. Landow L. Current issues in clinical trial design: superiority versus equivalency studies. Anesthesiology 2000;6(92): 1814–1820.
2. U.S. Department of Health and Human Services Food and Drug Administration. Non-Inferiority Clinical Trials to Establish Effectiveness—Guidance for Industry. https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf (lase accessed: Jan. 7, 2018).
3. Yang S and Berdine G. (2017) Non-inferiority trails. The Southwest Respiratory and Critical Care Chronicles, 5(21):50–52. DOI: 10.12746/swrccc.v5i21.424.