Testing the Wrong Hypothesis in Phase II Oncology Trials: There Is a Better Alternative

Commentary on Vickers et al., p. 972

Mark J. Ratain and Theodore G. Karrison

Although the Code of Federal Regulations (21 CFR 312.21) defines phase II studies as “controlled clinical studies,” the vast majority of phase II oncology trials have been single-arm investigations. Granted, a liberal definition of the word “controlled” would allow the use of historical controls; however, this is not the norm in other therapeutic areas (1).

The custom of single-arm, phase II oncology trials appears to date back to Gehan’s 1961 article (2), which proposed a method to calculate the minimum sample size required to estimate the response rate with a given degree of precision, under the caveat that the study would be terminated early if there was 95% confidence that the response rate was less than some target response rate of interest (usually 15-20%). In this classic design, 14 or 19 patients (corresponding to response rates of 20% or 15%) are initially treated, and if no responses are observed, the drug is considered inactive. On the other hand, if at least one response is observed, additional patients are enrolled and the response rate is estimated. Thus, the Gehan design combines elements of estimation and hypothesis testing.

In the 45 years subsequent to Gehan’s article, there have been a large number of articles about phase II design, most of which have focused on hypothesis testing, rather than estimation (3). It is important to emphasize for the average reader that the hypothesis testing framework is generally organized around the concept of a null and alternative hypothesis. Usually, the alternative hypothesis is what the investigator hopes is not true. Furthermore, these two hypotheses are usually constructed such that one or the other must be true. In the simplest hypothesis testing framework, one is trying to compare a standard and investigational therapy in a randomized study. In this context, the alternative hypothesis is that the investigational therapy is different (and ideally better) than the standard therapy, and the null hypothesis is that the two therapies are the same. The data analysis focuses on testing the null hypothesis. If the null hypothesis is rejected (i.e., \( P < 0.05 \)), then one concludes that the two therapies are different, and by definition, one can accept the alternative hypothesis.

In the Gehan design, the hypothesis being tested during the initial stage of the study is that the drug is active (response rate is \( \text{higher} \) than some minimal response rate of interest). Thus, this is the null hypothesis. If none of 14 (or 19) patients respond, this hypothesis is rejected. Therefore, the alternative hypothesis is that the drug is inactive. Thus, in this design, the investigators (and certainly the patients) would hope that the null hypothesis is true, which is an unusual scenario in clinical trials.

In 1982, Fleming (4) proposed the use of null and alternative hypotheses that do not imply that one or the other must be true, and this framework was also used in Simon’s 1989 design (5). The null hypothesis is that the response rate is less than or equal to some response rate \( (P'_A) \) that “does not warrant further investigation,” and the alternative hypothesis is that the response rate is greater than or equal to some response rate \( (P'_A) \) “which warrants further investigation,” with \( P'_A > P_A \). In this design, rejection of the null hypothesis is desirable and leads to the conclusion that the response rate is greater than \( P_A \) and that the data are not inconsistent with the alternative hypothesis. However, this should not lead to acceptance of the alternative hypothesis, and in many scenarios where the null hypothesis is rejected, the observed response rate is less than \( P'_A \). Thus, there are actually three possibilities: (a) the null hypothesis is true (response rate is less than \( P'_A \)); (b) the alternative hypothesis is true (response rate is greater than \( P'_A \)); and (c) neither the null nor the alternative hypothesis is true (response rate is between \( P_A \) and \( P'_A \)). Investigators need to recognize that rejection of the null hypothesis is consistent with either the explicit alternative hypothesis or the implicit third hypothesis (i.e., neither the null nor the alternative hypothesis is true).

This use of null and alternative hypotheses for uncontrolled phase II trials has been a frequent source of confusion for investigators. We generally expect to accept the alternative hypothesis if we can reject the null hypothesis, but this only applies if these are the only two possibilities. To better illustrate the opportunity for confusion, consider three possible implementations of the Simon design. In all three scenarios, \( P'_0 \) is 0.05 (5% response rate), but \( P_A \) is 0.20 (scenario A), 0.25 (scenario B), or 0.30 (scenario C). Using the National Cancer Institute calculator for the Simon design with \( \alpha \) (type I error) of 0.05 and \( \beta \) (type II error) of 0.10, one can calculate the minimal response rate to reject \( P'_0 \) for each scenario, yielding 5/41 (12.2%) for A, 4/30 (13.3%) for B, and 3/17 (17.6%) for C. In no case does the minimal response rate to reject \( P'_0 \) exceed
references

Testing the Wrong Hypothesis in Phase II Oncology Trials: There Is a Better Alternative

Mark J. Ratain and Theodore G. Karrison