STATISTICS IN MEDICINE
Statist. Med. 2000; **19**:763–766

EDITORIAL

Controlling alpha in a clinical trial: the case for secondary endpoints

Ralph B. D'Agostino, Sr

Boston University, 111 Cummington Street, Boston, MA 02215, U.S.A.

We have become masters of clinical trials. We design them beautifully, our protocols delineate objectives and categorize sharply efficacy variables as primary or secondary, teams of investigators execute them efficiently and then we analyse with the latest and most sophisticated statistical methods. Yet, all too often, they fail, being negative on the primary efficacy variable and positive on a secondary endpoint. Positive and negative refer solely to attaining or not attaining statistical significance. Statisticians try to interpret these trials. Some hold the position that the entire study is negative because the primary efficacy variable does not achieve statistical significance. For them, analyses of secondary variables are, at best, exploratory and no confirmatory conclusions can be made. This position has the logic of adhering to a strict code of alpha spending. If the primary efficacy variable is not significant, then all the alpha has been spent and no further confirmatory looks of the data are possible. Other statisticians oppose this stance as too rigid and argue to continue the analyses, and then judge the study as positive or negative depending on the importance of the significant secondary variables and how the results of all the analyses fit together. Neither approach is satisfactory.

The deliberations of the Cardiovascular-Renal Advisory Committee of the Food and Drug Administration (FDA) of the drug Carvedilol for treating heart failure is an excellent example of the above [1,2]. The primary efficacy variable related to exercise ability and was not statistically significant, however, all-cause mortality was ($p < 0.001$). In fact the Data and Safety Monitoring Committee stopped the study because of the mortality results. The FDA Advisory Committee's statistician, Lemuel Moyé, took the position that the study was negative because the primary efficacy variable was not significant, and the mortality results could not be used to make the study positive. The sponsor's statistician, Lloyd Fisher, argued that mortality was so important that it could not be ignored, the results of other variables were consistent with the mortality results and lack of significance on the exercise variable related to the appropriate decision to stop the trial because of the mortality results. This particular case was resolved by considering the above study in addition to supportive evidence from another independent positive study. Carvedilol was ultimately approved for the treatment of heart failure, interestingly without a mortality claim. This final FDA approval did not resolve the statistical issues.

In this issue of *Statistics in Medicine*, Professor Lemuel Moyé offers another strategy for dealing with the conceptual problem of the negative primary variable and positive secondary variables. He calls it the prospective alpha allocation scheme (PAAS). It involves an *a priori* allocation of alpha at two levels; the usual allocation for primary variables (Moyé calls this the

primary alpha) and an allocation of alpha for the secondary variables. The total alpha for the complete study is the experiment alpha. He also presents a notation to describe at which levels of the variables the significance occurs. With this scheme a study can be called positive even if significance is attained only on secondary variables. Similar concepts of alpha allocation have been suggested in other contexts, for example, for multiple comparisons [3,4] in studies with placebo and active controls. There the alpha for evaluating the sensitivity of a study (comparison of treatments with the placebo) is separated from the alpha for the comparisons of active treatments. Moyé argues the application of his PAAS in confirmatory clinical trials will resolve dilemmas such as seen with Carvedilol.

Two commentaries by Professors Gary Koch and Robert O'Neill accompany Doctor Moyé's article. Professor Koch notes correctly that the real problem is 'how to balance tolerable inflation of type I error against a more extensive structure for evaluating success or failure of a clinical trial'. He outlines nicely how Moyé's PAAS system fits into the context of other attempts to deal with multiple endpoints [5–7]. Of particular interest is his discussion of the resampling methods of Westfall and Young [8] and the use of existing statistical methods to aid in dealing with multiple endpoints [7–11].

Doctor O'Neill's approach is different. He first praises Moyé for his recognition of the problem and his contribution to addressing it, but then points to problems. These include the artificial classification of variables as primary and secondary, the problem of interpreting study results in light of a primary variable alpha and a new overall experiment alpha, and the overemphasis in the PAAS scheme of declaring each variable positive or negative in terms of alpha rather than examining the relation among the variables and the power of the study. He also makes useful suggestions to improve upon PAAS.

Moyé's basic point is the need for prespecification (that is, prospective identification) of efficacy variables and the allocation of alpha for them. Koch and O'Neill agree. The latter discuss how these can fit into multiple testing and resampling methods. They also emphasize the need to consider other features such as power. All three also draw attention to the difference between making a decision and drawing a conclusion [12]. The Moyé article and the commentaries make an excellent contribution to a serious contemporary problem in clinical trials.

THE CASE FOR SECONDARY VARIABLES

Moyé's suggestion of allocating some alpha to secondary variables actually has the effect of blurring the distinction between primary and secondary variables. Once some alpha is assigned to a variable, then there is no impediment to investigating it regardless of the outcome of analysis of primary variables. Moyé's scheme leads to the prospective designation of variables for analysis without concern for the outcome of other variables. Calling some variables secondary variables is arbitrary. This raises the question, do we need the distinction between primary and secondary? I believe the answer is a yes. Primary variables should relate directly to the study objectives. The design and power of the study should focus on them. Secondary variables have a number of important additional functions, all at levels different from the primary variables. There are at least six types or uses of secondary variables. Depending on the study secondary variables fit into one or more of these. The six classes are:

1. *Secondary variables can supply background and understanding of the primary variables.* For example, in a longitudinal study secondary variables can be the efficacy variables measured

on visits before the last visit, or in analgesic trials, where the primary efficacy variable is a sum of pain relief over time, secondary variables can be the pain relief at specific time points.

2. *For composite primary endpoints, useful secondary endpoints are the separate components of the primary variables and other related variables.* For example, in some of the major lipid and hypertension trials the composite endpoint is non-fatal myocardial infarction or coronary death. Two important secondary variables are the separate endpoints, myocardial infarction and coronary death. Also useful as a secondary variable is revascularization.
3. *Major variables for which the treatments under study are important but for which the study is underpowered constitute a major role for secondary variables.* Again in lipid trials, all-cause mortality is the ultimate outcome variable. Few of these studies are powered sufficiently for it, yet all should analyse it.
4. *Secondary variables can have the role of aiding in understanding the mechanism by which the treatment works or in supplying details of the processes of the conditions under investigation.* For example, in a cardiovascular trial with primary endpoint cardiovascular disease, a useful secondary measure could be a quantification of stenosis by a carotid ultrasound.
5. *Secondary variables can be variables that relate to subhypotheses that are important to understand but are not the major objective of the treatment*. For example, for a treatment directed at improving bone density, there may be the possibility that it increases HDL cholesterol. Studies of this agent would most likely consider as secondary variables cardiac outcomes.
6. *Secondary variables can be variables designed for exploratory analyses.* There is always something new to learn even in a confirmatory clinical trial. Exploratory analyses of prespecified secondary variables aid greatly in this learning.

The above list is neither exclusive nor exhaustive. It is put forth principally to demonstrate the broad array of roles that secondary variables fulfil. In the context of Moyé's article prospective alpha allocation seems mainly useful for uses 2 and 3 above. Uses 1, 4 and 5 probably have less concern with alpha, except in the situation where a desperate attempt is made to replace a negative primary variable with a minor secondary variable that happens to attain statistical significance. This situation needs to be discouraged. The exploratory variables of 6 should not introduce major problems for conserving alpha, unless, in a *post hoc* fashion, they are elevated to primary variables.

In this age of large-scale international clinical trials where it is common to have 10 000 or more subjects followed for three or more years, the need to gain more and more information from the studies is essential. It is not sensible to say the studies rise or fall based solely on a small number of outcome variables declared to be primary. Neither can it be the case that the interpretation should rest solely or mainly on exploratory data mining. The clear *a priori* declaration and correct analysis of major secondary endpoints coupled with appropriate subset analyses are needed.

The challenge to statisticians is to produce valid methods for analysing studies and obtaining uncontestable decisions. For dealing with secondary variables, the use of multiple endpoint statistical testing strategies, resampling methods and Bayesian methods exist and need further clarification and development in their application to confirmatory clinical trials. Doctor Moyé adds another contribution. The need for more is obvious.

REFERENCES

1. Fisher L. Carvedilol and the FDA approval process: the FDA paradigm and reflections upon hypothesis testing. *Controlled Clinical Trials* 1999; **20**: 16–39.
2. Moyé LA. P-value interpretation in clinical trials. The case for discipline. *Controlled Clinical Trials* 1999; **20**: 40–49.
3. D'Agostino RB, Massaro J, Kwan H, Cabral H. Strategies for dealing with multiple comparisons in confirmatory clinical trials. *Drug Information Journal* 1993; **27**: 625–641.
4. D'Agostino RB. Multiple comparisons in over-the-counter drug clinical trials with both positive and placebo controls. *Statistics in Medicine* 1991; **10**: 1–6.
5. Davis CE. Secondary endpoints can be validly analyzed, even if the primary endpoint does not provide clear statistical significance. *Controlled Clinical Trials* 1997; **18**: 557–560.
6. Koch GG, Gansky SA. Statistical considerations for multiplicity in confirmatory protocols. *Drug Information Journal* 1996; **30**: 523–534.
7. Koch GG, Davis SM, Anderson RL. Mythological advances and plans for improving regulatory success for confirmatory studies. *Statistics in Medicine* 1998; **17**: 1675–1690.
8. Westfall PH, Young SS. *Resampling-based Multiple Testing: Examples and Methods for p-value Adjustment*. Wiley: New York, 1993.
9. O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics* 1984; **40**: 1079–1087.
10. Lehmacher W, Wassmer G, Reitmeir P. Procedures for two-sample comparisons with multiple endpoints controlling the experimentwise error rate. *Biometrics* 1991; **47**: 511–532.
11. Hochberg Y. A sharper Bonferonni procedure for multiple tests of significance. *Biometrika* 1997; **75**: 800–802.
12. Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH. *Biostatistics in Clinical Medicine*. MacMillian Publishing Co.: New York, 1987.

