Drug Information Journal, Vol. 34, pp. 365–377, 2000
Printed in the USA.
0092-8615/2000

A DECISION ANALYTIC APPROACH TO DETERMINING SAMPLE SIZES IN A PHASE III PROGRAM

ALLAN PALLAY, MS

Department of Clinical Biostatistics, Wyeth-Ayerst Research, Philadelphia, Pennsylvania

In this paper we present a decision analytic approach to determining sample size in a set of Phase III drug efficacy trials sponsored by a pharmaceutical company. In this approach we built a model to predict the expected net present value of the drug at varying sample sizes. We then chose the sample size that maximizes that value. We took into consideration effects that sample size had on the probability of approval, the cost of the studies, the time it takes the drug to reach the market, and a variety of other factors. We found that increasing the sample size increases the chance that the drug will be approved. On the other hand increasing the sample size increases the cost of the studies and the time it will take for the drug to reach the market. The model weighs these factors to produce the expected net present value.

Key Words: Decision analysis; Sample size; Pharmaceutical industry

INTRODUCTION

WE PRESENT A DECISION analytic approach to determining the sample size in a set of Phase III efficacy drug trials sponsored by a pharmaceutical company. This approach involves building a model using clinical trials data and postapproval behavior of the market, and using this model to predict the expected net present value (ENPV) of the drug at varying sample sizes. We then chose the sample size that maximized that value. (The ENPV is the current monetary value of a cash payment which will be received at some time in the future. X amount of dollars to be received in the future is worth less than the same amount of money received today because present cash can be invested to increase its value. A present value algorithm determines how much a future cash payment is worth in dollars that are received today. The *net* present value subtracts out any costs associated with receiving the cash payment. An *expected* net present value multiplies the net present value by the probability that the money will actually be received.) In the model, we considered the effects of sample size on the probability of approval, the cost of the studies, the time it takes the drug to reach the market, and a variety of other factors. Cost and time to market are important factors that affect the ENPV and are affected by sample size, but are not taken into consideration in the usual statistical method of determining sample size. Although the importance of the cost of the studies, the time to market, and the probability of approval are intuitively clear, it is not clear how to weigh them in arriving at a sample size. For this reason a decision analytic model is very useful.

Decision theoretic approaches to determining sample size have been reported by a number of authors (1–7). The utility func-

Reprint address: Allan Pallay, MS, Department of Clinical Biostatistics, Wyeth-Ayerst Research, P.O. Box 42528, Philadelphia, PA 19101. E-mail: pallaya@war.wyeth.com.

tions in these studies, however, were not monetary. They were primarily related to deriving the maximum public health benefit as measured by maximizing the number of patients treated with the best drug among a group of drugs that were tested. The cost of experimentation was also taken into consideration in Canner (2), Sylvester (5), and Brunier and Whitehead (7).

Berry and Ho (9) used a decision theoretic model that had a monetary outcome. The monetary gain function, however, was hypothetical and very simple and the effect of time to market was not taken into account. Berry and Pallay (unpublished data; 1999) used a decision theoretic approach to determine sample size in a Phase III program and used a real monetary gain function derived from a marketing study. The monetary gain function, however, was taken as fixed and was not studied, and the effect of time to market was not taken into account.

In this paper we extend the range of application of decision analysis in the sample size question by including market considerations in our analysis. We do this by building a market model and using it to examine the effect various uncertainties have on the value of different sample size choices. As work in this paper reflects our real-life experience in dealing with a particular drug, the drug is referred to as "Drug X."

KEY CLINICAL AND MARKETING ISSUES

Clinical Issues

Drug X is an experimental treatment for a very serious disease in which death or severe disability is a common outcome. There are currently no competing drug treatments in the class of Drug X. Therefore, any level of treatment effect by Drug X over placebo will be an improvement over what is currently available to patients. Also, previous data showed virtually no evidence of significant adverse effects at the doses that will be tested.

Two essentially identical studies were done to assess the efficacy and safety of the drug, each with placebo and two doses of Drug X. The doses are the same in these two studies. The aim was to have two doses that worked, with the higher dose possibly giving greater efficacy and the lower dose possibly having less adverse effects; although significant adverse effects were considered unlikely at either dose. If the results of the studies are positive, the company will apply for approval at the Food and Drug Administration (FDA). Also, it was decided to complete the trials and to not stop early for substantial efficacy.

Marketing Issues

Although no drugs that would compete with Drug X are on the market, a number of competing drugs are in development at other companies. Available evidence suggests that these competitors may have similar efficacy to that of Drug X, and some of them may have a greater side effect burden. There is, however, substantial uncertainty about these points.

The market share that Drug X will take is mainly a function of the quality of the drug relative to its competitors and its order of entry into the market. The marketing experts believe that the order of entry into the market of Drug X is likely to be much more important than its relative quality, unless Drug X differs substantially in risk/benefit from the others.

INADEQUACY OF THE STANDARD SAMPLE SIZE FORMULA

The analysis of these studies involves the comparison of each Drug X dose group against placebo. We, therefore, consider a sample size calculation based on an individual comparison with no adjustment for multiplicity. An expert in the field said that an advantage over placebo of 0.33 units on the efficacy measure would definitely be clinically significant. Also, we specified the alpha and beta errors to be 0.05 (two sided) and 0.10, respectively, and the standard deviation to be 2.2 (based on previous studies). Using

these values, the sample size per group would be about 934 patients.

Standard sample size formula ignores effects of the cost of the studies. The time to market also is not taken into consideration. Both of these factors have the effect of reducing the sample size. Even if we could ignore cost factors, the main problem with standard sample size formula is specifying the minimum advantage over placebo that we want to detect (with specified error rates). The disease to be treated is very serious, with no competing therapies and with current evidence suggesting that Drug X will have only trivial adverse effects. In such a situation it could be argued that any level of effect of Drug X that is greater than the effect of placebo, no matter how small, is clinically significant.

OVERVIEW OF THE DECISION MODEL

The aim of the decision model is to show how the size of the sample affects the ENPV of the drug project. The model is a simulation which operates by simulating multiple possible clinical trial outcomes and associated market responses, to produce multiple possible net present values (NPVs). There are multiple possible NPVs because some of the elements of the model are expressed as random variables. The mean of these NPVs for a given sample size is the ENPV. The optimal sample size is taken to be the size that maximizes the ENPV.

Figure 1 gives an overview of how the simulation model works. It is instructive to think of the three ovals at the top of the diagram as the "clinical component," the bottom two ovals as the "cost component," and the rest in the middle as the "market component." The simulation begins by setting a sample size and sampling a value from the prior distribution of the efficacy parameter. These are used as input into the clinical trials simulation model. The output of this model is simulated results of the clinical studies which are input into the FDA approval model. The output of this model is "approved" or "not approved." This is the crucial input into the market component.

The simulation of the market component also uses the sample size as an input. It is used to determine the time it will take Drug X to be approved. This time, along with the efficacy level (based on the value sampled from the prior) are the inputs to the market share model. The output of the market share model (market share) and the output of the market size model (monetary size of the market) are inputs into the formula that calculates income.

The cost component consists of the cost of the studies (which is a function of sample size) and a set of costs associated with the production and marketing of the drug (which are independent of sample size). Approval, market share, market size, and costs are all inputs that are used to calculate income. Income is, in turn, used to calculate the net present value.

A DETAILED DESCRIPTION OF THE MODEL

We begin this section with a description of the statistical model that forms the basis of the clinical component. Next, we give a detailed description of each of the elements shown in Figure 1 and show how they are put together to calculate the income and NPV.

The Statistical Model

There were two studies each with two doses and placebo. Consider first the high dose compared with the placebo in one study. Let:

y = the observed advantage for drug X over placebo on the efficacy measure,
δ = the population advantage for drug X over placebo on the efficacy measure, and
N = the sample size per treatment group (with equal allocation in each group).

The statistical model is:

$$y = \delta + e,$$

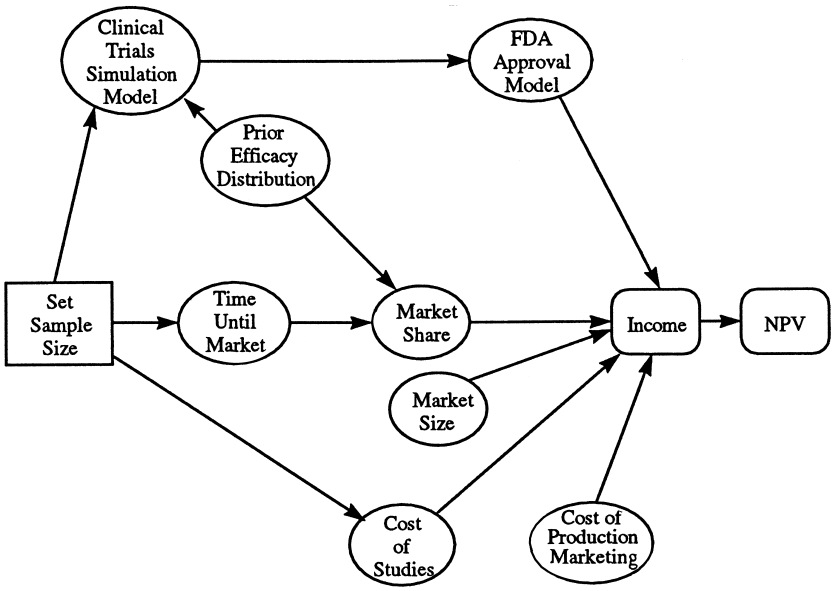

FIGURE 1. Overview of the decision model.

where e is the error term independent of y with a standard deviation = σ.

The same statistical model is used for the comparison of the low dose group versus placebo except that δ is assumed to be lower by a specified amount. The second study is modeled in the same way as the first.

Prior Distribution of the Efficacy Parameter

To determine the probability of getting a statistically significant result in a comparison of a dose group and placebo, it is necessary to assume a value of the population parameter (δ). In this model we assume a distribution of values of δ, which is, in essence, a Bayesian prior distribution of the efficacy parameter. The prior was set by the clinical group and the senior author. It is shown in Figure 2 and represents the best guess as to the distribution of the population efficacy parameter before beginning the study. Two other priors (pessimistic and optimistic views) were used in sensitivity analyses.

The output of this element of the model is a randomly sampled value of δ from the prior. It is used as input by the clinical trials simulation model. It will also be used as input into the market share model as will be described later.

The Clinical Trials Simulation Model

Consider first the simulation of the high dose group versus placebo and the sample size fixed at N. The simulation uses as input the sampled value from the prior distribution of the efficacy parameter (δ), and the sample size per group (N). The steps of the simulation are as follows:

1. A random value is drawn from the distribution of $\bar{e}$,
2. A Z type statistic is formed: $(\delta + \bar{e})/(\sigma/\sqrt{N})$, and
3. The Z statistic is classified as "significant" if Z is greater than 1.96, a "trend" if Z is greater than 1.44, and negative otherwise.

To simulate the low dose trial, 0.075 is subtracted from the sampled population parameter. This reflects the assumption that the lower dose will give somewhat less efficacy.

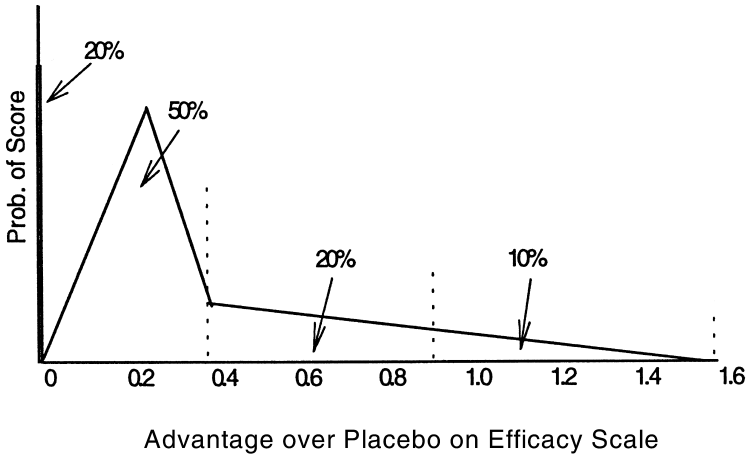

FIGURE 2. Prior on the population efficacy base model. A rough assessment of the clinical meaning of the advantage over placebo scores are as follows: 0 to 0.1 = very minimal efficacy; 0.1 to 0.2 = minimal efficacy; 0.2 to 0.3 = small; 0.3 to 0.5 = fairly good; 0.5 to 0.7 = quite good; 0.7 to 0.9 = very good; greater than 0.9 = outstanding.

The outcome of the clinical trials model for a single iteration of the simulation is a classification of the comparison of each of the two dose groups with the placebo group as "significant," "supportive," or "negative" for the two studies.

The FDA Approval Model

In the first part of the FDA approval model, each of the two studies is classified as positive, supportive, or negative. A study is classified as positive if the high dose is significant (regardless of the observed response of the low dose) or if the low dose is significant and the high dose is a trend. A study is classified as supportive if the high dose is a trend and the low dose is not positive (If the low dose were positive the study would be positive). All other outcomes were classified as negative. Using this algorithm on both studies, the following are possible outcomes: two positive, one positive and one supportive, or none positive.

Usually, two positive studies are required for FDA approval but given the seriousness of the disease and the lack of treatments something less may be acceptable. Therefore, we assumed that there is a 25% chance that only one positive study would be sufficient, a 50% chance that a single positive along with a single supportive study would be sufficient, and a 25% chance that two positive studies would be needed. Using input from the clinical trials model, the outcome for a single iteration of the FDA approval model is "approved" or "not approved."

Time to Market

The time to market (in months) for a given sample size was determined by first calculating the total number of patients that are needed across the two studies and then dividing by the expected enrollment rate per month. In addition, the time to start the study, the time to produce an application to the FDA, and the time until FDA approval were also estimated, and added to the time to do the studies. The outcome of a single iteration of the simulation is the number of months until the drug will reach the market. This number of months is the same for each iteration of the simulation since all of the elements are fixed. The number of months until

market changes (in linear fashion), however, when the simulation is run with a different sample size.

The Monetary Size of the Market

To estimate the monetary size of the market, the number of patients who might receive Drug X therapy is estimated for each year, beginning at the year the studies started (referred to hereafter as Year 1), and going out 15 years. For each of these years, this number is multiplied by the estimated price of the treatment to obtain the monetary size of the market for that year. The monetary size of the market is assumed to increase at a uniform rate up to the 12^{th} year (hereafter referred to as Year 12). The price of the drug is reduced by 70% in Year 13 because the patent will expire. The outcome from the market size model is the monetary size of the market for each year up to Year 15. These values are independent of the sample size and are constant for each iteration of the simulation.

The Market Share Model

The market share model is based on a model developed by Notvest (unpublished data; 1998). The aim of the market share model is to estimate the percent of the market that Drug X will receive for each year up until Year 15. It takes into account the time it takes Drug X to reach the market, possible competitors (the greater the number of competitors the less the share), and the level of efficacy of Drug X (the greater the efficacy the more the share). The level of efficacy for any iteration of the model is the value of the population efficacy parameter that is sampled from the prior distribution.

In the market share model we assumed the existence of four possible competitors (labeled drugs A, B, C, and D). Drugs A, B, and C are actual drugs under development at other companies. Drug D is another hypothesized competitor that is unknown to us but assumed to exist. The outcome from the market share model is the percent of the market that Drug X gets for each year until Year 15.

Costs

There are two categories of costs. The first is the cost of the two studies which is determined by the per-patient fee that is paid to investigators. The second category of cost includes costs of: goods sold, sales and marketing, general administrative, taxes, and other supportive studies. These costs are independent of the sample size.

Income and NPV

Figure 3 shows how the income is calculated for each of the first five years for an iteration of the model. The first line shows the monetary size of the market for each year (M_i). The next line shows the estimated proportion of patients who will receive Drug X as determined by the market share model (p_i). In this figure possible approval of the drug will occur in Year 3. Therefore, the proportion of the market is zero in Years 1 and 2. Note that the year that approval is applied for varies according to the sample size, with higher sample sizes making this year later in time.

Next are costs. In Years 1 and 2 these costs are the costs of the studies and other costs associated with the development program. In the later years these costs are those associated with production and marketing. Next is an indicator variable for drug approval, with $a = 1$ if approved and $a = 0$ if not approved. All of the above lines are combined in the next line as shown. Note that the costs in the first two years will be incurred regardless of approval. For the next three years, the revenues and the costs are only realized if the drug is approved ($a = 1$). To account for taxes 37.5% of the income (or loss) at each year is subtracted out.

The ENPVs were based on approximately 6000 iterations for each sample size. For the sensitivity analysis approximately 3000 iterations were used for each sample size.

	Year				
	1	2	3	4	5
Market Size ($ Millions)	M_1	M_2	M_3	M_4	M_5
Market Share (%)	0	0	p_3	p_4	p_5
Costs ($ Millions)	C_1	C_2	C_3	C_4	C_5
Approval (a = 1 or 0)	-	-	a	a	a
Pretax Income	$-C_1$	$-C_2$	$a(M_3p_3-C_3)$	$a(M_4p_4-C_4)$	$a(M_5p_5-C_5)$

FIGURE 3. Calculations of the pre-tax income for each year.

Sensitivity Analyses

The model in which the assumptions are based on the best guesses of the various experts is referred to as the "base model." This is in contrast to models where various alternative assumptions are used to check the sensitivity of the model to these assumptions. The values of the alternative assumptions were generated through discussions between the author and various experts. Most of the alternative values represented an upper and lower value level for what is felt to be reasonably possible.

AN EXAMPLE OF ONE ITERATION OF THE MODEL

The Clinical Component of the Model

Table 1 shows the results from one iteration of the "clinical component" of the model. It is based on a sample size of 250 per dose group. The simulation begins, as shown in the third column of the table, with a sample from the prior distribution of the population efficacy parameter (mean advantage over placebo). Note that this value is the same for the high dose in the two studies. To calculate the population value for the low dose, 0.075 was subtracted from the high dose value. Next an "error" term is added as described in the statistical model. The simulated observed score in the next column is the sum of the population and error scores, and the z score is the observed score divided by the standard error. As shown in the next column, for this iteration of the model, the high dose in Study 1 is significant, thus yielding a positive study. The high dose in Study 2 is a "trend," thus yielding a supportive study.

The next column, "FDA Criterion," shows the outcome of a sample from the random variable describing what might be the minimal requirement for approval by the FDA. In this case at least one positive and one supportive outcome are required, thus for this iteration, FDA approval would have been achieved. Note that had the sample for the "FDA Criterion" random variable yielded the need for two positive studies the outcome for this iteration of the simulation would have been "not approved."

The Market and Costs Component

In the first part of the "market component" of the model the number of months it will take Drug X to reach the market is calculated. With a sample of 250 patients per treatment group and three treatment groups in two studies a total of 1500 patients will be required. The enrollment rate is expected to be 112 patients per month so it is estimated that the studies will take about 13 months. Additional time will be required to start the studies, to

TABLE 1
An Iteration of the Clinical Component of the Model

Study	Dose	Population Parameter	Error Value	Observed Value	Z Score	Study/Dose Classification	Study Classification	FDA Criterion	Approved?
1	High	0.386	+0.10	0.486	2.47	Significant	Positive	1 Positive	Yes
1	Low	0.311	–0.02	0.292	1.48	Trend		and	
2	High	0.389	–0.06	0.326	1.66	Trend	Supportive	1 Supportive	
2	Low	0.311	–0.07	0.241	1.21	Negative			

TABLE 2
An Iteration of the Market Component of the Model

Year	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$ Size of Market (Millions)	136	279	428	586	751	923	1026	1051	1078	1105	1132	1161	359	363	383
Market Share (%)	—	—	—	44	52	46	40	37	35	34	33	32	32	32	31
Revenue	—	—	—	260	391	423	406	385	376	374	374	377	114	116	118
Costs	10	24	34	86	129	140	134	127	124	123	123	124	38	38	39
Before Tax Income	–10	–24	–34	174	262	283	272	258	252	250	251	252	77	78	79
After Tax Income	–6	–12	–15	68	91	87	74	63	55	48	43	38	10	9	8

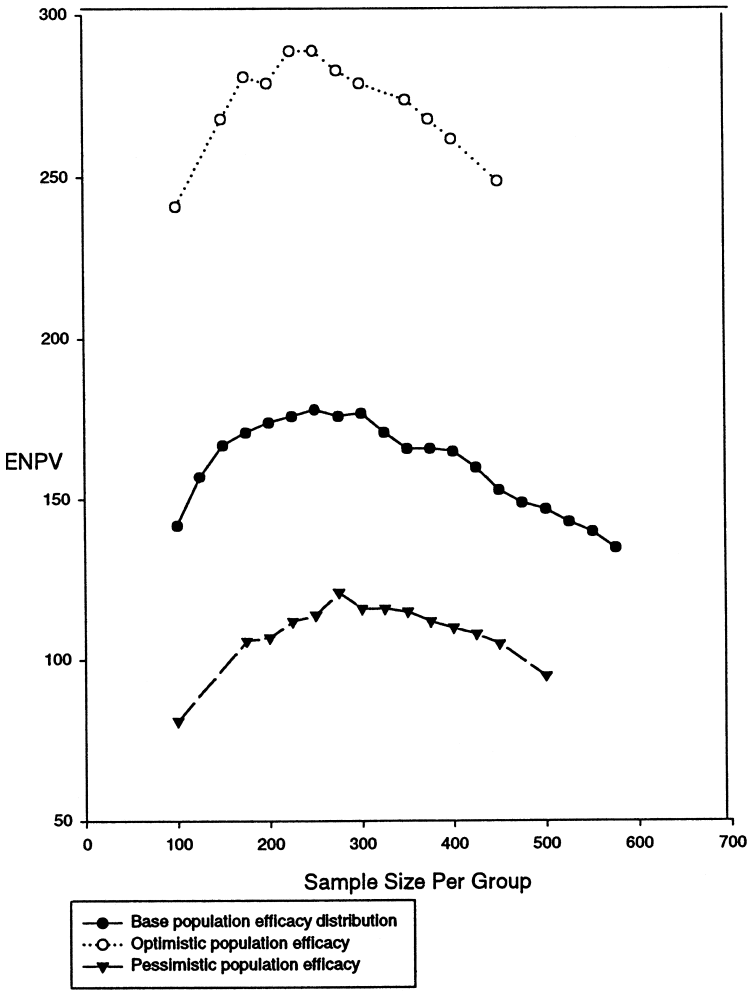

FIGURE 4. ENPV versus sample size for various prior distributions.

collect and enter the data upon completion, to write up and submit an application to the FDA, and for FDA review. With all of these times added in, the total time until market with 250 patients per group is 38 months. This means that income would potentially start flowing into the company in the fourth year.

The next parts of the model are displayed in Table 2. The second row shows the estimated monetary size of the market (in millions of dollars) for each of 15 years. This is equal to the number of patients who could potentially get Drug X treatment (if Drug X got 100% of the market) times the price. The size of the market drops considerably in Year 13 because the patent will expire and the price will be reduced.

The next row shows the share that is expected for Drug X. This value came from the market share model. The product of these two lines gives the revenue. Next are the various costs that were outlined in the last section. The revenue minus the costs gives the income and then this income is subjected to taxes to get the net income or cash flow. To account for additional cash flow after the

15th year, $67 million was added as a final value in the cash flow. This cash flow was entered into a present value formula to give the ENPV. For this iteration of the model the ENPV was $574 million. Note that had the clinical component of the model yielded a "not approved" outcome there would be no positive cash flow.

Figure 4 is a graph of sample size versus ENPV. The middle plot is generated from the base model. The maximum ENPV occurs at about 250 patients per group. The distribution of the NPVs for a run of the model with 250 patients is shown in Figure 5.

DYNAMICS OF THE MODEL

The following set of analyses were aimed at gaining a deeper understanding of the key elements that caused the ENPV to reach a maximum at the sample size of about 250. One goal of achieving this understanding is to explain the somewhat unexpected result that the maximum ENPV occurred at a much lower sample size than the sample size derived from the use of the standard statistical sample size formula.

Effect of Sample Size on Probability of Approval and Monetary Value if Approved

There are many components in the model but the key dynamic is as follows. Sample size has a perfect direct linear relationship with time to market. Increasing sample size (and therefore the time to market) decreases the amount of money that will be made if the drug is approved. Cost of the studies increases with sample size but analyses reveal that the effect of this factor is relatively small compared with the others. On the other hand, increasing sample size increases the probability that the drug will get approved, and, therefore, increases the probability that any money will be made at all. The functional relationship among these factors is shown in Figure 6. In this graph the NPV of the drug, if approved, is plotted against sample size and scaled based on values on the y-axis at

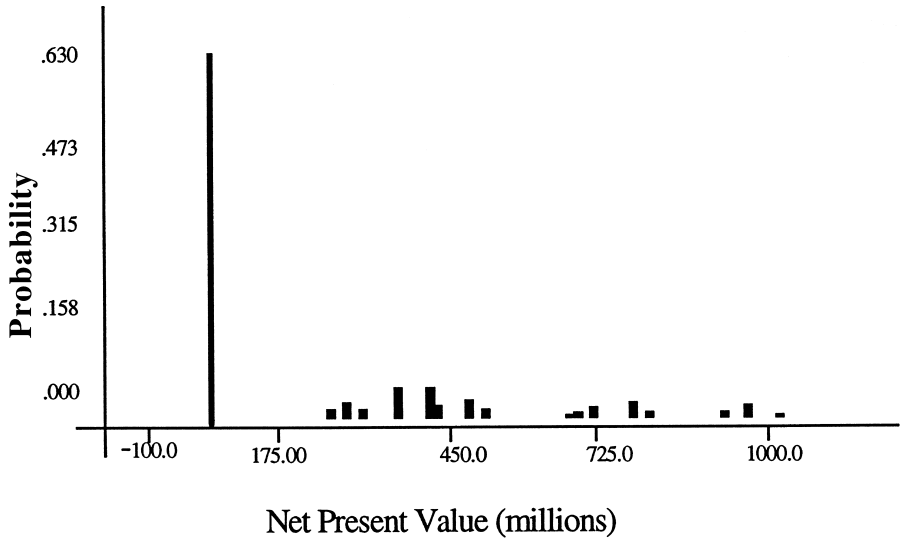

FIGURE 5. Distribution of net present values based on a sample of 250 per group and the base model.

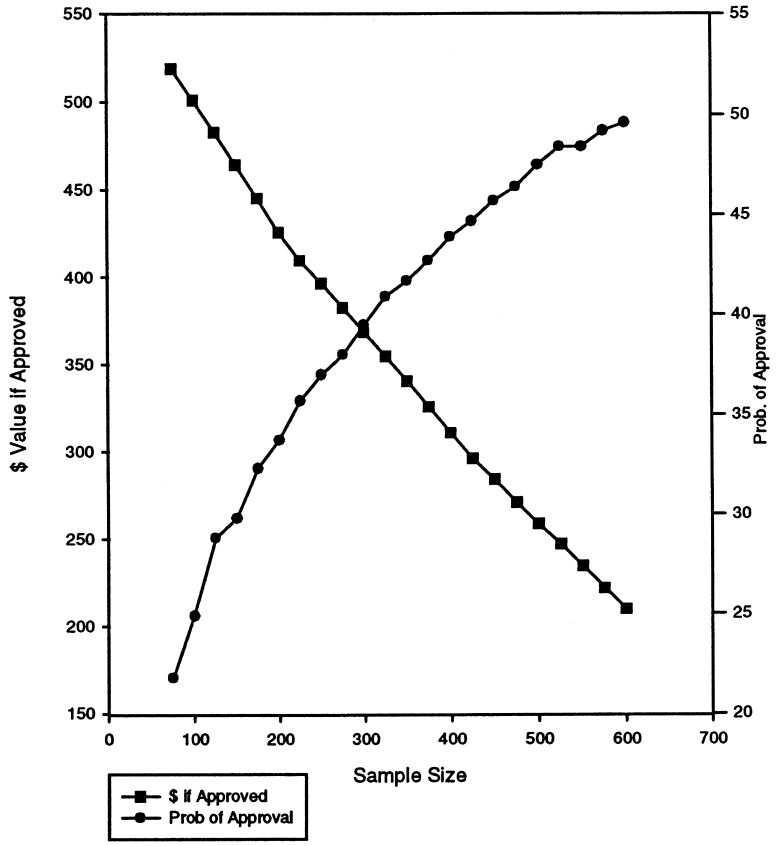

FIGURE 6. The sample size versus dollar value is approved and the probability of approval for three fixed competitors.

the left side of the plot. The probability of approval (referred to hereafter as "prob-approve") is plotted against sample size on the same graph but is scaled based on values shown on the y-axis at the right side of the plot. (The values in the graph are based on a slightly simplified version of the model in which the random variables in the market side of the model are fixed at single values. The first three competitors—A, B, and C—definitely enter the model, but the last competitor, D, does not. Fixing these values makes it easier to create the plots and analysis shows that the change does not substantially change the results or the underlying dynamics of the model.)

The plot helps explain why the sample size chosen by the decision analysis is so much lower than that chosen by using the standard formula. First, the standard method does not take the time until market into account at all. (Recall that time until market is a linear function of sample size.) In addition, the negative impact on the monetary-value-if-approved of increasing sample size is relentlessly linear, while the positive impact of increasing sample size on the prob-approve diminishes with increasing sample size.

Another factor that helps explain the substantial difference between the sample sizes given by the decision analysis and the standard formula relates to the relationship between prob-approve and the prior distribution of the efficacy parameter. The probability of approval is largely a function of the probability of getting a significant difference between a Drug X dose group and the placebo group.

The decision approach incorporates the possibility that the efficacy parameter is so high or so low that increasing sample size makes very little difference in the probability of approval. In those regions of true efficacy, large samples tend to decrease ENPV. Thus, the optimal sample sizes calculated using the decision model are much less than those calculated with the standard approach.

DISCUSSION

Increasing sample size increases the chance of approval, but it also reduces the amount of money that can be made, mainly because of the delay in getting the drug to market. The importance of these two factors is clear, but it is not clear how to weigh them in arriving at a sample size. For this reason an analytic model is extremely useful. On the other hand, it is also clear that the model did not capture all of the complexities of the clinical and commercial situation. For example, the considerations used by the FDA and other regulatory agencies to determine approval involve an analysis of many factors, some of which no one could predict before seeing the data. Also, the market model is an obvious simplification of all the forces that determine economic value. For these reasons, we did not see the sample size that is derived from this model as *the* final answer. Rather, it is seen as a "baseline value" in which the *key* pieces of information have been integrated.

Acknowledgments—I thank Clint Korver from On Your Mind Inc. for his many contributions to the development of the model. He served as a consultant on this project. I also thank the management of Clinical Biostatistics at Wyeth-Ayerst Research (Richard Entsuah, Assistant Director; Gary Littman, Senior Director; and Robert Starbuck, Assistant Vice-President) for their support. Finally, I thank Mike Dazenski (Wyeth-Ayerst New Products Marketing Group) for substantial input into the market model.

REFERENCES

1. Colton T. A model for selecting one of two medical treatments. *J Am Stat Assoc.* 1963a;58:388–400.
2. Canner PL. Selecting one of two treatments when the responses are dichotomous. *J Am Stat Assoc.* 1970;65:293–306.
3. Day NE. Two-stage designs for clinical trials. *Biometrics.* 1969;25:111–118.
4. Donner A. The use of auxiliary information in the design of a clinical trial. *Biometrics.* 1977;33:305–314.
5. Sylvester RJ. A Bayesian approach to the design of phase II clinical trials. *Biometrics.* 1988;44:823–836.
6. Palmer CR. A comparative phase II clinical trials procedure for choosing the best of three treatments. *Stat Med.* 1991;10:1327–1340.
7. Brunier HC, Whitehead J. Sample sizes for phase II clinical trials derived from Bayesian decision theory. *Stat Med.* 1994;13:2493–2502.
8. Berry DA, Wolff MC, Sack D. Decision making during a phase III randomized controlled trial. *Control Clin Trials.* 1994;15:360–378.
9. Berry DA, Ho CH. One-sided sequential stopping boundaries for clinical trials: A decision-theoretic approach. *Biometrics.* 1988;44:219–227.