Adaptive Trials and Bayesian Statistics in Drug Development

Donald A. Berry

University of Texas M.D. Anderson Cancer Center
1515 Holcombe Blvd, Houston TX 77030

Introduction

Designs of phase II and III clinical trials are usually static in that the sample size and any prescription for assigning treatment, including for randomization protocols, are fixed in advance. Results observed during the trial are not used to guide its course. There are exceptions. Some phase II cancer trials have two stages, with stopping after the first stage possible if the results are not sufficiently promising. And most phase III protocols specify interim analyses that determine whether the trial should be stopped early for sufficiently strong evidence of a difference between competing treatment arms. However, traditional early stopping criteria are very conservative and so few trials stop early.

The simplicity of trials with static designs makes them solid inferential tools. Their sample sizes tend to be large, at least in comparison with alternatives discussed in this article. And they usually consider two therapeutic strategies, or arms, thus enabling straightforward treatment comparisons. I do not mean that static trials always give clear answers as to whether one arm is better than the other, only that they usually allow for an unambiguous quantification of the uncertainty regarding whether one arm is better.

Despite their virtues, static trials lead to slow and unnecessarily costly drug development. Millions of dollars and many years can be devoted to developing a single drug, one that may not make it to market. For a company developing a moderate number of drugs (say 20 or more), this circumstance is tolerable because costs are balanced by profits from other drugs. Smaller companies are at the mercy of the prevailing attitudes toward drug development and risk going belly up.

The tradition of drug development is singular—we develop drugs one at a time. The number of drugs available for development is increasing exponentially. Focusing on a single drug while a gazillion others are sitting on the sidelines waiting to be evaluated is enormously inefficient. The standard types of errors in drug development are false positives and false negatives.

(continued)
But these errors apply to drugs actually being tested. There is another kind of error that applies to drugs not under investigation: false neutrals! And it applies to every such drug. A drug not being developed has no chance of helping anyone. Medical research should be indicted when it fails to develop therapies. Sure, resources are limited. But that is precisely the point. When resources are limited we should approach their allocation in a sensible way. And what makes sense today may well be different from the ways of the past.

Pharmaceutical companies and the medical establishment generally must be able to consider hundreds or thousands of drugs for development at the same time. Static trials inhibit simultaneously processing many drugs. Nor can they efficiently address dose-response questions in the context of the development of many drugs. Dynamic designs that are integrated with the drug development process are necessary for moving medical research along at a pace faster than the snail’s.

The purpose of this article is to describe a family of designs that are dynamic in the sense that observations made during the trial can affect the subsequent course of the trial. The general class of designs is called adaptive or sequential. Among the adaptations possible are: (i) stop early, (ii) extend the trial beyond its planned sample size if the conclusion is still unclear, (iii) drop treatment arms or doses used, (iv) add arms or doses, (v) change the proportions of patients assigned to the various treatment arms and (vi) shift seamlessly (without stopping accrual) into a later phase of drug development. In all of these, the accumulating results are monitored as closely as possible to better inform decisions.

A focus of this article is on clinical trials, but the ideas apply at least as forcefully in preclinical trials. A main bottleneck of the drug development process occurs at the level of the preclinical animal toxicity/carcinogenicity studies. There are many opportunities for using adaptive designs in the preclinical area that will efficiently identify the best drugs to move forward in humans.

Depending upon one’s goals and one’s statistical philosophy, there may be a price paid for such flexibility. There are two principal philosophical views: frequentist and Bayesian. Although the frequentist approach is traditional in pharmaceutical research, times are changing and the Bayesian approach is being used increasingly in drug development. The next section is a brief primer on statistical philosophy. The ideas are necessary for understanding the statistical aspects of adaptive trials.

**Frequentist vs. Bayesian Approaches**

The principal difference between the two approaches is the interpretation of probability. Of particular importance are the quantities about which one can make probability statements. Bayesians associate probabilities with any quantity that is unknown. Frequentists restrict probability to quantities that can be observed repeatedly—hence, long-term frequency. Bayesians can make probability statements about true success rates, for example, while frequentists cannot. Suppose the success rate using therapy A is $\theta_A$ and using therapy B is $\theta_B$. A typical Bayesian conclusion is: "Given the results of this trial, the probability that $\theta_A$ is greater than $\theta_B$ is 85%." Frequentists assume particular values of $\theta_A$ and $\theta_B$ and calculate probabilities of possible data under these assumptions. An example of a frequentist statement of a $p$-value is: "If $\theta_A = \theta_B$ then there is a 10% probability of observing a result as or more extreme than the one observed in this trial." The numerical values in these statements are arbitrary and unrelated.

Indeed, for the same experimental evidence, a Bayesian probability that therapy A is superior may be the same as the analogous frequentist $p$-value or it may be much larger or much smaller.

A critical distinction between the two approaches is manifest in the above comparison, but it is so subtle that it seems innocuous to nonstatisticians. Moreover, it is not universally appreciated within the statistical community. Bayesian probabilities apply to the actual results obtained in the trial while frequentists add in "tail areas," probabilities of results "more extreme" than those actually observed. These results were possible under the design used, but they happened not to be observed. For example, if 25 out of 40 patients are successes on therapy A and 15 out of 30 patients are successes on therapy B then a Bayesian calculates the probability distributions of $\theta_A$ and $\theta_B$ given these data. From these posterior distributions the Bayesian can find the probability that $\theta_A > \theta_B$, that $\theta_A > \theta_B + 0.10$, that $log(\theta_A(1-\theta_A)/(\theta_B(1-\theta_B))) > 0.10$, etc.

Life is more complicated for frequentists. One reason is that "more extreme" is seldom clear. Suppose the results above arose from a sampling scheme in which exactly 40 patients were to be assigned to therapy A and 30 patients were to be assigned to therapy B. One might reasonably view "more extreme" as having a smaller likelihood ratio than that for the actual observations. The following table shows the exact rejection region for the likelihood ratio test corresponding to 25 successes of 40 on therapy A versus 15 successes of 30 on therapy B. For the indicated number of successes on therapy B (the top number in each pair), reject the null hypothesis if the number of successes on therapy A is greater than or equal to the value indicated in the table (the bottom number in each pair). The values observed in the example are shaded.

<table>
<thead>
<tr>
<th>$S_B$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_A$</td>
<td>1</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>11</td>
<td>13</td>
<td>14</td>
<td>16</td>
<td>17</td>
<td>19</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
</tbody>
</table>

The $p$-value is the probability of the region indicated in this table, calculated under the assumption that the two success rates are equal, $\theta_A = \theta_B = \theta$. This probability depends on $\theta$, but it is nearly constant at about 0.15 (one-sided) over the likely range of values of $\theta$. And luckily, the asymmetry kick in early—a chi-square or normal test also gives a $p$-value of about 0.15.

The $p$-value will be different, and perhaps very different, for other designs. Consider an adaptive sampling scheme.
Suppose that time to response is sufficiently short that each patient's response is available for use in assigning treatment to the next patient. Suppose that the trial was to stop after 70 patients had been treated and that treatments were allocated using "play-the-winner" (Robbins 1952). In this design the first patient is assigned randomly to receive treatment A or B. Thereafter, the same therapy is assigned to the next patient if the previous observation was a success and to the opposite therapy if the previous observation was a failure ('switch on a loser'). (This design may not be very good in any sense, but it is easy to describe and it serves to make my point.) At any time, the difference between the numbers of failures on the two therapies is 1 or 0—as it was at the end of the trial in the example (15 − 15 = 0). What is the p-value? In an actual trial in which a similar assignment rule was used, there have been at least 14 different p-values published, and these run the gamut from close to 0 to 1. And many frequentists feel that calculating a p-value is not possible in the circumstances of a design such as this. Life is even more difficult if the stopping rule had been complicated. It is always possible to find a p-value by conditioning on the null hypothesis and simulating the trial, but such a calculation requires that one decide what values are "more extreme." Moreover, sometimes various aspects of the sampling scheme or stopping rule are unknown, and this can make frequentist inferences impossible. It is not surprising that many frequentists are less than enthusiastic about adaptive designs.

A great advantage of the Bayesian approach is that at any given time the Bayesian has (or can find) the current probability for any hypothesis. And it is easy for the Bayesian to update probabilities of the various hypotheses after making an observation. An important aspect of this calculation is what it does not entail: There is no need to adjust these probabilities if the decision to take the observation depends on previous observations. This is different from frequentist calculations in which one must identify a rejection region (an "extreme" set) among the possible outcomes. The probability over any such region depends on the experimental design. So a frequentist must specify a design and follow it unwaveringly, for otherwise frequentist inferences are not possible. The Bayesian is subject to no such diligence and can even change designs in mid-trial. This makes the Bayesian approach to design more flexible than the frequentist approach. The inflexible nature of the frequentist approach slows the rate of preclinical and clinical research, which is especially noticeable with the current availability of multitudinous treatments and treatment combinations.

The Bayesian approach can lead to an efficient design. Regardless of how efficient, frequentists will want to know a design's operating characteristics (significance level and power). As indicated above, these can always be found using simulation when a rejection region is specified (whether from posterior probabilities or nominal significance levels or ad hocery). If the operating characteristics turn out to be undesirable then the design can be modified accordingly. For example, if the overall significance level is greater than 0.05 and a regulatory agency insists that that is too high, the rejection region can be made smaller to effect a smaller significance level. If the statistician adjusts the various design aspects to achieve acceptable frequentist operating characteristics, then the design is frequentist. In effect this statistician is using the Bayesian approach as a tool for developing designs with satisfactory frequentist properties.

Adaptive Designs

Adaptive designs are being used in a variety of trials at The University of Texas M. D. Anderson Cancer Center (MDACC). The adaptive aspects are varied. I will describe some of them here.

Continuous Reassessment Method in Phase I

The purpose of phase I cancer trials is to identify the maximum tolerated dose (MTD). The most commonly used phase I designs are variants of the so-called "3+3" design. Patients are admitted in groups of 3. If none experience toxicity then the dose is increased one level for the next group of 3. If 2 or 3 experience toxicity then the next lower dose is the MTD. If 1 of the 3 experiences toxicity then 3 more patients are added at the same dose level. If 2 or more of the 6 patients experience toxicity then again the next lower dose is the MTD (Dixon and Mood 1948, Wetherill 1963). This design is adaptive, but it is crude, and it is likely to assign ineffective doses and to select an MTD that is ineffective. Moreover, such a design ignores important information that accrues during the course of a trial. In particular, dose assignments are not based on sufficient statistics. An alternative approach uses Bayesian probabilities: the Continual Reassessment Method (CRM) of O'Quigley, et al. (1990). This scheme is also adaptive. Each patient is assigned to the dose having probability of toxicity closest to some predetermined target value; this probability is calculated from the data collected up to that point. The CRM more effectively and ethically finds the MTD than does the 3+3 design. The CRM is the standard design used in phase I trials at MDACC. But it is itself rather crude and we are improving it in a number of ways.

Adaptive Dose-finding in Phase II

The standard phase II dose-finding design is to allocate a fixed number of patients to each dose in a grid. In retrospect the investigators usually wish they had assigned patients in some other fashion. Perhaps the dose-response curve was shifted more to the left or right than anticipated. If so then assignment of many patients to one end or the other was a wasted effort. Or perhaps the slope of the dose-response curve is greater than anticipated and the response of patients assigned to the flat regions of the curve would have been more informative if the assigned doses were in the region where the slope is apparently greatest. Or perhaps results for the early patients made it clear that the dose-response curve was flat and that the trial could have stopped earlier. Or perhaps the results of the trial show that the standard deviation of response is greater or less than...
anticipated and so the trial should have been larger or could have been smaller.

The approach of Berry et al (2001) is to proceed sequentially, analyzing the data as it accumulates—see also Malakoff (1999). There are two stages of the trial, first dose-ranging and then a confirmatory stage, if the latter is warranted. The dose-ranging stage continues until a decision is made that the drug is not sufficiently effective to pursue future development or that the optimal dose for the confirmatory stage (phase III) is sufficiently well known. (Switches to phase III can be effected seamlessly—see below.) The example trial of Berry et al (2001) involves a neuroprotective agent for stroke. Accrual began in November 2000 and is ongoing as of August 2001. Each entering patient is assigned the dose (one of 16, including placebo) that maximizes information about the dose-response relationship, given the results observed so far. This dose can be in the region of greatest apparent slope, or it could be placebo or a high dose. But future patients are not assigned doses within a region where accumulating evidence is suggesting that the dose-response curve is flat.

In the dose-ranging stage, neither the number of patients assigned to any particular dose nor the total number of patients assigned in this stage are fixed in advance. The dose-ranging sample size will be large when the drug has marginal benefit, when the dose-response curve is gently sloping, or when the standard deviation of the responses is moderately large. It will tend to be small if the drug has substantial benefit, if the drug has no benefit, if the dose-response curve rises over a narrow range of doses, or if the standard deviation of the responses turns out to be small. (In addition, and somewhat non-intuitively, the dose-ranging stage will be small if the standard deviation of responses is very large. The reason is that a sufficiently large standard deviation implies that a very large sample size would be required to demonstrate a beneficial drug effect. The required sample size may be so large that it makes it impossible to study the drug and so the trial stops in the dose-ranging phase before substantial resources go down the drain.)

In the stroke trial considered by Berry et al (2001) the ultimate endpoint is improvement in stroke scale from baseline to 13 weeks. If the accrual rate is large then the benefit of adaptive assignment is limited by delays in obtaining endpoint information. To minimize the effects of delayed information, each patient’s stroke scale is assessed weekly between baseline and week 13. Within-patient measurements are correlated, with correlations greater if they are closer together in time. We incorporate a longitudinal model into the analysis of the trial and do Bayesian predictions (using multiple imputation) of ultimate endpoint based on current patient-specific information, and we update probability distributions of treatment effects accordingly.

Adaptive dosing is more effective than is the standard design at identifying the right dose. And it usually identifies the right dose with a smaller sample size than when using fixed dose assignments. Another advantage is that many more doses can be considered in an adaptive design. (Even though some doses will be little used and some might never be used, these cannot be predicted in advance.) An adaptive design therefore has some ability for distinguishing responses at adjacent doses and for estimating nuances of the dose-response curve.

The circumstances of the stroke trial are similar to those in many other types of trials. Finding the right dose is a ubiquitous problem in pharmaceutical development, and it is done neither well nor efficiently. The adaptive nature of the stroke trial would be less advantageous if we did not exploit early endpoints. Many diseases and conditions are characterized by the availability of information about how a patient is doing (local control of the disease, biomarkers, etc.) before reaching the primary endpoint. Finally, the possibility of moving seamlessly into phase III depending on the phase II results exists for many types of drugs.

Seamless Phases II and III

An unfortunate convention is categorizing drug development into phases. We go from one phase to the next when we think we know something: the MTD from phase I or the appropriate dose from phase II to be used in phase III. In the Bayesian approach one never regards a quantity to be perfectly known. Instead, the Bayesian carries uncertainty along with whatever knowledge is available. Phases of drug development are arbitrary labels that describe a process that is—or should be—continuous.

One of the consequences of partitioning drug development into phases is that there are delays between phases. For example, there is a pause between phases II and III to set up one or more pivotal studies. As mentioned above, the design of the stroke trial allows for avoiding such a pause. At each timepoint, say weekly, the algorithm that guides the conduct of the trial does a decision analysis and recommends either (i) continuing the dose-ranging stage of the trial, (ii) stopping the trial for lack of efficacy (inadequate slope of the dose-response curve or, more accurately, evidence of a positive dose-response that is insufficient to justify continuing the trial), or (iii) shifting into confirmatory stage. This shift can be made seamlessly, with no break in accrual. Indeed, it is theoretically possible to effect such a shift without informing the investigators: they would continue to randomize doses, but unbeknownst to them, the only doses being assigned would be the phase III dose and placebo. (Although this seamless switch is an option of the algorithm in the stroke trial, this option is not being used in the actual trial.)

At MDACC we have designed a trial (Thall et al 2001) that encompasses both phases II and III in one. If there is a switch to phase III, this switch is seamless. (The trial is in the planning stage but has not yet begun.) The disease is non-small cell lung cancer (NSCLC). The drug is Ad-p53, an adenovirus that carries a wild type of gene p53 that, it is hoped, restores programmed cell death to the tumor. The trial incorporates a single dose of the drug. (It could incorporate other doses in a manner similar to that of the stroke trial, but we have found it more acceptable in the scientific community to effect one innovation at a time.) The anticipated effect of the drug is on local control. We model sur-
vival as it depends on local control and as it depends on treatment. (Though the possibility is remote, we allow for the drug to have a beneficial effect on survival that is not mitigated by local control.) So local control is a surrogate endpoint in a way similar to the way early stroke score is a surrogate endpoint in the stroke trial. But the clear focus is on survival as the main endpoint and the utility of the surrogate endpoint must be demonstrated by the results actually observed in the trial. We exploit any relationships that exist, but do not assume that such relationships exist. We analyze the data in the trial frequently and adapt to the accruing evidence.

The seamless aspect is as follows. Initially, only MDACC patients are accrued to the trial. Think of this as phase II. If the accumulating data are sufficiently strong in suggesting that the drug has no effect on local control or survival then the trial stops. If the data suggest that the drug may have an impact on local control and that this impact translates into a survival benefit then the trial will be expanded to include other centers and the accrual rate will increase accordingly. During such an expansion, patients continue to accrue at MDACC so that there is no down time in local accrual while other centers gear up for joining the trial. This is efficient use of patient resources because the responses of patients accrued early at MDACC contribute to the eventual inferences about survival. These patient responses are the most informative of those enrolled in the two phases because their follow-up times are the longest. The trial continues until (i) stopping occurs for futility, (ii) the maximum sample size of 900 is reached, or (iii) the Bayesian predictive probability of eventually achieving statistical significance becomes sufficiently large. Should (iii) occur, accrual ceases and the drug company prepares an application for marketing approval to regulatory agencies.

The sample size of a conventional phase III trial with the desired operating characteristics is 900. We take this to be the maximum sample size in the seamless design. Actual accrual is very likely to be much less than this maximum sample size and on average it will be about half as large. On the other hand, incorporating the same number of interim analyses in a conventional design using an O’Brien-Fleming stopping boundary allows for only a slight decrease in average sample size. Under any hypothesis, null or alternative, the Bayesian design occasionally leads to a relatively large trial (close to 900 patients). However, a pleasant aspect of a Bayesian design is that the sample size is large precisely when a large trial is necessary. Conventional trials may well (and sometimes do!) come to their predetermined end with an ambiguous conclusion. In a Bayesian approach one may choose to continue such a trial to resolve the ambiguity, and this option has substantial utility. (Carrying this argument to the maximum sample size, there may be times for which stopping at 900 is ill advised, but for logistical reasons we felt the need to specify a maximum size.)

Reductions in sample size result from two characteristics of the seamless design described above. First are the frequent analyses to assess the predictive probability of eventual statistical significance. The second is the explicit modeling of the possible relationship between local control and survival. Of the two, the second is more important.

A conventional drug development strategy involves running a phase II trial that addresses local control, digesting the results, and if the results are positive, starting to develop phase III trials with survival as the primary endpoint. As indicated above, in comparison with such a strategy, a seamless approach can greatly reduce sample size. In addition, a seamless design minimizes pauses between phases and so the total drug development time is greatly shortened.

**Adaptive Allocation**

The adaptive designs discussed so far are motivated by the desire to learn as efficiently and as rapidly as possible. Another kind of adaptive design aims to treat patients as effectively as possible. These designs use adaptive allocation in which patients are more likely to be assigned to treatments that are performing better. In addition to making clinical trials more attractive to patients and thereby increasing participation in clinical trials, such strategies have the interesting side effect of efficient and rapid learning!

As of this writing more than a dozen trials at MDACC have been designed and are being conducted using adaptive allocation. Our standard approach is to randomize treatment assignment, but to shift the weights toward better performing arms as the trial proceeds and the results accumulate. Many of these trials have more than two arms. The arms are sometimes distinct therapies, and sometimes they are closely related. An example of the latter is an MDACC trial involving five doses (including 0) of a drug (pentostatin). The goal is to inhibit graft-versus-host-disease (GVHD) in leukemia patients who are receiving bone marrow transplants. The problem is that the drug may inhibit successful engraftment of the transplant, and such inhibition may be related to dose. (Engraftment is essential for survival.) We use a combination endpoint: survival at 100 days free of GVHD. The conflict between engraftment and freedom from GVHD means that the dose-response curve may not be monotone. In particular, it may increase for small doses and then decrease. Initially we assign doses in a graduated fashion, climbing the dose ladder slowly. But as doses become admissible, we assign patients to those that have been performing well.

Consider a patient who qualifies for the trial. To decide which pentostatin dose to assign we calculate the current (Bayesian) probabilities that each admissible dose is better than placebo. This calculation uses all available information. We allocate doses randomly, with weights proportional to these probabilities. (The first version of this idea is almost 70 years old, dating to Thompson [1933]. An historical footnote is that Thompson’s 20-page paper focused almost exclusively on the computation problem of evaluating the probability that $\theta_1 > \theta_2$ where these parameters have independent beta distributions. The era of computers renders such a calculation trivial, but in Thompson’s time it was an enormous hurdle.) We consider other allocation algorithms, including assigning in proportion to powers of

(continued)
these probabilities. The assignments we consider involve some amount of randomization, but patients are more likely to receive doses that are performing better. Doses that are doing sufficiently poorly become inadmissible in the sense that their assignment weight becomes 0. When and if we learn that the drug is effective, we stop the trial. When and if we learn that the drug is ineffective then again we stop the trial. Patients in the trial benefit from data collected in the trial. The explicit goal is to treat patients more effectively, but a happy side effect is that we learn efficiently. We evaluate each design’s frequentist operating characteristics, which are quite acceptable. Although such a design is radical in medical research, one day we will look back and wonder how we could ever have done otherwise.

Thompson-like designs are similar to play-the-winner rules in that their goal is to treat patients in the trial more effectively. They are also similar in that both are ad hoc and neither is optimal in any sense that I know. Various optimality criteria have been suggested for evaluating adaptive trials. The most common is maximizing an expected sum of observations over the course of the trial, such as the total number of successes, possibly with later observations discounted relative to the next observation (Berry 1972, 1978, Berry and Fristedt 1985). Decision problems of this type are called bandits. For example, two-armed bandits involve two treatments. Finding optimal bandit strategies means solving dynamic programming problems. The solutions can require intensive computation, but finding on-line solutions during a trial is not out of the question, even for some rather complex settings.

A characteristic of optimal strategies for commonly considered objectives in bandit problems is that they are deterministic. One of the vagaries of clinical trials is that the prognosis of patients may fluctuate over the trial’s course. This would not be a serious problem if prognoses were well understood and easy to measure. Usually, neither is true. Or more precisely, we seldom know whether they are true. A treatment may do relatively well during a period when it happened to be assigned to patients who have relatively good prognoses. When following a deterministic strategy, such a treatment may be used for some number of patients in a row and its performance may therefore be artificially inflated. An easy fix is to mix a randomized assignment with an optimal assignment: assign the optimal arm with probability r and an arm chosen randomly with probability 1−r. At MDACC we have not yet progressed to this point, however, we recognize the importance of having some level of randomization, which is why we use Thompson-like strategies.

**Process or Trial? Evaluating Many Drugs Simultaneously Using Adaptive Allocation**

The greatest room for innovation and for improving drug development is effectively dealing with the enormous numbers of potential drugs that are available for development. The notion of processing drugs one at a time is ingrained in the pharmaceutical culture. This notion has to change. Companies that are able to process many drugs simultaneously and do so effectively will survive and others will not.

Many different drugs should be evaluated in the same preclinical experiment or collection of experiments. Information should be updated frequently or even continually. The extent to which any particular drug is used and the order of drugs used will depend on the available data. Drugs that are apparently more promising will move faster through the preclinical setting. Drugs that give disappointing data will languish. And the sample sizes of drugs whose promises and toxicities are not clear will tend to be large so as to better resolve uncertainties.

These ideas and imperatives apply as well to clinical trials. As an example, at MDACC we are building the foundation for a phase II trial for evaluating drugs that is more a process than a trial. The idea is a straightforward extension of the adaptive assignment strategies described above. We start with a number of treatment arms plus a control—possibly a standard therapy. We randomize to the arms and learn about their relative efficacy as we go. Arms that perform better get used more often. An arm that does poorly enough gets dropped. An arm that does well enough graduates to phase III, and if it does sufficiently well it might even replace the control. As more arms become available, we add them to the mix.

The result is that better arms move through quickly and poorer arms get dropped. An advantage to patients in the trial is that they are provided with better treatment (when the arms are not equally good, or equally bad). The advantage to patients outside the trial is that they get access to better drugs more rapidly.

**Discussion and Conclusion**

There are obvious benefits from updating one’s state of knowledge as relevant evidence accumulates and proceeding on the basis of full information. So why are adaptive designs not more common? The answer has to do with the dominance of the frequentist approach in the history of medical research. The Bayesian view has made occasional inroads into attitudes among medical researchers, but only recently has its influence been felt in pharmaceutical development. Changes will not be drastic or immediate. In the next few years we will see Bayesian approaches used increasingly, but as a tool, with justifications following a more or less traditional frequentist course. As time passes and as researchers and regulatory folk become more accustomed to Bayesian ideas, they will be accepted increasingly on their own terms.

What do regulatory agencies think about all of this? A good source for the FDA’s view of using Bayesian methods is Farr-Jones (2001). There is no single FDA view, of course. The Center for Drug Evaluation and Research (CDER) tends to be more conservative than the Center for Biologics Evaluation and Research (CBER) or the Center for Devices and Radiological Health (CDRH). Even the CDER is open-minded. But Farr-Jones quotes the CDER’s Robert O’Neill as indicating that any submission to the CDER will require frequentist analyses to support Bayesian analyses. This view is consistent with the near-future view that I sug-
gested above. The far future for the CDER exists today in the CDRH, where Bayesian designs and analyses are accepted as standing on their own. The danger is that small device companies with little or no statistical sophistication get on the bandwagon and put forward scientifically inferior designs and analyses under a Bayesian banner. There is as much Bayesian junk as there is frequentist junk. Actually, there's probably more of the former because, to the uninitiated, the Bayesian approach seems like it provides a free lunch. For example, a novice Bayesian may accept historical patient responses as being exchangeable with current patient responses. Depending on the circumstances this may be lousy science. (Whether historical patient responses can be combined into a single analysis with current patient responses and how to effect such a combination is among the most difficult of problems facing a Bayesian).

The FDA's attitude toward adaptive assignment strategies and seamless phase II/III designs is less clear than is their attitude toward Bayesian analyses generally. I am heartened by reactions from individual FDA personnel, but some adaptive ideas are sufficiently radical that they will become generally accepted only after they have been used successfully in a variety of settings. The stroke and NSCLC trials and the MDACC trials mentioned above are first steps.

The stroke and NSCLC trials and the MDACC trials mentioned above are first steps.

References

Thall P, Inoue LYT, Berry DA (2001). Seamlessly expanding a randomized phase II trial to phase III. (Submitted for publication.)

Some comments on "Adaptive Trials and Bayesian Statistics in Drug Development"
by Donald A. Berry

George Chi,
H.M. James Hung,
and Robert O'Neill

We welcome the opportunity to comment on Donald Berry's stimulating article 'Adaptive trials and Bayesian statistics in drug development'. Dr. Berry addresses several major themes and we do not intend to comment in depth on all of them. These themes cover, for instance, the contrast between the so called 'static trials' and flexible trials; the contrast between frequentist and Bayesian approaches in the interpretation of probabilities and clinical trial results; the contrast between adaptive designs as they relate to seamless phase II and III designs; the use of designs for more efficient drug discovery of multiple candidate drugs; and briefly discussed is his perception of the view of regulatory agencies on all this.

First, a comment on the learning and confirming paradigm as it is practiced in drug development. The concept of learning as the data are accumulating is natural and appealing. It has been key to most, if not all, human discoveries. Drug discovery is certainly not an exception. Statistical methods applied in this process have been so rich that it is often difficult to find a perfect match between the practical case problem in hand and methods/designs. Examples are plenty, such as re-testing after initial testing, continuous model correcting, analysis results being repeatedly interpolated and extrapolated in view of past analysis results, etc. All these are critical components of learning.

In drug development, aside from the screening of many multiple candidates, sponsors focus on a specific drug and begin to craft a clinical hypothesis or multiple hypotheses that eventually are intended to support claimed indications for the product. For drug efficacy, the basic hypothesis is one that the experimental drug is effective at some dose(s) in targeted patient populations under proposed conditions of use. This is certainly not the only hypothesis. The clinical questions of interest involve many aspects: treatment effects on what endpoints; at what dose; what is the effective dose to be recommended for the average patients; what
is the duration of effect and when and if the effect is lost, etc. The list can go on as more questions are explored or advanced. The use of a single Phase III trial with one comparison whether it is called confirmatory or not is too simplistic and indeed the number of studies needed to convincingly demonstrate a product’s efficacy is challenging to plan. Thus, informative Phase I and II studies are vital not only to save resources by stopping futile studies but also to provide more accurate information for the planners to use better design parameters in designing Phase III trials.

It is surprising how many diseases and therapies proposed to treat them are so difficult to characterize from an endpoint perspective, even at a late stage of phase III confirmatory trials. These trials often are still in the learning phase for poorly understood diseases and responses to characterize them. Successful confirmatory trials are critical to the establishment of substantial evidence of efficacy. While the goal at later phases of drug development is to be parsimonious in selection of endpoint(s) that characterize the disease and treatment response, it is surprising in many situations how problematic it is to demonstrate a convincing treatment effect, or to demonstrate these treatment effects repetitively in several trials. All appropriate strategies that would help to maximize the learning phase and better utilize its findings to minimize failed trials in the confirmatory phases and avoid inconclusive statistical evidence are welcome. Suffice it to say that flexibility in trial design is a welcome goal if later confirmatory trials have increased chances of success.

The opportunities for the greatest flexibility in study design exist in the earlier phases of drug development, even in the initial randomized controlled efficacy trials. These trials are not just exploratory in nature, and may have confirmatory components to them in that their evidence is considered seriously in conjunction with later phase confirmatory trials. Non-static, or outcome dependent sample size sensitive trials have a place here.

Even now, many trials in the later stages of drug development are not static. In fact, there is an increasing use, especially for drugs to treat serious diseases with life threatening morbidity and mortality outcomes, of study designs utilizing interim analysis and sequential monitoring strategies, and data monitoring boards charged with implementation of the accruing results. There is a rapidly developing body of flexible designs that address many of the questions touched on by Dr. Berry, but these have been approached from a frequentist perspective [see below]. Indeed, the concern of some is not so much the technical limitations to methodology for adaptive flexible designs as the practical aspects of operational implementation of them so as to not introduce bias. Bias comes about because of operational situations such as patient unblinding, sharing of comparative interim treatment results, design changes to parties that may adversely influence current or future trial results, etc. Adaptive allocation in these trials has its own set of concerns.

Methods have been proposed by various authors to deal with these problems within the context of the frequentist approach for both the group sequential trial setting and the traditional fixed sample size setting [Bauer, Brannath, Posch (2001)]. For sample size adjustment, see the references, such as, Bauer (1994), Proschan & Hunsberger (1995), Chi & Liu (1999), and Liu & Chi (2001) for the two-stage setting, and Cui, Hung, & Wang (2000), Wang, Hung, Tsong, & Cui (2001), Jennison and Turnbull (2001) in the group sequential setting. These authors have considered combining certain phase II and phase III trials within the context of a confirmatory trial from the frequentist perspective. Yang, Sridhara, Chen & Chi (2001), Sridhara, Yang, Chen and Chi (2001) and Shi, Ouyang, Quan, & Lin (2001) have recently proposed combining certain phase II and phase III cancer trials for accelerated approval of new drug products. There are many other relevant references not cited here. Most of these new methodologies have been proposed under the frequentist approach. This is because such combined phase II and phase III trials under the heading of two-stage adaptive design are proposed as confirmatory in nature.

Dr. Berry suggests that the adaptive approach to phase III trials generally reduces the average sample size needed when compared to the traditional static design. The alleged saving in sample size suggested is misleading if he compares so-called Bayesian adaptive design strategies to the traditional static design. Certainly, for some situations there will be savings in sample size for trials adopting the Bayesian adaptive design strategies, but these savings need to be compared with the frequentist adaptive design strategies referenced above.

Despite the arguments to the contrary, the frequentist approach to the design and analysis of confirmatory trials has many tried and tested advantages, and the consideration of prospective use of frequentist strategies within the flexible design paradigm attests to this. The frequentist approach is less assumption dependent and can provide the statistical strength of evidence required for a confirmatory trial that may be lacking in a more assumption dependent Bayesian approach. As an additional and useful perspective for interpretation of uncertainty, the Bayesian approach has a place but it could be argued that it does not usually add much to the decisions when sufficient amounts of data are available.

Phase II encompasses that area of clinical trials, usually randomized, that are intended to demonstrate some aspect of efficacy of a drug. Phase III is also intended to demonstrate the efficacy of a medical product, but often under a more extensive set of circumstances. There is not as much a distinction between these phases as Dr. Berry claims. The extent to which the learn-confirm paradigm is not being followed because Phase III trials are being conducted simultaneously with Phase II is a problem of proper planning. The idea of seamless Phase II-III trials is worth exploring, but it should be made clear that the total evidence from such an approach may not reach the level of at least two independent trials each convincingly demonstrating a treatment effect, a usual standard for efficacy. Better use of a Phase II trial as it merges into a Phase III trial may
be sensible for some drugs and diseases where the end-points are usually the same in both phases of study.

Dr. Berry makes mention of stroke trials that are currently being carried out at MDACC. His seamless trial design appears problematic if the MDACC patients are not representative of the target stroke patient population in general, and if failure to detect a treatment effect at MDACC prevents an otherwise effective treatment from being studied, or cause delay in the treatment being marketed in the future. In other words, if a new treatment is deemed not effective locally, then it may substantially decrease its chances of ever being proved effective.

Again, if the proof of efficacy is in the confirmatory second stage, then it may not matter how the first stage is done as long as the final analysis is based only on data from the second stage. But if this is the case, then we don't see the advantage of the so-called two-stage design in this stroke trial.

Given that all the ten or so stroke trials in the recent past have ended in failure, it is doubtful that the exploratory analysis done at the first stage of this trial would produce a positive confirmatory second stage result. And that a single confirmatory study would be convincing evidence of efficacy. It is important to better understand why all these trials have failed.

The use of adaptive allocation is much less than perhaps it should be, but there are some reasons. Treatments with variable and very modest effects are not best tested under outcome dependent adaptive allocations, because imbalances in patient covariates and other potential predictors, or poorly understood relationships associated with the assignment mechanism may be just as responsible for observed effects as the treatment itself. Moreover, it is not just because of the difficulty in computing p-values or controversies in quantifying strength of statistical evidence from such allocation schemes, but rather the potential operational biases associated with design implementation. When the allocation ratio is lopsided between comparative groups and it may be known as part of the design, operational biases can creep in because of educated guesses about the relationship of responses and treatment unless the study endpoint can only be measured by an objective means, such as, death. The concern for operational biases in confirmatory trials should never be dismissed, as evidenced by emphasis on the need for valid blinded endpoint adjudication processes.

Difficulty in the quantification of strength of statistical evidence in adaptive designs should not be an excuse for less usage. On the other hand, operating procedures need to be established with such designs to avoid biases and research is needed for this purpose. Moreover, for any design there needs to be a means to assess the sensitivity of the results of analysis to the expected or unexpected auxiliary events (some of them can be detrimental), such as, differential or informative dropouts or censoring, that is critical to the interpretation of the study results. Adaptive design is certainly not an exception to this concern which impact static designs but may be easier to deal with here.

CDER's perspective on the utility of adaptive designs and Bayesian approaches is based to a large degree on its experience with implementing regulations for evidence of efficacy. Bayesian thinking is involved in many regulatory decisions, but it is not formal Bayesian methods that are usually relied on to shed light on such decisions. Frequentist methods applied to empirical data in order to quantify uncertainty in the evidence and to quantify measures of treatment effect have served the community well. Frequentist adaptive designs can also have appeal. Bayesian strategies offer a helpful additional perspective used in decision making. Safety assessment is one area where frequentist strategies have been less applicable. Perhaps Bayesian approaches in this area have more promise.

**Selected References:**


**Disclaimer:** The views expressed in this commentary are those of the authors and not necessarily those of the Food and Drug Administration.
Rejoinder to discussants
George Chi, H.M. James Hung and Robert O’Neill

by Donald A. Berry

I want to thank Drs. Chi, Hung and O’Neill for discussing my article and for their erudite comments. I agree with most of what they say. I find their optimistic comments about using adaptive designs particularly encouraging. I will react to some of their comments here and I will clarify some things I did not say well in my article.

The discussants indicate that "Certainly, for some situations there will be savings in sample size for trials adopting the Bayesian adaptive design strategies, but these savings need to [be] compared with the frequentist adaptive design strategies referenced above." I agree. But the reader should not come away with the notion that frequentists have done it all (any more than they should think that Bayesians have done it all or could do it all). The references in question are rather tame in the context that I tried to present. Stopping trials early and expanding trials based on accumulating information are not revolutionary ideas (although the latter is seldom used). None of the references include adaptive allocation, dropping arms, adding arms, combining arms, etc., while a trial is ongoing.

I agree with the discussants’ concerns about operational biases. These are not insurmountable and can be appropriately addressed, and I think the discussants would agree with this. We need experience in addressing these biases but we will be able to get it right.

Regarding "The idea of seamless Phase II-III trials is worth exploring, but it should be made clear that the total evidence from such an approach may not reach the level of at least two independent trials each convincingly demonstrating a treatment effect, a usual standard for efficacy." I cannot disagree with "may not reach." But such trials can be designed so that they do reach such a level. They continue: "Better use of a Phase II trial as it merges into a Phase III trial may be sensible for some drugs and diseases where the endpoints are usually the same in both phases of study." In my view it is unwise and inefficient to change endpoints from one phase to the next. Consider the seamless phase II/III study of Ad-p53 in non-small cell lung cancer (NSCLC) that I described in my article. We consider both local control and survival throughout the trial, with the observed relationship between the two serving to augment information about the ultimate endpoint, which is survival. The information about this relationship enables early decisions and promotes efficiency. Now of course this information is sparse, and the algorithm implicitly recognizes this, but there is some information, and it is important to use that information in deciding how best to proceed.

The discussants refer to the importance of extending clinical investigation beyond a local setting. I agree. (I should correct them regarding stroke trials and MDACC. We run cancer trials and not stroke trials at M. D. Anderson Cancer Center. I designed the stroke trial described in my article for a pharmaceutical company and MDACC was not involved.) Indeed, the NSCLC trial that I described starts out at MDACC and, depending on the accumulating data, expands into other centers. The discussants are concerned that the inability to show that a drug is effective locally will "substantially decrease its chances of ever being proved effective." No doubt this is true. But the concern does not seem very important to me. The opposite concern is more worrisome: treatments shown to have benefit in one locale may fail to show benefit more generally. However, neither concern holds greater force in the context of adaptive methods than when using more conventional methods.

The discussants say "we don't see the advantage of the so-called two-stage design in this stroke trial." In view of the parenthetical remark in the previous paragraph, I am not sure whether they’re talking about the stroke trial or the MDACC trial in NSCLC. I’ll address both. The point of the first stage of the stroke trial is to identify the appropriate dose for carrying into the second, confirmatory stage. The point of the first stage of the NSCLC trial is to proceed cautiously (phase II) until the available data suggest that the drug has a benefit on local control that likely translates into a survival advantage. Too many drug development programs blast through phase III only to come up empty. A goal of the stagewise process is to ramp up development (phase III) when the data are suggesting that it is appropriate, to bail out when the data are sufficiently discouraging, and otherwise to continue investigating in a cautious mode. An additional advantage is that the size of the second stage is informed by the results from the first stage.

Regarding the sorry state of clinical trials in stroke, I am not suggesting that adaptive designs can convert a sow’s ear into a silk purse. A lousy drug is a lousy drug. But I am saying that adaptive designs are good at recognizing a sow’s ear because it continually examines the object in question. I am also saying that finding the right dose or the right dose schedule or the right combination therapy is easier if the early stages of a trial investigate a variety of therapeutic options, again adaptively.

The discussants have correctly seen that my article is not about differences between the Bayesian and frequentist philosophies. In most of what I described, I used the Bayesian approach as a tool in building adaptive designs that have good frequentist properties. However, the discussants do contrast the two approaches. I am pleased to see their appreciation of using Bayesian methods in decision analysis. And I cannot disagree with "As an additional and useful perspective for interpretation of uncertainty, the Bayesian approach has a place but it could be argued that it does not usually add much to the decisions when sufficient amounts of data are available." But they also say "The frequentist approach is less assumption dependent and can provide the statistical strength of evidence required for a confirmatory trial that may be lacking in a more assumption dependent Bayesian approach." I disagree with this, and emphatically so. Indeed, I can make arguments that the opposite is true. I am reasonably well read in statistical

(continued)
philosophy, but this purported difference between the Bayesian and frequentist approaches is new to me.

In defense of the frequentist approach, the discussants indicate that "Frequentist methods applied to empirical data in order to quantify uncertainty in the evidence and to quantify measures of treatment effect have served the community well." I don't have major objections to this statement, and in fact I think frequentist statistics deserves full credit for putting science into medical research. But the statement gives pause to the statistician in me. What's the control? "Well" compared with what? Could the community have been better served? More to the point, can the community be better served today? We are entering an era of drug development with a gazillion drugs to investigate. Is the old-fashioned two-arm, balanced, highly powered study appropriate today? I submit not. Will we miss out on some good drugs if we stray from today's norm? Of course we will, but we'll miss good drugs regardless of how we proceed. The problem is that we're missing out on some great drugs that are sitting in the queue waiting their turn at clinical investigation. It is as much a crime to not investigate a good drug as it is to throw it out after an inadequately powered examination. An adaptive procedure that considers 100 drugs in 2,000 patients is much better at finding a 1-in-a-100 drug than is a fixed procedure that assigns 200 patients to each of 10 drugs. Indeed, a drug that is sufficiently good will be found for sure by an adaptive procedure but will be lost 90% of the time by an old-fashioned procedure.

Again, I very much appreciate the discussants' comments. I look forward with great anticipation to regulatory agencies working with industry and academia in developing new and better designs for clinical trials.

---

**Section News**

**Letter from the Chair**

**Jeff Meeker**  
*Chair*

**To our past Corporate Members:**

The Biopharmaceutical Section developed a corporate membership soon after the ASA Constitution change that created separate funds for each Section. Because of concerns by ASA that they would loose some of their own corporate members, it was agreed that one requirement for membership in the Biopharmaceutical Section would be corporate membership in ASA. Over the years, we have had a varying number of corporate members, ending with seven in 2000. We recently found that at some point, the Committee on Membership decided to eliminate corporate memberships for all Sections and Chapters. As the only group effected by that decision, we were not asked for input and no one informed us of the decision after the fact. We found out when one company, trying to join our Section, had difficulty in enrolled in the Section. We contacted five of the seven corporate members we had at the latest listing. They also did not know about it, although none of them had paid the Section's dues for 2001. Based on the seven corporate members we had, this decision cost us $2000 income per year.

ASA is now allowing as a replacement an annual "sponsorship" under the Development policy approved last year by the board. We would be free to solicit what funds we could (under $2000 per sponsor); regardless of whether the organization was a corporate member of ASA. The only requirement is that this be totally administered by the Section; ASA staff will not help. We can provide sponsors with whatever benefits are allowable under current tax laws, such as a list on our Section's web pages, publication in Biopharmaceutical Report, or posting elsewhere.

The Section Executive Committee intends to develop such a program during the early part of next year. Anyone wishing to provide input may contact one of the Section's officers.

**To All—Thanks.**

This has been a banner year for our Section in many ways and I would like to thank everyone for their hard work. Most importantly, Keith Soper, our 2001 program chair; Len Oppenheimer, our 2002 program chair who is putting together an excellent program for next year; Greg Enas and Anna Nevius for the 2001 work shop (which is delayed until January, 2002); Kathy Monti for the round tables at JSM; Tom Bradstreet and his committee for the Student Paper Competition; and Anne Cross and her helpers for the Contributed Paper Awards. This group put together an excellent scientific program.

On the publication side, thanks to Demissie Alemayehu, our Publications Officer, Kannan Natarajan, Biopharmaceutical Report editor; Webmaster Kalyan Ghosh; and mail list coordinator Sally Greenberg.

Administratively, chair-elect Bob Small and past chair Tom Capizzi have provided excellent advice; secretary/treasurer Sally Greenberg and Ram Suresh have taken care of the administrative details on which the Section depends; and Executive Committee members Lukas Makris, David Carlin, Kathy Monti, Greg Campbell, Anne Cross, and Tom Bradstreet provided thoughtful input. Council of Sections representatives Nancy Smith, Ralph Harkins, and Avital Cnaan made sure our input was made. Add to this list Dave Carlin, our Membership Committee chair, and Bob Davis, our Fellows Nomination Committee chair.

There are many others who helped as committee members, session organizers and chairs and several of the other usually thankless jobs within the Section. For everyone's efforts, thank you.
Highlights of Executive Committee Meeting

Tuesday, August 7, 2001
Atlanta, GA

Jeff Meeker
Chair

- The 2001 JSM program included 5 invited sessions, 6 topic contributed sessions, 9 regular contributed sessions, and 9 luncheon roundtable discussions. The Executive Committee noted that many of the sessions were in rooms that were too small, particularly the topic contributed sessions.
- The Section awarded four Student Paper Awards for 2001. The 2000 Best Contributed Paper Awards were also awarded.
- The FDA/Industry Workshop is September 24-25 in Bethesda, Maryland.
- The section will have 3 invited sessions at the 2002 ENAR meetings in Washington, D.C.
- The Section expects to have 4-5 invited sessions at the 2002 JSM in New York.
- Dues were increased from $7 to $8 for 2002 for regular members. Student membership will remain at $1.
- The first issue of Biopharmaceutical Report was to be mailed the first week of August and features an article on Population PK. The second issue is scheduled for October. Proposals to switch to online publication are being developed.
- The 2001 Proceedings will be issued on a CD-Rom. The cost is $25 for all Proceedings, not just the Biopharmaceutical Section’s. It was reported that the Section will realize either a reduced profit from the Proceedings or possibly even a loss. The problems seem to be no real ability to estimate the number of Proceedings that will be sold and an underestimate of the cost. Estimates as high as $28 per copy were reported. If the cost is greater that $25 per copy, the Sections will have to pick up the loss.
- A proposal for a new survey of Section members was rejected due to the high cost. It was returned to the Membership Committee to develop a new proposal.

Highlights Executive Committee Meeting

Tuesday, October 23, 2001
Alexandria, Virginia

Jeff Meeker
Chair

- The 2002 budget was finalized. Unfortunately, there is a significant shortage for the Industry/FDA workshop, which will require an increase in registration fees.
- We recently discovered the ASA Committee on Membership had voted to eliminate Corporate Membership in Sections and Chapters, beginning in 2001. We were the only group to have such a membership, but neither we nor our Corporate Members were informed of the action. This has resulted in a loss of income of approximately $2000 per year, based on the last year we had Corporate Members. However, ASA has used the Development Procedures approved last year by the ASA Board to allow Sections to develop a "Sponsorship" program. The advantage to us is there is no longer a requirement for the Sponsor to be a Corporate Member of ASA. However, ASA has indicated the program would have to be totally our responsibility and they will not support it. The Executive Committee will develop procedures during 2002.
- The Executive Committee discussed issues with rescheduling the FDA/Industry Workshop to January, 2002. We were assured by ASA staff there will not be a significant impact on the costs of the Workshop as a result of the change.
- The Executive Committee approved a change in the paper used in the Biopharmaceutical Report to a less expensive paper.
- The Executive Committee agreed to provide support to the Merck/Temple Conference, similar to other conferences we support (Deming Applied Statistics Conference and Mid-West Biopharmaceutical Workshop). However, we would not be listed as a cosponsor. There would be no financial involvement of the Section.
- Changes were approved in the Manual of Operations. These primarily involve a stronger structure for the Membership Committee and the recreation of the Education Committee. The old Education Committee had become the Industry/FDA Workshop Committee.
- ASA staff demonstrated a new financial and membership information system that can be accessed by the Chair and Secretary/Treasurer. It provides up-to-date information which should significantly improve our ability to monitor both areas.
Program Chair Report

Len Oppenheimer

ENAR

Joanna Shih (NIH) is the 2002 Program Chair for the Spring 2002 ENAR meetings in Alexandria, VA. There were 28 available/competition Invited Paper sessions. Four sessions were submitted and three sessions were accepted. The accepted sessions were as follows:

1. "Some Statisticians' Perspectives on FDA Advisory Committee Meetings." A panel discussion organized by Janet Wittes (Statistics Collaborative)

2. "Decision Analysis in the Pharmaceutical Industry* organized by Jerry Nedelman (Novartis)

3. "Statistical Issues in the design and Analysis of Extensions to Clinical Trials* organized by Matilde Sanchez (Merck Research Laboratories)

JSM

The 2002 meeting will be held August 11 - 15 in New York City. The theme for the meeting is "Statistics in an Era of Technological Change."

Ten Invited Paper session proposals have been submitted for consideration for sponsorship by the Biopharmaceutical Section. The section has 4 allocated sessions and can also submit 2 sessions for the "competition slots". Three additional competition slots* will be allocated. Final decisions on the selections will be made by the end of October.

Deadlines for all abstracts is February 1, 2002. Topic Contributed Paper sessions are still being solicited. All such sessions are guaranteed acceptance. In addition volunteers are being solicited to chair sessions.

Short Courses

Diane Fairclough has tentatively agreed to teach a short course on "Quality of Life: Design and Analysis of Clinical Trials." She will be publishing a CRC Press book on this topic within the next few months.

Attempts are also underway to find a second potential short course.

2002 JSM Short Courses, Topic Contributed Paper Sessions, Chairing a Session

It's not too early to start thinking about next year's Joint Statistics Meetings in New York City (August 11-15, 2002).

There's still plenty of time to organize a Topic Contributed Paper Session, a collection of contributed paper presentations and discussions (if desired) that share a common topic. There are five presentations and or discussants, each 20 minutes long. As with all contributed papers -- abstracts are due by February 1, 2002.

In addition volunteers are need to chair both contributed and topic contributed paper sessions.

Suggestions for topics/instructors for short courses that would be of interest to Biopharmaceutical Section members are welcomed.

Please contact Len Oppenheimer 2002 Program Chair for the Biopharmaceutical Section if you would like to (i) organize a Topic Contributed Paper Session, (ii) would like to chair a session, (iii) would like to teach or attend a specific short course. He can be reached as follows:

Len Oppenheimer
Merck Research Laboratories
P.O. Box 2000, RY 33-408
Rahway, NJ 07065
732-594-5490
leonard_oppenheimer@merck.com

CORPORATE MEMBERSHIP

The following are the current Corporate Members of the Biopharmaceutical Section. To become a Corporate Member of the Biopharmaceutical Section, the organization must first be a Corporate Member of the ASA. These members provide funds that allow the Section to provide services to all our members. The Biopharmaceutical Section gratefully acknowledges their support.

• Warner-Lambert Company-MI, Ann Arbor, MI
• The Cambridge Group Ltd, Westport, CT
• Sugen Inc, South San Francisco, CA
• Trilogy Consulting Corporation, Waukegan, IL
• Statistical Solutions Ltd, Cork, Ireland
• Scirex Corporation, Bloomingdale, IL
• R.W. Johnson PRI, Raritan, NJ
• Merck & Company Inc PA, West Point, PA
• Pfizer Inc, Groton, CT

For individual membership list, please visit us at: http://www.best.com/~asabp/
Midwest Biopharmaceutical Statistics Workshop Meeting Report

May 21-23, 2001
Stacy David

The twenty-fourth annual Midwest Biopharmaceutical Statistics Workshop was held in Muncie, Indiana on May 21-23, 2001. The number of participants was 167. The majority of the participants were from biopharmaceutical sector while the remaining participants were from academia (faculty and students).

Dr. Frank Rockhold, Senior Vice President and Director, Biomedical Data Sciences in GlaxoSmithKline Pharmaceuticals Research and Development, was the Tuesday night banquet speaker. He touched on the history & future of statistics in the Biopharmaceutical Industry and contrasted public opinion of the field with other industries (e.g., Tobacco).

A short course was offered for the 2nd year at the workshop. Dr. Joe Schaefer from the Pennsylvania State University presented a course entitled, "An Overview of Multiple Imputation". Later the same afternoon, Dr. Richard De Veaux from Williams College lead the plenary session on Data Mining. Both sessions received high praises on the evaluation forms.

Two concurrent sessions and one joint session were held during the workshop. On Tuesday morning, May 22, sessions on "Current Issues in Clinical Pharmacology Statistics", "Data Mining" and "Current Issues in Statistical Applications for Manufacturing and Development" were held. A Memorial session was held on Tuesday afternoon in memory of Robert P. Rathmacher. Talks during this session spanned Methods for Safety Assessment in Clinical and Preclinical Studies. The Wednesday morning sessions included: "Gene Expression Analysis via DNA Microarrays", "Post-Marketing Surveillance, Pharmacovigilance, Phase IV Studies" and "Health Economic and Outcomes Research".

Around 10 participants, including students from neighboring universities, represented the poster sessions. The poster session was well received and the first Charlie Sampson Poster Award was presented to Xiaohua Zhang for his poster on "Cancer Prediction Using Gene Expression Data". This award is presented to the student with the poster judged as best based on content, originality, and contribution to biopharmaceutical issues. Xiaohua received a plaque and a monetary award.

The workshop committee was chaired by Stacy R. David, Lilly Research Laboratories. The other members of the committee were Ken Gerald, past chair, Applied Logic Associates, Inc.; Mike Lutz, GlaxoSmithKline; Dhammika Amaratunga, R W John Pharmaceutical Research, and Mohammad Hoseyni, Proctor & Gamble Pharmaceuticals, as program co-chairs; Linda Leonard, Pharmacia Corporation, poster session chair; Charlie Sampson, retired from Eli Lilly, chairman emeritus; Amanda Clancy & Amy Rosen, Eli Lilly, registrars; Kerry Barker, Pharmacia Corporation, treasurer; Mir Ali, Ball State University, local arrangements chair; and John Schollenberger, Quintiles, publicity.

Overall, the workshop was well received by those filling out program evaluation forms. Over 96% of the respondents felt that the topics were relevant and that the workshop had the proper mix of topics.

The twenty-fifth annual Midwest Biopharmaceutical Statistics Workshop will be held in Muncie, Indiana on May 20-22, 2001. Stan Young from GlaxoSmithKline and Walt Offen from Eli Lilly will be serving as the workshop committee co-chairs.
March 1, 2002
Hotel reservation form available online
April 1, 2002
Final audiovisual requirements deadline
April 1–May 1, 2002
Revisions can be made to abstracts online
May 1, 2002
Preliminary Program available online
May 15, 2002
Registration materials available online
June 1, 2002
Draft manuscripts due to session chairs for all regular and topic contributed papers and any invited papers with discussants
July 5, 2002
Last day for early bird registration forms to arrive at the office
July 6, 2002
Advanced registration fees apply
July 19, 2002
Hotel reservations and advanced registration deadlines
July 20, 2002
On site registration fees apply

Joint Statistical Meetings
Get READY for HOT FUN in NEW YORK CITY

August 11-15

New York HILTON & TOWERS
SHERATON New York Hotel
Let’s Hear from You!

If you have any comments or contributions, contact Editor Kannan Natarajan, Director, Biostatistics, Bristol-Myers Squibb PRI, P.O. Box 5400, Princeton, NJ 08543; Phone 609-818-4299; email: kannan.natara-jan@bms.com; Associate Editor Neal Thomas, Principal Statistician, Biostatistics, Bristol-Myers Squibb PRI, 5 Research Parkway, Wallingford, CT 06492; Phone 203-677-7270; email: neal.thomas@bms.com; or Past Editor Demissie Alemayehu, Biometrics Director, 235 East 42nd Street, Bldg 205/4, Pfizer Inc, New York, NY 10017; Phone 212-573 2084; e-mail: alemd@pfizer.com; or Past Editor Demissie Alemayehu, Biometrics Director, 235 East 42nd Street, Bldg 205/4, Pfizer Inc, New York, NY 10017; Phone 212-573 2084; e-mail: alemd@pfizer.com;

The Biopharmaceutical Report is a publication of the Biopharmaceutical Section of the American Statistical Association.

The Biopharmaceutical Report is a publication of the Biopharmaceutical Section of the American Statistical Association.

© 2001 The American Statistical Association
Printed in the United States of America