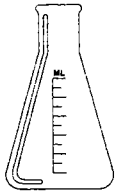


20

Biopharmaceutical Section



American Statistical Association

Biopharmaceutical Report

Volume 3, No. 1

Summer 1995

Chair: *Lilliam Kingsbury*

Co-Editors: *Curt Wiltse and Bill Huster*

Adaptive Designs in Clinical Trials: Lilly Experience

Douglas E. Faries, Roy N. Tamura, and John S. Andersen

Lilly Research Laboratories, A Division of Eli Lilly and Company, Indianapolis, IN 46285

1. Introduction

Any clinical trial that incorporates information obtained during the study to change the design of the study in some fashion can be considered an adaptive clinical trial. With this broad definition of an adaptive clinical trial, all current clinical trials are adaptive because all trials must carefully monitor each patient's safety information; the sudden occurrence of serious adverse events will lead to termination of a trial. Many clinical trials also incorporate sequential analysis techniques (Whitehead 1992) or interim pilot studies (Wittes and Britain 1990) that monitor accruing efficacy data to determine a stopping rule for the trial. In this article, however, we focus only on trials that adapt the individual patient allocation rule based on information accrued during the trial. By being able to change the allocation rule based upon the observed data, adaptive designs have the potential to provide more ethical treatment to individual patients in the study. In studies where a failure of treatment is serious and the potential benefit of a successful treatment is great, minimizing exposure to the ineffective therapy is crucial and it is difficult to ignore the potential benefits of an adaptive design.

In spite of the potential advantages, adaptive designs have either not been used in clinical trials or have been designed using ad-hoc criteria. Although there are many theoretical articles in the literature on randomized adaptive trials, we are aware of only three real-life clinical trials. One of the trials is the University of Michigan extracorporeal membrane oxygenation (ECMO) trial (Bartlett et al. 1985; Cornell, Landenberger, and Bartlett 1986). The other two trials are Lilly-sponsored randomized multicenter trials with which we have been associated. In phase I cancer trials, an ad-hoc adaptive design has been widely used for many years (Von Hoff, Kuhn, and Clark 1984), even though the poor operating characteristics for this design have been well documented (O'Quigley et al. 1990). In this article, we discuss some of our own experiences in the design and implementation of both adaptive randomized trials and dose escalation trials. We hope that by illustrating our own experiences (good and bad), we can provide more insight to other applied statisticians than by summarizing the substantial statistical literature on adaptive designs.

Contents

FEATURED ARTICLE

**"Adaptive Designs in Clinical Trials: Lilly Experience"
FARIES, TAMURA, AND ANDERSEN 1**

| | | |
|------------------|------------------|----|
| Discussion | LACHIN | 6 |
| Discussion | SIMON | 6 |
| Discussion | TEOH | 9 |
| Discussion | AL-OSH AND DUBEY | 9 |
| Rejoinder | | 11 |

BIOPHARMACEUTICAL SECTION NEWS

| | | |
|--|--|----|
| Letter from the Editors | | 11 |
| Letter from the Past Chair | | 12 |
| Minutes of August 15, 1994 ASA Biopharmaceutical Section Executive Committee Meeting | | 12 |
| Topics for the Biopharmaceutical Roundtable Luncheon | | 14 |
| Request for Names of ASA Fellow Candidates | | 15 |
| Business Meeting Announcement | | 16 |
| Notes of March 25, 1995 ASA Biopharmaceutical Section Executive Committee Meeting | | 17 |
| Book Review: <i>Randomization and Monte Carlo Methods in Biology</i> , Reviewed by Boris Iglewicz | | 17 |
| Corporate Members | | 18 |
| Section Executive Committee and Section Representatives | | 19 |
| Let's Hear from You! | | 20 |

In Section 2, after a discussion of the ECMO trial, we discuss our own choice of design for a two-arm adaptive study, for a four-arm adaptive study, and for a phase I cancer dose escalation design. In Section 3, we discuss some of the logistical difficulties associated with the two-arm and four-arm adaptive trials. We conclude with some additional thoughts and opinions on the future of adaptive trials in Section 4.

2. Adaptive Clinical Trials: ECMO and Our Experience

2.1 ECMO

The first published clinical trial that adapted the individual patient allocation rule based on the observed data was a study in persistent pulmonary hypertension of newborn infants reported by Bartlett et al. (1985). Historical data suggested that the mortality rate using conventional medical therapy among such infants was at least 80%. However, using extracorporeal membrane oxygenation (ECMO), a mortality rate of at most 20% had been reported at several centers. A clinical trial was planned in order to provide a scientific comparison of ECMO and conventional therapy. Failure of treatment in this indication was death, and the potential existed for a dramatic improvement in survival using ECMO. Due to these ethical concerns, the researchers chose to run an adaptive study using the randomized play-the-winner design proposed by Wei and Durham (1978).

The randomization process in the randomized play-the-winner design can be described by an urn containing balls labeled with a treatment assignment. Patients are randomized by sampling with replacement from the urn and assigning the treatment associated with the label on the selected ball. The urn for the ECMO study contained one ball of each type at the start of the study. Thus, at the beginning of the study, patients had a 50 percent chance of receiving either treatment. When a patient randomized to a treatment arm survived, a ball with the same treatment label was added to the urn. When a patient did not survive, a ball with the other treatment label was added to the urn. A treatment difference in efficacy would result in a higher proportion of balls in the urn for the superior treatment and thus a higher probability that a patient would be randomized to the superior treatment. The stopping rule was to continue randomization until 10 balls of one type were added to the urn. This rule was selected in order to have at least a 95% probability of selecting the superior therapy assuming at least a 40% difference in survival rates. Control of Type I error, however, was not a factor in the selection of the stopping rule.

Using the randomized play-the-winner rule, the first patient was assigned to ECMO and lived. The second patient received conventional therapy and died. The next 10 infants were each assigned to ECMO and survived. Thus, the final results were one failure on conventional therapy and 11 successes on ECMO. These results generated much controversy in the literature (Ware 1989; Begg 1990) and several discussants were critical of the adaptive design because of the results of this study. The lack of consensus on

the issue resulted in a second trial in which ECMO again outperformed conventional therapy (Ware 1989).

In our opinion, the ECMO situation provided an appropriate setting for an adaptive trial. We also feel that adaptive designs in general were unfairly criticized based on the results of this trial. The presence of only one patient on conventional therapy contributed to the lack of agreement in various statistical analyses. However, the sample size in the ECMO trial was chosen to select the truly superior therapy as opposed to demonstrating statistical significance between therapies. When the opportunity arose, we designed an adaptive trial that would avoid some of the controversy with sample size.

2.2. A Two-Arm Adaptive Study in Depression

Researchers in psychiatry are interested in biological markers for major depressive disorder. There is some evidence in the literature that patients with reduced rapid eye movement latency (REML) respond well on antidepressant treatment and poorly on placebo compared with the general depression patient population (Coble et al. 1979; Rush et al. 1989). Our psychiatrists were interested in testing the efficacy of fluoxetine in patients with reduced REML. We felt that the potential existed for a large treatment difference, and we wanted to limit the number of patients treated with placebo. At the same time, a comparison with placebo was preferred scientifically in order to assess the treatment effect in this population of patients. A randomized play-the-winner design was selected.

Although we chose a similar randomization algorithm to that used in the ECMO trial, there were many differences in design between this study and the ECMO study. Patients were stratified into two groups based on REML measured at baseline (reduced REML, nonreduced REML) and were randomized to active treatment or placebo using two independent randomized play-the-winner urns. To avoid having a treatment arm with extremely few patients, we randomized the first six patients within each stratum in a nonadaptive balanced fashion. Response was defined as a reduction of 50% or more in a patient's total score on the 17-item Hamilton scale for depression (HAMD17) from baseline to endpoint. We used simulations to determine a target sample size of 50 randomized patients in the reduced REML stratum. This number of patients provided at least 90% power to detect a difference in response rates between 70% on active treatment and 20% on placebo using a one-sided .05 level test and assuming a dropout rate of at most 20%. In order to compute the power, we assumed 70% and 20% response rates for active treatment and control, a fixed delay time for responses, and generated simulated data under the actual design used. Other simulations, assuming equal response rates on both treatments, were run to confirm the .05 significance level of the test. The simulations also demonstrated that the number of patients assigned to placebo would likely be reduced, but the case of an extremely low sample size was unlikely. The difference in total sample size between this study and a traditional balanced study that would have required 48 reduced REML patients was small.

Unlike the ECMO study, in which responses were observed fairly quickly, a determination of response for

depression is typically not available until the end of a six- to eight-week treatment period. In order to update the urn in a more timely fashion, a surrogate response (see Tamura et al. 1994 for more details of the surrogate), rather than the response at endpoint, was used for updating the urn. Both the surrogate and the final response were based on a patient's change from baseline in the HAM-D17. The surrogate response did not affect the definition of final response used in the analysis of the data.

The primary analysis of the data was based upon the posterior probability that the true response rate on active treatment was greater than that on placebo. A posterior probability of greater than .95 was considered a statistically significant result. Independent uniform prior distributions for the true response rates on each treatment were used in the calculations. Secondary analyses included a comparison of mean changes in the HAM-D17 using noninformative prior distributions. This Bayesian approach, which is based on the likelihood and not influenced by the design of the study, was accompanied by a frequentist approach. Randomization tests, which incorporate the design of the study and make no distributional assumptions, were also used to test for treatment differences in response rates and HAM-D17 changes. For the randomization tests, patient responses and time to response were held fixed and the distribution of the test statistic under the null hypothesis was generated under the actual allocation rule used in this study.

The study was successful in demonstrating efficacy of fluoxetine in reduced REML patients. Inferences from both the Bayesian and frequentist analyses, despite the different assumptions, were in agreement. However, the allocation of patients to treatment was more balanced than we expected. This was due to many factors: the observed treatment difference was slightly smaller than expected, the delay to surrogate response was longer than expected, differences between the surrogate and final responses produced a more balanced urn early in the study, and finally, in small samples, the play-the-winner allocation rule results in highly variable treatment group sizes. In a follow-up simulation study, we found that given the observed rates of response and the actual design used, balance in the treatment groups was not extremely unlikely for a study with only 45 reduced REML patients. Further details of the design, analysis, and results of this study are given in Tamura et al. (1994).

2.3. A Four-Arm Adaptive Study in Depression

The results of our first trial encouraged us to start a second adaptive trial in patients with major depressive disorder. As the analysis of data from this study is not yet complete, only a brief discussion of the design is given here. The design was similar to the two-arm study in that depressed patients were stratified into two groups based on REML, sample sizes were selected using simulations, and both a Bayesian and frequentist analyses were planned. The second trial, however, had four treatment groups. Most of our research efforts in planning this study dealt with how to design an adaptive study with more than two treatment arms.

Wei (1979) proposed extending the two-arm randomized play-the-winner design by adding balls of the

successful treatment to the urn when a success is observed and balls of each of the other treatment arms to the urn when a failure is observed. For instance, in a four-arm trial with $\alpha = 3$, $\beta = 1$, three balls of a successful treatment are added to the urn when a success is observed, and one ball of each of the other treatment arms is added if a failure is observed. Intuitively, a success or failure on one treatment should not influence the probability of selecting a second treatment over a third treatment. However, when a failure on one treatment is observed, this rule would change the proportions of balls between the other treatments. For instance, suppose that the contents of the urn at a given point in time in a four-arm trial were (2,4,1,2) for treatments (A,B,C,D). Under the $\alpha = 3$, $\beta = 1$ rule, if a Treatment C patient failed, then the contents of the urn would be (3,5,1,3). Using the previous contents, a patient was twice as likely to receive Treatment B compared with Treatment A. However, even though a failure on Treatment C provided no information regarding the efficacy of Treatment B versus Treatment A, a patient's chance of receiving Treatment B relative to Treatment A is now reduced. For these reasons, we developed a different randomization scheme (Andersen, Faries, and Tamura 1994).

In our scheme, the same number of balls (e.g., one) is added to the urn for a failure as for a success. If a success is observed, one ball of the treatment is added to the urn. If a failure is observed, then one ball is divided among the other treatments and added to the urn. In order to preserve the current relationships in the urn, the ball for a failure is divided among the other treatments in proportion to the number of balls currently in the urn for those treatments. For instance, if the current urn is (2,4,1,2) for treatments (A,B,C,D) and a Treatment C patient fails, then the urn will be updated to (2.25,4.5,1,2.25). Note that a patient still has twice the chance of receiving Treatment B as Treatment A.

Although we prefer our design to that of Wei (1979), one drawback to our approach is that the urn status is not invariant to the order of responses. Given the same set of responses, the proportion of balls in the urn for each treatment could be different depending upon the order in which the responses are observed. The early patients tend to have a slightly greater impact on the contents of the urn than later patients. One possibility we are considering to correct this problem is to base the distribution of balls in the urn on summary statistics, such as the posterior probabilities that each treatment is superior to the other treatments. In this way, the proportion of balls in the urn would be invariant to order of the responses.

2.4 A Dose Escalation Design for Phase I Cancer Trials

The primary objective of a phase I trial in cancer is to determine the maximum tolerated dose (MTD) of the compound. Because cancer compounds are typically quite toxic, the first patient must be treated at a low dose level (often well below the level expected for efficacy) and escalation should proceed cautiously. However, participants of these trials are cancer patients who have failed other therapies (not normal volunteers) and thus one must administer dose levels that have the potential for efficacy.

These trials present an ethical dilemma. The design must start at a conservative dose level, escalate slowly enough to avoid toxic dose levels, but escalate quickly enough to minimize the number of patients treated at dose levels with little potential for efficacy.

For many years, an ad-hoc dose escalation design (Von Hoff, et al. 1984) has been used for phase I cancer trials. In this traditional design, the first three patients are treated at a low dose level. If no unacceptable toxicity is observed, the next three patients are treated at the next higher dose level. If one of the three patients has unacceptable toxicity, then an additional three patients are treated at the same dose level before the next higher dose level can be administered. If at least 2 patients at a dose level suffer unacceptable toxicity, the study ends. This design has been criticized in the literature (O'Quigley et al. 1990) for its inefficiency in estimating the MTD and for treating a large number of patients at doses with little potential for efficacy.

O'Quigley et al. (1990) proposed a Bayesian strategy, the continual reassessment method (CRM), in which an updated estimate of the MTD is calculated after each patient's response is observed. The estimate is then used to guide the dose escalation process. In the CRM, a prior distribution for the MTD and a dose toxicity model are selected prior to the study. The mean of the prior distribution is used as the dose level for the first patient. After each patient's response is observed, an updated estimate of the MTD is calculated from the posterior distribution of the MTD. Patients entering the study are treated at the current estimated MTD. O'Quigley demonstrated by simulations that by using updated estimates of the MTD to guide the escalation, more patients in the trial will receive dose levels close to the MTD and fewer patients will receive levels with little potential for efficacy. He also demonstrated through simulations that the CRM provides a more accurate and less variable estimate than the traditional design.

We are currently using a modification of the CRM in our phase I cancer trials. These modifications are designed to lessen the chance of administering toxic dose levels, while retaining much of the improvement in efficiency and reduction in use of suboptimal dose levels provided by the CRM. The first modification is to use a conservative, low starting dose as in the traditional design. Our oncologists considered the potential for toxicity in the first patient to be too great to start with anything other than a conservative dose. In addition, we select a set of predefined dose levels as in the traditional method. As in the CRM, we calculate an updated estimate of the MTD after each patient's response and use the estimate to guide the dose escalation. However, our modifications limit the maximum dose increase to one dose level. In addition, we require three patients to be treated at any dose level where moderate toxicity is observed and we require three patients to be treated at the initial dose level in order to avoid exposure to an extremely toxic dose level. Except at the initial dose level, it is still possible to administer the next higher dose level after only a single-patient exposure, if the estimated MTD suggests an increase and no moderate toxicity was observed. Our simulation studies indicate that the modified CRM will continue to produce a more efficient estimate of the MTD and also provide a more ethical selection of dose levels for the patients involved in the study. More details of the basic design were provided by Faries (1994).

3. Logistics in Conducting Adaptive Trials

In this section we discuss some of the logistical issues involved in our two-arm and four-arm randomized adaptive trials. The logistical issues involved in the phase I cancer trials are fewer because these trials are typically unblinded and conducted at a single site.

The greatest difference between adaptive trials and traditional trials is the amount of communication required between the sponsor and the investigators. In a traditional multicenter, double-blind trial, the randomization sequence for each investigator is a priori generated and kits of clinical trial material are then constructed according to these randomization sequences. Each kit is unique to a single patient and contains all the medication for the patient for the duration of the trial. The kits are constructed according to the randomization sequence and the investigator is then instructed to assign kits sequentially as patients are randomized into the trial. Patients can therefore be randomized and assigned medication throughout the trial with little if any communication between the investigator and the sponsor.

In an adaptive trial, randomization is dependent on accruing data; therefore, many randomization sequences need to be generated throughout the trial. The investigator can no longer assign kits sequentially and must contact the sponsor for the proper kit assignment for each patient. The efficacy data must be communicated to the sponsor in a timely fashion; otherwise the benefits of adaptive randomization are lost. These requirements dictate that the information flow between sponsor and investigator be accurate and timely. In our two randomized adaptive studies, we requested that each investigator call the sponsor after every patient visit with the results of the HAMD17 score for that patient. A clinical research associate blinded to treatment recorded this information and determined whether a patient was a responder or a nonresponder. If the patient could be classified as a responder or nonresponder, the information was then passed to an unblinded clinical research associate. The unblinded clinical research associate updated a computer program that produced a new randomization sequence. The unblinded clinical research associate then passed a new sequence of kits assignments based on the new randomization sequence to the first clinical research associate. Thus, an updated randomization sequence of kits was available at all times and employees in contact with the investigational sites remained blinded.

This system worked reasonably well but did suffer two problems. First, it required two research associates to be on-call at all times because many of the investigators randomized patients on weekends and after work hours. This was difficult since it required the associates to keep randomization logs with them at home. Changes in job assignments for associates of this trial and the training of backup personnel created difficulties. A second problem we detected was that some investigators did not always call in response data after a patient was randomized. Thus, our associates had to prompt the investigators for the missing data and some patients' response data was never received in time to update the urn. This was more of a problem during the later part of the four-arm trial.

After the four-arm trial was initiated, we began studying

ways in which the logistical difficulties could be handled. We have since developed an automated touch-tone phone system that is being tested in an ongoing pilot study. Similar types of phone systems have been used in other clinical trial settings (Vaughn et al. 1994). In our system, the investigator calls into the system after each patient visit and inputs the patient's HAMD17 response. We eliminated the patient kits of clinical trial material, and provided only cartons of blinded clinical trial material to the investigators. The clinical trial material from a carton can be dispensed to any patient. The investigator calls the system after every patient visit, and after entering the required patient information, is given the proper bottle numbers of clinical trial material to dispense. The investigator is required to call into the system in order to obtain therapy for each patient at every visit, thus rapid acquisition of the relevant response data is guaranteed. Computerization of the randomization rule eliminates the potential for human error. We were initially worried that investigators would object to using the phone system after every patient visit; however, this has not been a problem in our pilot study. Our investigators of the pilot study were pleased to eliminate the storage of large numbers of unused kits and were also pleased with the 24-hour availability of the system. The only difficulty that one of our investigators experienced was temporary discontinuations in local telephone service within their offices. In these cases, the site personnel used another telephone in a different building. We are currently investigating whether a personal digital assistant computer in conjunction with the phone system could be used as a backup in case of disruptions in telephone service.

4. Discussion

The use of placebo-controlled trials to indicate whether an effective therapy exists is coming under increasing scrutiny for ethical reasons (Rothman and Michels 1994). Despite this criticism, the scientific need for a comparison with placebo often exists. An adaptive design that includes placebo therapy, active therapy, and test therapy may offer a compromise between a traditional, balanced placebo-controlled trial and an active comparator-controlled trial. The scientific comparison with placebo is retained, yet exposure to placebo would be lessened if the test and/or active comparator proves to be efficacious. In some situations an unbalanced, yet fixed, randomization might produce the same results as a design that is adaptive on response. However, in most situations we do not know which treatment or which dose will prove to be superior in the patient population for a particular study. Thus, choosing an unbalanced randomization proportion in advance is difficult.

The design of randomized clinical trials has been relatively stable for the past thirty years. However, in the past decade, the advances in computer technology and data access have been dramatic, which makes the logistics of adaptive trials much more feasible. It seems obvious that the technological advances will continue for the near future. Given these improvements, clinical trials will likely become more adaptive in all aspects. The allocation rule for patients in future trials might be simultaneously adaptive on many different types of accrued data (baseline data, efficacy, and safety data). We statisticians need to actively influence the designs of such trials.

In this article, we have tried to give some of our

experiences in the planning and conduct of adaptive trials. We have been involved in adaptive trials for only the past three years, which is significantly less than our experience with traditional balanced clinical trials. Much more needs to be learned; however, we feel that the only way to gain experience is to conduct such trials and learn from our successes and failures. We encourage our clinical colleagues in the biopharmaceutical industry to do the same.

References

- Andersen, J., Faries, D., Tamura, R. (1994), "A Randomized Play-the-Winner Design for Multi-Arm Clinical Trials," *Communications in Statistics, Part A—Theory and Methods*, 23, 309-323.
- Bartlett, R. H., Roloff, D. W., Cornell, R. G., Andrews, A. F., Dillon, P. W., and Zwischenberger, J. B. (1985), "Extracorporeal Circulation in Neonatal Respiratory Failure: A Prospective Randomized Study," *Pediatrics*, 76, 479-487.
- Begg, C. B. (1990), "On Inferences from Wei's Biased Coin Design for Clinical Trials," *Biometrika*, 77, 467-484.
- Coble P. A., Kupfer, D. J., Spiker, D. G., Neil, J. F., and McPartland, R. J. (1979), "EEG Sleep in Primary Depression," *Journal of Affective Disorders*, 1, 131-138.
- Cornell, R. G., Landenberger, B. D., and Bartlett, R. G. (1986), "Randomized Play-the-Winner Clinical Trials," *Communications in Statistics, Part A—Theory and Methods*, 15, 159-178.
- Faries, D. (1994), "Practical Modifications of the Continual Reassessment Method for Phase I Cancer Clinical Trials," *Journal of Biopharmaceutical Statistics*, 4, 147-164.
- O'Quigley, J., Pepe, M., and Fisher, L. (1990), "Continual Reassessment Method: A Practical Design for Phase I Clinical Trials in Cancer," *Biometrics*, 46, 33-48.
- Rothman, K. J., and Michels, K. B. (1994), "The Continuing Unethical Use of Placebo Controls," *New England Journal of Medicine*, 31, 394-398.
- Rush, A. J., Giles, D. E., Jarrett, R. B., Feldman-Koffler, F., Debus, J. R., Weissenburger J., Orsulak, P. J., and Roffwar, H. P. (1989), "Reduced REM Latency Predicts Response to Tricyclic Medication in Depressed Outpatients," *Biological Psychiatry*, 26, 61-72.
- Tamura, R. N., Faries, D. E., Andersen, J. S., and Heiligenstein, J. H. (1994), "A Case Study of an Adaptive Clinical Trial in the Treatment of Out-Patients with Depressive Disorder," *Journal of the American Statistical Association*, 89, 768-776.
- Vaughn, M., Holland, K., Wagner, W., Nufer, L., Patterson, T., Ray, C., and Davies, H. W. (1994), "The Use of Touch Tone Technology for In-House Monitoring of a Large Treatment IND," presented at the Drug Information Association Meetings, Washington, DC.
- Von Hoff, D. D., Kuhn, J. and Clark, G. M. (1984), "Design and Conduct of Phase I Trials," in *Cancer Clinical Trials*, eds. M. E. Buyse, M. J. Staquet, and R. J. Sylvester, New York: Oxford University Press, pp. 210-220.
- Ware, J. H. (1989), "Investigating Therapies of Potentially Great Benefit: ECMO," *Statistical Science*, 4, 298-340.
- Wei, L. J. (1979), "The Generalized Polya's Urn Design for Sequential Medical Trials," *The Annals of Statistics*, 7, 291-296.
- Wei, L. J., and Durham, S. (1978), "The Randomized Play-the-Winner Rule in Medical Trials," *Journal of the American Statistical Association*, 73, 840-843.

Whitehead, J. (1992), *The Design and Analysis of Sequential Clinical Trials*, New York: Ellis Horwood.

Wittes, J. and Brittain, E. (1990), "The Role of Internal Pilot Studies in Increasing the Efficiency of Clinical Trials," *Statistics in Medicine*, 9, 65-72.

Discussion

John Lachin

Professor of Statistics

Director, The Biostatistics Center

The George Washington University

Although there has been a plethora of statistical, theoretical, and methodological research on the properties of response-adaptive designs, these designs are still only rarely used. The investigators at Eli Lilly and Co. are among the few who are actively placing these designs to the test. I hope that such pioneering efforts will lay the groundwork for translating the rich theoretical literature into clinical research practice.

The ultimate purpose of response-adaptive designs is simple and appealing. Rather than subject all patients in a study to an equal (or fixed) probability of receiving either treatment, the objective is to provide patients who enter later in the trial with a higher probability of being assigned to receive the apparently better treatment. In this way later patients benefit from the interim accrued information provided by earlier patients. Such designs are considered more ethical than purely randomized designs.

Such designs may also have other tangible benefits. They are likely to be more appealing to prospective volunteers and could facilitate recruitment. They also may promote patient adherence and completeness of follow-up, providing a higher quality study.

Response-adaptive designs have been debated in the medical literature for years. The points raised by Weinstein (1974) and the counterpoints of Byar et al. (1976) are as relevant today as they were 20 years ago. Although theoretical statisticians have explored the properties of these designs and developed proper statistical tests under such designs, these designs still have their potential flaws.

The principal problem, from my perspective, is the requirement that the stream of patients entering such a trial constitute a "random" sample from a single population—that is, one in which the expected characteristics do not vary over time. For such trials to provide an unbiased assessment of treatment effects it is necessary to assume that the patients who enter later in the study are drawn from the same population as those who entered early. Therefore, the challenge to those who conduct such trials is to implement recruitment strategies such that this assumption is satisfied.

Such designs are also operationally challenging. Implementation in a stratified, double-masked, multicenter trial with "delayed" outcome assessments requires greater administrative resources than does the simple randomized design.

However, successfully addressing all of these issues is for naught if the results of such studies are not accepted by the scientific community, irrespective of acceptance by the statistical community. The objective of any clinical trial is to

Biopharmaceutical Report, Summer 1995

obtain accurate and compelling information on the effects of a treatment. If the study and its results are not perceived in this light, then the study will not have the intended impact.

Rosenberger and Lachin (1993) recently reviewed the "state of the art" methodologically of response-adaptive designs. Clearly, such designs are not yet widely accepted, in part due to limited exposure and in part due to the unfavorable reaction to earlier attempts. However, there are settings in which we believe it is now appropriate to consider such designs. We concluded that these designs could now be successfully implemented in trials where there is a single hypothesis with timely ascertainment of the outcome, but where the outcome is not life-threatening. From this perspective, the Lilly trials of fluoxetine in the treatment of patients with reduced REML are ideal.

Like any "new" idea, the widespread acceptance of these designs will take time and experience. It is only through the careful and deliberate use of these designs that statisticians, physicians, editors, and regulatory scientists will become comfortable with their use and will be able to identify those settings in which their use is most appropriate. The investigators at Lilly are to be commended for contributing to this experience.

References

- Byar, D.P., Simon, R.M., Friedewald, W.T., Schlesselman, J.J., DeMets, D.L., Ellenberg, J.H., Gail, M.H., and Ware, J.H. (1976), "Randomized Clinical Trials: Perspectives on Some Recent Ideas," *New England Journal of Medicine*, 295, 74-80.
- Rosenberger, W.F., and Lachin, J.M. (1993), "The Use of Response-Adaptive Clinical Trials," *Controlled Clinical Trials*, 14, 471-484.
- Weinstein, M.C. (1974), "Allocation of Subjects in Medical Experiments," *New England Journal of Medicine*, 291, 1278-1285.

Discussion

Richard Simon

Chief, Biometric Research Branch

National Cancer Institute

The authors are to be congratulated for their attempts to adapt and apply new statistical methodology to real-world clinical trials. The large literature on adaptive treatment assignment methods has found few applications to date for reasons that have been discussed previously (e.g., Simon 1977, 1989). Most previous models have been based on either a decision theoretical framework or on ranking and selection theory. Neither approach has been well suited for phase III clinical trials. For example, Spiegelhalter and Freedman (1988) conclude: "... in general the consequences of a phase II trial are difficult to formalize, and that a 'medical trial is not, in any clear-cut fashion, a decision procedure.' Essentially, although a private decision may exist concerning termination, the consumers require conclusions in order to make their own decision." Ranking and selection theory has found applications in phase II clinical trials (Simon, Thall, and Ellenberg 1994),

but is generally not appropriate for phase III trials, where inference about the null hypothesis is generally of interest. Faries et al. have tried to modify and apply adaptive methods in phase III trials in a manner that provides for inference of the type usually required.

The authors have defined adaptive trials as those in which the treatment allocation rule is modified based on information accrued during the trial. The adaptive stratification designs widely used in cancer clinical trials (e.g., Pocock and Simon 1975) satisfy this definition. These procedures use a restricted randomization procedure to balance the treatment groups marginally with regard to many prognostic factors. The probability of allocation for a given patient depends on the covariate values for that patient and on the covariate values and treatment assignments of previously registered patients. The primary motivation usually given for the development of what I would call "outcome adaptive designs" is an appeal to ethics. Faries et al. assert that "...adaptive designs have the potential to provide more ethical treatment to individual patients in the study. In studies where failure of treatment is serious . . . minimizing exposure to the ineffective therapy is crucial and it is difficult to ignore the potential benefits of an adaptive design." This argument requires careful consideration, however. "Treatment" is not subject to ethical scrutiny; behavior of an individual is. In order to address the ethics of randomized and outcome adaptive trials, we must consider the actions of individuals involved in such trials and the information on which those actions are based. Also, behavior is usually viewed as either being consistent or inconsistent with particular ethical principles, not as being rated on a continuum as more or less ethical as implied above.

The physician registering a patient in a clinical trial must have the best interest of that particular patient as his/her primary consideration. The "respect for persons" ethics upon which modern medical research is based means that the autonomy of that particular patient must be the primary concern of that physician. If that physician believes that substantial interim data supports the superiority of treatment A, then it is difficult to see how the physician can fulfill his or her obligation to respect the autonomy of that patient without sharing that information with the patient. Merely weighting the randomization in favor of A does not resolve the ethical problem of the physician.

Keeping physicians who enter patients blinded from interim outcome data avoids compromising their ethical position, uninfluenced by whether or not an adaptive design is used.

The ethical position of the members of the data monitoring committee (DMC) is more complex. It is widely viewed that the primary purposes of a data monitoring committee are "ensurance both of participant safety and of study integrity (and of course adjudicating the inevitable occasional conflicts between these goals)" (Friedman 1993). To the extent that the credibility or influence on medical practice of a trial is reduced by the use of outcome adaptive methods, their objective of ensuring study integrity may be compromised. If the use of outcome-adaptive methods reduces the number of patients assigned to what turns out to be the inferior treatment but increases the total number of patients assigned to what turns out to be the inferior treatment by increasing the total number of patients treated and the time till an answer is obtained, announced, and the new treatment made available to patients not on the trial, then utilitarian ethics, the greatest good for the

greatest number, will not have been enhanced. In general, however, there will be other more consequential delays and this will not be a serious consideration.

From the point of view of the responsibility of the DMC members to protect the volunteers participating in the trial, if it is ethical to assign treatment with an outcome-weighted randomization, then it must be ethical to assign treatment with equal probability. Consequently, it is not clear that outcome-adaptive treatment assignment has a positive impact on the ethical position of members of the data monitoring committee.

The major concern about the use of outcome-adaptive methods has always been that these methods compromised the strength of inference provided by randomized clinical trials. Proposed analyses of adaptive trials are generally based on models of independent and identically distributed errors given treatment and possibly known covariates. Simulations or analytical derivations usually used to show that correct inference is preserved using adaptive methods beg the question because the validity of the models cannot be ensured. Obviously if one's model can be assumed correct, then there is no need for randomization. One of the main purposes of randomization is to guard against conscious or subconscious selection bias which would invalidate most models. But one cannot assume the absence of selection bias. Rubin (1978) showed that ignorability of the treatment assignment rule, ensured by randomization, was equally important for Bayesian analysis of causal effects. Rosenbaum and Rubin (1984) also showed that "Bayesians can face heightened sensitivity to prior assumptions when a data-dependent [stopping] rule is used, even if the rule is ignorable. . ."

Faries et al. provide two types of analyses for comparing treatment groups in their two-arm trial of patients with depression. One is a Bayesian analysis based on the assumption that the observations are independently and identically distributed. This analysis is subject to the limitations described above. Simon et al. (1977) showed an example of how time trends with regard to an unknown covariate can inflate the type 1 error to .18 instead of the .05 indicated by the model. Generally, known prognostic factors account for only a small proportion of the variability in outcome, and so imbalances in unknown determinants resulting from selection bias may distort results. We can, however, adjust for the propensity score, the probability of assigning the new treatment to a patient. Rosenbaum and Rubin (1983) proposed using this adjustment in observational studies, and it can also be useful for the analysis of outcome-adaptive clinical trials.

A second potential problem with the proposed Bayesian analysis is the use of independent flat priors for the response probabilities for the two treatments. This places substantial prior weight on large treatment differences. This assumption implies that the prior probability that the difference in response probabilities exceeds x in absolute magnitude is $1-2x$. For a difference of 20 percentage points (common for cancer trials), the prior probability of a larger difference is .60. For the trial in question, such an optimistic prior may be appropriate, but in many other cases it will not.

The second form of analysis described by the authors is attractive and similar to the method I proposed for use with adaptive stratification designs (Simon 1979). "It is possible, though cumbersome, to perform the appropriate randomization test generated by a nondeterministic adaptive stratification design. One assumes that the patient responses,

covariate values, and sequence of patient arrivals are all fixed. One then simulates on a computer the assignment of treatments to patients using the adaptive stratification design actually employed and the treatment assignment probabilities actually employed. Replication of the simulation generates the approximate null distribution of the test statistic adopted, and the significance level. One need not make the questionable assumption that the sequence of patient arrivals is random. . ."

Faries et al have modified this approach for outcome adaptive designs with time delays in observing response. Because they have based treatment assignment on a shorter-term "surrogate" endpoint, however, the method is subject to certain potential biases. Suppose, for example, that the new treatment influences the "surrogate" but not the true outcome, and that there is an unknown time trend toward the registration of better-prognosis patients who will be assigned the new treatment. The overall response rate for the new treatment on the true endpoint will tend to be greater than for the control treatment because of the unknown prognostic imbalance. This problem can be avoided if the intermediate endpoint used is actually a good surrogate for the true endpoint. For cancer treatments, however, it is common for new treatments to have substantial effects on intermediate endpoints such as tumor shrinkage, but little or no effect on survival.

One last comment on the Lilly experience with outcome-adaptive designs for phase III trials concerns the absence of mention of sequential analysis. Because the objective is to reduce the number of patients given a treatment that ultimately is shown to be inferior, it would seem appropriate to compare the results for outcome adaptive designs to those for standard sequential designs using equal randomization rather than to fixed sample size designs.

Many clinical trials provide little opportunity to use outcome-adaptive designs. In some cases there are long time delays in observing response and no good intermediate endpoints. There may be multiple endpoints of importance, covariates to deal with, and subset effects of concern. There are additional concerns in using this approach in an open label clinical trial. In such cases, investigators may be aware of whether the active treatment is doing better than placebo. Knowledge that outcome-adaptive assignment is being used could induce time trends in the types of patients registered for the trial. The description by Faries et al. provides very useful examples, however, of serious and innovative attempts to use this body of methodology in certain types of phase III clinical trials.

The Lilly group is also to be congratulated for attempting to adapt the continual reassessment method for use in phase I cancer clinical trials (Faries 1994). Korn et al. (1994) expressed concern about the safety and time required to complete phase I trials conducted using the originally proposed continual reassessment method. As Faries et al point out, because the prior distribution of the dose-toxicity relationship may be very diffuse, starting out with the dose for which the expected incidence of dose-limiting toxicity is the target level (e.g., 1/3), may be very risky. In such studies the likelihood of serious toxicity at an excessive dose is much greater than the likelihood of patient benefit. The study of continual reassessment type methods, both Bayesian and frequentist, goes back at least to Zacks and Eichhorn (1975, 1981) and Anbar (1984). More recently, the design of phase I

cancer trials has been considered by Storer (1989), O'Quigley et al, (1990), Gatsonis and Greenhouse (1992), and others. Although the traditional methods leave room for improvement, it should be recognized that the phase I trial is not an end in itself. Patients in phase I trials are often more debilitated than patients on later phase clinical trials and stay on study for only a few (e.g., 1-3) courses of treatment. Consequently, many phase I trials can at best furnish a reasonable dose for initial use in subsequent trials. Precise estimation of the dose at which some specified proportion of phase I patients will experience dose limiting toxicity is not the real objective. Statistical designs for safely expediting the conduct of phase I trials and for limiting the number of patients treated at homeopathic doses are, however, important and warrant investigation and testing.

References

- Anbar, D. (1984), "Stochastic Approximation Methods and Their Use in Bioassay and Phase I Clinical Trials," *Communications in Statistics A*, 13, 2451-2467.
- Faries, D. (1994), "Practical Modifications of the Continual Reassessment Method for Phase I Cancer Clinical Trials," *Journal of Biopharmaceutical Statistics*, 4, 147-164.
- Friedman, L. (1993), The NHLBI Model: A 25 Year History in *Practical Issues in Data Monitoring of Clinical Trials*, eds. S.S. Ellenberg, N. Geller, R. Simon, and S. Yusuf, *Statistics in Medicine*, 12, 425-431.
- Gatsonis, C., and Greenhouse, J.B. (1992), "Bayesian Methods for Phase I Clinical Trials," *Statistics in Medicine*, 11, 1377-1389.
- Korn, E.L., Midthune, D., Chen, T.T., Rubinstein, L.V., Christian, M.C., and Simon, R.M. (1994), "A Comparison of Two Phase I Trial Designs," *Statistics in Medicine*, 13, 1799-1806.
- O'Quigley, J., Pepe, M., and Fisher, L. (1990), "Continual Reassessment Method: A Practical Design for Phase I Clinical Trials in Cancer," *Biometrics* 46, 33-48.
- Pocock, S.J., and Simon, R. (1975), "Sequential Treatment Assignment With Balancing for Prognostic Factors in the Controlled Clinical Trial," *Biometrics*, 31, 103-115.
- Rosenbaum, P.R., and Rubin, D.B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.
- (1984), "Sensitivity of Bayes Inference With Data-Dependent Stopping Rules," *The American Statistician*, 38, 106-109.
- Rubin, D.B. (1978), "Bayesian Inference for Causal Effects: The Role of Randomization," *The Annals of Statistics*, 6, 34-58.
- Simon, R. (1977), "Adaptive Treatment Assignment Methods and Clinical Trials," *Biometrics*, 33, 743-749.
- (1979), "Restricted Randomization Designs in Clinical Trials," *Biometrics*, 35, 503-512.
- (1989), "Adaptive Allocation in Clinical Trials," 1989 *Proceedings of the Biopharmaceutical Section*, Alexandria, VA: American Statistical Association, pp. 23-25.
- Simon, R., Hoel, D.G., and Weiss, G.H. (1977), "The Use of Covariate Information in the Sequential Analysis of Dichotomous Response Experiments," *Communications in Statistics—Theory and Methods*, 8, 777-788.
- Simon, R., Thall, P., and Ellenberg, S.S. (1994), "New Designs for the Selection of Treatments to be Tested in Randomized Clinical Trials," *Statistics in Medicine*, 13, 417-429.
- Spiegelhalter, D.J., and Freedman, L.S. (1988), "Bayesian

- Approaches to Clinical Trials," in *Bayesian Statistics 3*, eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M. Smith, Amsterdam: North-Holland.
- Storer, B. (1989), "Design and Analysis of Phase I Clinical Trials," *Biometrics*, 45, 925-937.
- Tamura, R.N., Faries, D.E., Andersen, J.S., and Heiligenstein, J.H. (1994), "A Case Study of an Adaptive Clinical Trial in the Treatment of Out-Patients With Depressive Disorder," *Journal of the American Statistical Association*, 89, 768-776.
- Zachs, S., and Eichhorn, B.H. (1975), "Sequential Search of Optional Dosages: The Linear Regression Case," *Statistical Design in Linear Models*, 609-628.
- (1981), "Bayes Sequential Search of an Optimal Dosage: Linear Regression With Both Parameters Unknown," *Communications in Statistics*, 10, 931-953.

Discussion

Nick Teoh

Schering-Plough Research Institute

It is very encouraging to see that complex adaptive designs are still being implemented in clinical trials since the days of the ECMO studies. The authors deserve a lot of credit for carrying out simulations to better understand the potential pitfalls of using such designs and proactively modifying the design to avoid these.

In terms of the trial outcome, it is interesting to note that approximately equal numbers of patients were eventually assigned to both treatment arms. The authors mentioned several factors, including the delay in treatment response which prevented the urn from being updated through the adaptive mechanism on a timely basis. It may be better if the design allows for batches of patients to be treated and the urn be updated only after the outcomes of each batch of patients are known. It would seem that updating of the urn on a patient-by-patient basis may not be very efficient, as the increment in information provided by a single patient's outcome is insufficient to warrant immediate modification of the treatment assignment probabilities. The trial was also quite fortunate to produce results that yield similar conclusions whether randomization or Bayesian methods were used to analyze the data. Again, this is largely due to the modes difference in treatment effects and the near balance of treatment samples. It will be very instructive to see how one resolves the discrepancy in study results when divergent conclusions arise from the use of different statistical methods, since I do not expect the controversy between Bayesians and frequentists to go away soon.

The incremental benefit that derives from the use of an adaptive design such as the CRM in the context of a Phase I study may not, however, be as easily justified. This type of trial usually employs a small sample size of 15-30 subjects with dose-limiting toxicity (usually a binary outcome) as the primary endpoint. In addition, ethical concerns mandate starting the trial at a very low dose level with a prespecified number of patients, three being the most common number in typical cases. Escalation of the dose also is usually restricted to a single dose-step at a time, up to a maximum allowable dose. Thus, given these various constraints, it is quite unlikely for

one to gain a substantial increase in precision in estimating the MTD using adaptive designs. In this regard, one should be cautious in interpreting comparative simulation results since all such simulations are based on particular models. It may be quite tenuous and unwieldy to address the problem of model misspecification using simulations of small-sample experiments.

Aside from the above statistical considerations, I share the authors' concern that logistical difficulties may be by far one of the most daunting aspect of implementing adaptive designs. A carefully thought-out operational strategy involving back-up personnel and/or interactive phone systems such as the one described by the authors are crucial to the success of such a project. It will be prudent for one to anticipate potential pitfalls in such trials, even if they may seem hypothetical in the beginning.

Discussion

M. Al-Osh and S. Dubey

Division of Biometrics, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Rockville, MD

The attractive feature of the response-adaptive randomization method is its potential to provide the better treatment to the majority of patients in the study. Faries et al. discussed Lilly experience in implementing this design in two clinical trials. Besides the Lilly trials¹, the extracorporeal membrane oxygenation (ECMO) study, reported by Bartlett et al. (1985), was the first trial that used this design. The design and analysis of the ECMO trial has been the subject of controversy among researchers. Consequently, many discussants argued against the use of the response-adaptive design in actual clinical trials.

Among the problems limiting the application of the response adaptive-design, which Faries, Tamura, and Andersen addressed are: (1) having too few patients in one arm of the trial; (2) delayed response; and (3) the appropriate statistical analysis for this method. Lilly's approaches to these limitations (see Tamura, et al. 1994), when implementing this design in their depressive disorder trial, were: (1) randomize the first six patients within each stratum in a nonadaptive fashion; (2) use surrogate response instead of the final response for randomization purposes; and (3) use a Bayesian approach as the primary analysis for comparing alternative treatments, and a frequentist one as a secondary analysis.

Lilly's work in this area is a welcome addition to the increasing literature on the subject. However, their approach fails to provide fully satisfactory solutions to the problems associated with the application of the response-adaptive method.

To circumvent the problem of having few patients in one arm of the trial, Tamura et al. (1994) assigned the first six patients within each stratum of their trial by using a randomized block design. Then they used independent randomized play-the-winner rules starting with the seventh patient within each stratum. Because for Lilly's trial an eight-

¹The views expressed herein are those of the authors and not necessarily of the U.S. Food and Drug Administration.

week waiting period is needed for the final measurements, the authors used a surrogate response based on the results of two consecutive visits after at least three weeks of therapy. Implementing these modifications in their depressive disorder trial resulted in a nearly balanced actual treatment allocation, as they stated. Furthermore, Lilly's analysis of this trial showed that the differences between the Bayesian and randomization inferences were small.

Clearly, it is undesirable for this design to have balanced treatments allocations because this would rule out the attraction of this design in practice. The treatment allocation in Lilly's depressive disorder trial, as presented in Tamura et al., (1994, table 3), is summarized as:

| | | strata | |
|-----------|---|--------|----|
| | | 0 | 1 |
| Treatment | 0 | 21 | 22 |
| | 1 | 23 | 23 |

Farewell et al. (1993) showed that in the adaptive-allocation design, in contrast to the fixed-sample-size design, the marginal totals have an impact on the difference between treatments. Consequently, it may not be surprising with this almost-balanced design to have only small differences in inference between the Bayesian and the frequentist approaches. For the response-adaptive method, the interest would be in comparing the two approaches of inference under what is expected from this design—that is, few patients in one arm of the trial.

The goal of the response-adaptive design would be best served when the difference in the effectiveness between the two treatments is large. However, one would expect that the larger this difference is, the higher the proportion of successes of the superior treatment would be, and consequently; the farther the patients' allocation from the balanced design. But the attraction of assigning few patients to the inferior treatment in this case is challenged by: (1) loss of statistical power in testing for treatment differences; and (2) having zero or very few patients in one arm of the trial may make treatment comparisons questionable, as Andersen et al. (1994) remarked. The results of Lilly's simulation experiment (Andersen et al. 1994), which compared different urn configuration of the response-adaptive design, showed that the most balanced design across treatments had the greatest power to detect treatment differences and vice versa.

One way to avoid both extremes in allocating patients to treatments in this design is by updating the urn through the addition of different numbers of balls for the two treatments, depending on the results of the previous trial, instead of adding one ball to one treatment.

Lilly's use of a surrogate, instead of the actual, response for randomization purpose seems to be the appropriate course or action for delayed-response trials. However, the utility of using this surrogate response hinges on its ability as a predictor for the response at the endpoint. If the correlation between the surrogate and final responses is not large, then basing the randomization on the surrogate response may lead to treatment allocations not intended by the method. For many trials, it may be very difficult to predict the patient's response on the basis of the

information obtained within a short time of the patient's admission to the trial.

The statistical inference for adaptive designs has been one of the major factors for their infrequent use in practice. For the ECMO trial, which used this design, the analysis was controversial. This trial resulted in 11 patients allocated to the new treatment, all successes, and one patient allocated to the control treatment, a failure. Begg (1990) reported the following serious discrepancies in the P values for comparing the two treatments in this trial: (1) an analysis that ignored the design and presumed complete randomization yielded P value of .001; (2) an analysis in which both margins were fixed yielded $p=.083$, when complete randomization was assumed; and (3) when the urn-sampling allocation was taken into account, the P value was: .28 for the analysis that conditioned on both margins and .62 for the analysis that conditioned on the observed sequence of responses and the observed treatment totals.

Farewell et al. (1993, p. 26) summed-up the problems of the statistical inference in these trials by concluding "These [conditional and marginal analyses] are statistical difficulties that seem inherent in the adaptive treatment allocation. They therefore seem resistant to a statistical solution."

In using the Bayesian perspective for analyzing data from this design, there are several issues that need to be addressed. One of these is the choice of the prior distributions, which is known to arise in any Bayesian analysis. For an actual clinical trial, choice of these priors should reflect site-to-site population variability as well as different investigators' standards. Furthermore, one needs to consider the robustness of the inference to changes in prior distributions and/or patient populations. Another issue that arises in this analysis is the adjustment for imbalance with respect to (unmeasured) prognostic factors, which is expected to arise in this design, as will be discussed subsequently. The proposed solutions for choice of the prior distributions do not eliminate concerns about the robustness of the inference for the design.

There are other practical issues pertinent to the application of response-adaptive randomization that were articulated by several authors (see, e.g. Armitage 1985; Simon 1977; and references therein). One of these issues is expected imbalances with respect to prognostic factors. In commenting on this, Armitage (1985, p. 21) stated: "Adaptive schemes . . . depart from the standard forms of randomization which provide a safeguard against the possible imbalance of prognostic characteristics in different treatment groups. . . . Indeed, the very existence of a changing allocation proportion may force a change in patient characteristics by discouraging the entry of some types of patients and encouraging that of others: the attitude of investigators towards the admission of marginally suitable patients may, for example, be quite different when they suspect that most patients will receive the better of the two treatments from their attitude at the outset."

The issue of selection bias was also considered by Simon (1977, pp. 746-747) when he stated that "play-the-winner assignment permits the physician to know the treatment assignment before the next patient enters the study. Such knowledge can prejudice the physician with respect to the

type of patient he will admit. This effect has been called selection bias. . . . It would be a giant step in the wrong direction to advocate the use of designs completely susceptible to such bias."

In the response-adaptive design, unlike the traditional design, it is extremely difficult to check for selection bias in the trial. Furthermore, it is almost impossible to check for imbalance with respect to prognostic factors because knowledge of these factors was not required for carrying out the randomization, and consequently these factors were not measured during the trial. For this design, imbalance in prognostic factors is expected to arise when characteristics of the small number of patients assigned to the inferior treatment are not representative of those of the majority of patients assigned to the superior treatment. In particular, this imbalance is expected when there is time trend in the patient population.

Among the other problems related to having a small number of patients in one arm of the trial is the credibility which the clinician, the end-user of the results of the trial, would place on the results of the trial in the presence of departure from the balanced design. Furthermore, the ethical justification for assigning few patients to the inferior treatment by this method has been questioned by many authors (see, e.g., Royall 1991).

In summary, the motivation underlying the proposal of reducing the number of patients receiving an inferior treatment is appealing for the use of the response-adaptive randomization method. However, there are many serious reasons for the reluctance in using this method in practice. Finally, the credibility of this method hinges on adequately answering the points raised in this article.

References

- Andersen J., Faries, D., and Tamura, R. (1994), "A Randomized Play-The-Winner Design for Multiarm Clinical Trials," *Communications in Statistics—Theory Method*, 23, 309-323.
- Armitage, P. (1985), "The Search for Optimality in Clinical Trials," *International Statistics Review*, 53, 15-24.
- Bartlett, R., Roloff, D., Cornell, R., Andrew, A., Dillon, P., and Zwischenberger, J. (1985), "Extracorporeal Circulation in Neonatal Respiratory Failure: A prospective Randomization Study," *Pediatrics*, 76, 479-487.
- Begg, C.B. (1990), "On Inferences From Wei's Biased Coin Design for Clinical Trials" (with discussion), *Biometrika*, 77, 467-484.
- Farewell, V.T., Viveros, R., and Sprott, D.A. (1993), "Statistical Consequences of an Adaptive Treatment Allocation in a Clinical Trial," *The Canadian Journal of Statistics*, 21, 21-27.
- Faries, D., Tamura, R., and Andersen, J. (1995), "Adaptive Designs in Clinical Trials: Lilly Experience," preprint.
- Royall, R. (1991), "Ethics and Statistics in Randomized Clinical Trials," *Statistical Science*, 6, 52-88.
- Simon, R. (1977), "Adaptive Treatment Assignment Methods and Clinical Trials," *Biometrics*, 33, 743-749.
- Tamura, R., Faries, D., Andersen, J., and Heiligenstein (1994), "A Case Study of an Adaptive Clinical Trial in the Treatment of Out-Patients With Depressive Disorder," *Journal of the American Statistical Association*, 89, 768-776.

Rejoinder

Douglas E. Faries, Roy N. Tamura, and John S. Andersen

Lilly Research Laboratories, A Division of Eli Lilly and Company, Indianapolis, IN 46285

We thank each of the discussants for sharing their thoughts concerning adaptive trials. Dr. Teoh raises the idea of adapting randomization probabilities after batches of patients' data becomes available. We think this is an attractive idea and could be implemented in conjunction with the use of data monitoring committees (DMC). The ability to change the randomization probabilities might be useful to a DMC because of the complex ethical position it holds.

Dr. Simon argues that behavior is either consistent or inconsistent with a particular ethical standard, and not rated on a continuum. We are uncomfortable with this position. Ethical standards are rarely straightforward and are likely to change over time. One reason that adaptive trials are controversial is that they challenge the status quo. In this situation, Dr. Lachin's advice in his closing paragraph seems most appropriate.

Drs. Al-Osh and Dubey argue that adaptive designs are particularly susceptible to selection biases. Although both adaptive and nonadaptive studies are subject to selection biases, we feel it is unlikely that the adaptive design itself will introduce selection bias as long as randomization and appropriate blinding are maintained. The suggestion by Drs. Al-Osh and Dubey of updating the urn with different numbers of balls for each of the treatments is interesting and warrants further research.

Letter From the Editors

We'd like to introduce ourselves. We're the new co-editors of the *Biopharmaceutical Report* for the 1995-97 term, Curt Wiltse and Bill Huster. We are planning to have two issues per year—summer and winter. Since this is your newsletter, we encourage your contributions. You can contribute in one of several ways: write the lead article, provide a book review, submit comments concerning this issue, send letters to the editor, etc. We look forward to hearing from you!

Editorial Board

Bill Huster
Eli Lilly and Company
Co-Editor

Curt Wiltse
Eli Lilly and Company
Co-Editor

Eric Sampson
American Statistical Association
Layout and Design

Section News

Letter from the Past-Chair

Biopharmaceutical Section Accomplishments

September 1993 - December 1994

1. Applied Statistics Conference (December 1993)
 - Sponsored three three-hour tutorials and one two-day short course.
2. ENAR Meetings (April 1994)
 - Sponsored five invited paper sessions at the ENAR meetings in Cleveland.
3. Midwest Biopharmaceutical Workshop (May 1994)
 - Co-sponsored the workshop and contributed to eight sessions.
 - Attendees evaluated the conference very highly.
4. Joint Statistical Meetings (1994)
 - Conducted a half-day workshop for new biopharmaceutical statisticians
 - Co-sponsored a short course on resampling-based multiple testing
 - Held three invited sessions (two invited, one competition winner)
 - Organized five special contributed paper sessions
 - Organized 13 regular contributed paper sessions
 - Organized two poster sessions
 - Organized eight luncheon and two non-luncheon roundtables
 - Awards were provided for the three best presentations at Section-sponsored sessions
5. Winter Conference (January 1995)
 - Co-sponsored with the NC Chapter of ASA.
 - The theme of the conference was the "Interface Between Statistical Science and other Scientific Disciplines."
 - Presented a half-day workshop overviewing the role of a statistician in the pharmaceutical industry.
6. Publication of Proceedings
 - Proceedings of major statistical conferences for 1993 are published, and arrangements are well underway for publishing the proceedings for 1994 conferences.
7. *Biopharmaceutical Report*
 - Identified new co-editors, Bill Huster and Curt Wiltse, to succeed Tuli Cnaan.
8. Contributed \$500 to a Data Analysis contest to be held at St. Cloud State University.

Biopharmaceutical Report, Summer 1995

9. Provided \$1000 Outstanding Paper Awards to four graduate students.
10. Reduced Section annual dues by \$2, from \$11 to \$9.
11. Established a subcommittee to evaluate the current Section Charter and to recommend any changes that are deemed appropriate.
12. Continued efforts to search for statisticians to serve on FDA advisory committees.

Submitted by: Dr. Robert Starbuck, Past-Chair
Biopharmaceutical Section

Minutes of ASA Biopharmaceutical Section Executive Committee Meeting August 15, 1994 Toronto, Canada

Attendees:

| | |
|-----------------|-------------------|
| Barbara Bailar | Lillian Kingsbury |
| Janet Begun | Ken Koury |
| Norman Bohidar | Gary Neidert |
| Mike Boyd | Bruce Rodda |
| Nguyen Dat | Denise Roe |
| Bob Davis | Bob Starbuck |
| Charles Davis | Wayne Weng |
| Sally Greenberg | |

Chairman Bob Starbuck announced the newly elected officers. They are: Gary Neidert, Chairman Elect; Steve Snapinn, Program Chairman Elect; and Janet Begun, Publication Officer. Bob also announced the appointment of Bill Huster and Curtis Wiltse as Editors of the *Biopharmaceutical Report* and Liang Yuh as liaison to the Muncie Meeting.

Certification

Barbara Bailar announced that a motion at the ASA Board to enter into discussions with other societies, e.g., the ASQC and the IIE, for participation in a joint program for certification of interested statisticians was subsequently tabled for consideration at the December meeting of the Board of Directors to give an opportunity for input from the whole membership of the Association. The Biopharmaceutical Section was just under 50% affirmative in recommending certification.

1994 ASA Annual Meeting in Toronto

Mike Boyd reported that the Section sponsored a half-day workshop for new biopharmaceutical statisticians. At the workshop, Bruce Rodda and Bob Starbuck spoke on the role of statistics in drug development, Gordon Pledger reviewed FDA

guidelines, and Howard Smith discussed technical writing. In considering whether to repeat the workshop at next year's meeting the Section will review the course evaluations from this year, which the ASA Continuing Education office will provide. Among names of possible presenters suggested were Mike Boyd, Ken Koury, Lillian Kingsbury, and Liang Yuh.

The Section sponsored invited paper sessions in Toronto on pharmacoeconomics issues, non-linear mixed effects, and individual bioequivalence. Ken noted that the Section got its fair share of large meeting rooms this time. Liang Yuh said that eight luncheon round table discussions would be held.

Student Awards

Chuck Davis reported that four Section-sponsored student awards would be presented at the business meeting.

1995 ASA Meeting In Orlando

Joe Heyse is putting together proposals for invited sessions at the 1995 meeting.

Assignment: Anyone who wants to suggest a Section-sponsored short course needs to fill out the requisition forms and have them signed by Bob Starbuck.

Applied Statistics Conference

Norm Bohidar reported that the three-day conference at Merv Griffin's Resorts Casino Hotel in Atlantic City will run from December 12 to December 14, 1994, followed by two two-day short courses.

1995 Winter Conference, Raleigh, North Carolina

Ken Koury reported that the 1995 Winter Conference is being sponsored by the Section and the North Carolina Chapter of the ASA. The meeting will be held Thursday, January 5 through Sunday, January 8 at the North Raleigh Hilton in Raleigh, North Carolina. The theme of the conference is "The Interface between Statistical Science and other Scientific Disciplines." Sessions will address the collaboration of statisticians with other scientists in research, development, and other complex problem-solving settings. Underlying themes include education, cross-training, and recognition of statistical science.

The program committee consists of Ken Koury, Nich Teoh, Liang Yuh, and Joe Heyse from the Biopharmaceutical Section, Perry Haaland from the North Carolina Chapter and Satya Dubey from the FDA.

Midwest Biopharmaceutical Statistics Workshop

Liang Yuh reported that the Program Chair for the 1995 Workshop will be Bob Rathmacher, while Pat O'Meara, Dave Pyne, and Gary Stevens will organize sessions. Bradley Efron of Stanford will address the plenary session on an "Overview of Bootstrap Methods."

Budget

Bob Davis presented the proposed 1995 Budget. Three major changes were approved. First, the committee approved reduced Section dues of \$9.00 (\$8.00 for the Section, \$1.00 for ASA) from the previous \$11.00.

The committee approved \$2000 support for the Winter Meeting and a Section-sponsored mixer at the Orlando meeting, with a suggested cost range of \$2000 to \$4000.

Bob Davis reported that the Section had a cash-on-hand balance of \$66923.36 as of June 30.

Biopharmaceutical Membership

The Section's membership total was 1555 as of July 31, 1994. Of these, 1319 were full members, 205 student members, 11 corporate members, and 20 in other classifications.

ASA Fellow Nominations

Bruce Rodda reported that three Section members, later identified as Paula Norwood, Gordon Pledger, and Steve Ruberg, were named ASA Fellows in 1994. Barbara Bailar noted that the nomination presentations were superior to those of previous years. John Schultz will be an incoming member of the Committee on Fellows.

Candidates for Section Officers

Bob Starbuck reminded the committee that nominations for Section officers will be discussed at the November meeting. The Secretary-Treasurer position, which carries a three-year term, will be on the ballot in 1995.

Publications

Chris Gennings reported that the production of the 1994 section Proceedings has begun. The ASA suggested a pre-publication price of \$25, to which we have agreed. The proposed member price (post publication) is \$33 and for nonmembers, \$50.

As in years past, we have taken the approach that it is beneficial for us to produce the proceedings "jointly with ASA" due to their experience and continuity (and lack of these on our part). Therefore, we have agreed to their suggested page charges (\$15/page) and reprint prices. We will include papers in the proceedings from presentations made at meetings/conferences other than the joint meetings (ENAR, Midwest Biopharm Stat Workshop, Applied Stat Conference). We have agreed to give a 20% discount to member/nonmembers who purchase "complete proceedings sets."

We have invited 54 speakers to publish their papers in our proceedings. These speakers are from all Biopharm-sponsored invited presentations made at ENAR '94, the Midwest Biopharm workshop in May, 1994, and the Applied Stat Conference in Atlantic City in December, 1993.

ASA wants us to decide if we want our proceedings included in a Directory of Published Proceedings produced by InterDok Corporation. Part of the agreement is that we would give InterDok a 20% commission of volumes InterDok sells directly to non-members. The impact is expected to be an increased level of Proceedings sales. ASA will enter into an agreement that allows either party to cancel the agreement with 60 days notice. The committee approved the InterDok proposal.

Council of Sections

Denise Roe and Harji Patel reported that the Council of Sections has addressed several issues during the past year. A summary of these issues follows.

1. Board of Directors' actions on modifying the management structure of the ASA Headquarters office.

The Council provided input from section chairs and section representatives concerning the process that the Board of Directors used to reach the decision made at the Cleveland ENAR meeting. Discussions of the potential restructuring and its approval process are ongoing. The proposal to split the duties at Headquarters was referred to an outside consultant whose report is due in December.

Assignment: Bob Starbuck will send a letter to the ASA Board requesting that we have an opportunity to see the management consultant's report before the Board makes any final decision on reorganization.

2. Sharing of potential profit or loss from CE courses that formerly would have been canceled due to low enrollment.

Previously, the ASA office canceled short courses that did not meet minimum enrollment criteria. This policy has been changed, but sharing of potential profit/loss from section-sponsored courses has not been addressed. The interim recommendation was that the decision to cancel a course should be made by the section, after input from the ASA office on the expected loss from the course based on current enrollment. The section could weigh the potential benefits of the course versus the expected dollar loss. Additional detailed information on the costs of presenting a short course was requested. Discussion of this topic is ongoing.

3. Streamlining the Council of Sections Structure.

A recommendation was made for streamlining interaction between the sections and the Council of Sections representatives to the ASA board. This recommendation may eliminate the position of Section Representative(s) and replace these individuals with the current chair or past-chair. Discussion of this restructuring is ongoing.

4. New Section on Health Policy Statistics.

The Council of Sections representatives voted on approving the establishment of a new section on Health Policy Statistics. The results were to be announced at the Toronto meeting.

Biopharmaceutical Section Charter

Bob Starbuck asked the group if our charter was still on target. He established a committee to evaluate the current charter and recommend any changes deemed appropriate. Sally Greenberg will chair the Charter Committee with Lillian Kingsbury and Gary Neidert as committee members.

Membership Survey

The membership survey was tabled until further notice.

Status of Biopharmaceutical Section Committees

Gary Neidert noted that several standing committees are not active. Although Finance seems under control and Membership seems to be growing without a committee, Continuing Education might need more attention.

Paper Presentation Awards (Post-meeting note)

Lillian Kingsbury announced that the Section's awardees for best biopharmaceutical presentation in 1993 were Ron Helms, \$500; Bruce Turnbull, \$250; and John Weaver, \$250.

Topics for the Biopharmaceutical Roundtable Luncheon at the 1995 Annual Meeting

The following topics will be discussed at the Biopharmaceutical Roundtable Luncheon sessions at the 1995 ASA Annual Meeting in Orlando:

- | | | |
|----|--------------------------------|---|
| 1 | Barber, B.L. | Statistical Issues in Quality of Life |
| 2 | Chinchilli, V. | Issues in Bioequivalence |
| 3 | Chuang-Stein, C. | Testing Dose-Response Relationship Based on Ordinal Data |
| 4 | D'Agostino, R. | Raw Value, Change from Baseline and Percentage Change from Baseline. Which should we use and why? |
| 5 | Hung, J., and Nevius, E. | Dose Response for Continuous Response Outcome in Clinical Trials |
| 6 | Kazempour, K. | Covariate/Baseline Adjustment in Clinical Trials |
| 7 | Offen, W., and Helterbrand, J. | Strategies for Nominal Alpha Level Adjustments for Multiple Endpoints |
| 8 | Pledger, G. | Use of Serum Drug Concentrations in Design or Analysis of Clinical Trials |
| 9 | Salsburg, D. | What Can be Done About "Poor" Investigator Sites in Multiclinic studies? |
| 10 | Ting, N. | Subset Analysis for Phase II/III Studies |

1. Statistical Issues in Quality of Life

There is a growing interest in the use of health-related quality of life (HRQOL) measures for the purpose of comparing two or more pharmaceutical interventions on the basis of safety and efficacy. A typical HRQOL instrument will consist of several domains characterizing dimensions of quality of life thought to be affected by the underlying disease, or the interventions used in the trial, or both. Each domain will usually consist of one or more items or questions.

During this roundtable we will discuss the development stage of a HRQOL instrument, aspects of designing a validation study to assess the measurement characteristics of an instrument, and how to define clinically meaningful change in an instrument.

Beth L. Barber, Biostatistics and Research Information Systems, Health Economic Statistics, Merck Research Laboratories, 10 Sentry Parkway, BL 3-2, Blue Bell, PA 19422; (610) 397-7314 (phone); (610) 397-2931 (fax)

2. Issues in Bioequivalence

Current statistical issues in bioequivalence including distribution-free approaches, comparison of intrasubject

variances, and individual vs. population bioequivalence.

Biostatisticians who participate in the design and analysis in the design and analysis of bioequivalence trials will benefit from this discussion, as well as biostatisticians who want to learn about bioequivalence.

Vernon M. Chinchilli, Ph.D., Professor of Biostatistics, Penn State, College of Medicine The Milton S. Hershey, Medical Center, Hershey, PA 17033; (717)531-4262 (phone); (717)531-4359 (fax)

3. Testing Dose-Response Relationship Based on Ordinal Data

We will review the methods commonly used to test for a dose-response relationship when the data are recorded on an ordinal scale. The merit and the strength of these methods will be discussed.

Christy Chuang-Stein, Ph.D., Upjohn Laboratories, Biostatistics, 7000 Portage Road, Kalamazoo, MI 49001; (616) 384-9654 (phone); (616)385-5286 (fax)

4. Raw Value, Change from Baseline and Percentage Change from Baseline: Which one should we use and why?

In data arising from medical trials we often have both baseline and final outcome data measured on the same variable; for example severity scores before and after treatment with an analgesic. In these cases the final value can be analyzed directly. However, it is possible to use the baseline data to improve the precision of the statistical procedures or to adjust for the lack of balance across groups at baseline. The question then arises as to which form the data should take for analysis. For example, should it be as a difference score or as a percentage of the baseline. For this roundtable we hope to attract individuals who have considered this question not only from the mathematical point of obtaining efficient statistical procedures, but also from

the practical point of analyzing data in a form useful for interpretation.

Ralph B. D'Agostino, Ph.D., Professor of Mathematics/Statistics, and Public Health, Boston University College of Liberal Arts, 111 Cunningham Street Boston, MA 02215; (617)353-8091 (phone); (617)353-8100 (fax)

5. Dose Response for Continuous Response Outcome in Clinical Trials

In most NDA submissions, multidose studies constitute a part of efficacy data base. The commonly used method of analysis is analysis of variance (ANOVA) that aims at estimating and testing the effect of each dose. It is well recognized that useful and reliable dose-response information is essential for providing instruction for use in drug label. This discussion session is proposed for the participants to exchange thoughts and ideas on the following:

- Goals of dose response trials
- Types of dose response trials
- Competing methods of analysis
- Linkage between design and analysis

H.M. James Hung, Ph.D, Division of Biometrics, Center of Drug Evaluation and Research, Food and Drug Administration, HFD-713, Room 6060, Woodmont H, 1451 Rockville Pike, Rockville, MD 20852; (301)594-5436 (phone); (301)594-6593 (fax). Edward S. Nevius, Ph.D, Division of Biometrics, Center of Drug Evaluation and Research, Food and Drug Administration, 5400 Fishers Lane, Rockville, MD 20857; (301)443-4594 (phone); (301)443-9279 (fax)

6. Covariate/Baseline Adjustment in Clinical Trials

The purpose of this discussion is to discuss the importance/usefulness of using "covariates" in analysis of clinical

Request for Names of ASA Fellow Candidates

Bob Starbuck

Past Chair, Biopharmaceutical Section

The Biopharmaceutical Section would like to see all deserving members of the Section receive the prestigious recognition of ASA Fellow. The number of Section members receiving this recognition has been quite small relative to the size of the Section, and may be partially due to an inadequate number of nominations. Accordingly, the Section is encouraging its membership (via this article) to prepare nomination packages for worthy candidates for the ASA Fellow award.

For each candidate, an individual is also needed to perform the vital task of preparing a nomination package for the candidate. This person should be acquainted with the candidate and have the motivation needed to complete the task of assembling the nomination package.

Each year, the Biopharmaceutical Section prepares a letter of recommendation for each of the most outstanding candidates from the Section. If you are preparing a nomination package and would like the Section to consider writing a letter of recommendation for that candidate, please contact: Dr. Robert R. Starbuck, Assistant Vice President, Clinical Biostatistics & Data Management, Wyeth-Ayerst Research, 1787 Jefferson Downs, West Chester, PA 19380-6448; fax: 610-341-2092

The Guidelines for Nominating ASA Fellows can be obtained from ASA Headquarters, 1429 Duke Street, Alexandria, VA 22314-3402.

trials, to share experiences, and to enhance our knowledge of their usage. More specifically, baseline and prognostic variables (i.e. covariates) can be used in controlled clinical trials to increase statistical power. Some investigators use them to adjust for baseline imbalances in baseline characteristics, and some "pre-test" for covariates and they suggest to use the covariates "only" if there is "an imbalance at baseline." Usage of covariates that are incorporated in the analyses after the data are collected and were not specified at the design stage can be controversial. Evaluating the underlying assumptions of analysis of covariance and the impact of violating them are of concern. In some circumstance the impact of violating the assumptions may be so large that including a covariate is worse than not including it, in the sense of precision. These topics and other relevant topics will be discussed in this session.

Kazem Kazempour, Ph.D., Division of Biometrics, U.S. Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857; (301)594-6259 (phone); (301)594-6289 (fax)

7. Strategies for Nominal Alpha Level Adjustments for Multiple Endpoints

Much has been published on how to adjust P values when one has multiple endpoints. The simplest and most well-known method is the Bonferroni adjustment, where the computed P values are compared to the alpha level divided by the number of endpoints. This method has been extended, for example by Holm's sequentially rejective Bonferroni procedure, and other methods that take into account the correlations between the multiple endpoints. These techniques are appropriate when at least one of the multiple endpoints must achieve statistical significance to conclude that a treatment effect exists.

What has not been discussed much in the literature is the case where *all* the multiple endpoints must achieve statistical significance. In the case of two endpoints that are uncorrelated, it is easy to show that one should test each hypothesis at the nominal level of .2236 to maintain an overall Type I error rate of .05.

A clinical example of a case where an effect must be demonstrated in two outcome variables is in the Alzheimer's therapeutic area. In this case, one must demonstrate a positive drug effect upon cognition, and also a positive effect on a clinical global scale.

The discussion will include what to do when there are multiple endpoints and at least one must be "significant," multiple endpoints where all or more than one must be "significant," and multiple comparisons of three or more treatment groups.

Walter W. Offen, Ph.D., Senior Research Scientist, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285-2233; (317)276-4865 (phone); (317)277-3220 (fax). Jeffrey D. Helterbrand, Ph.D., Senior Statistician, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285-2233; (317)277-2241 (phone); (317)277-3220 (fax)

8. Use of Serum Drug Concentrations in Design or Analysis of Clinical Trials

Phase II/III drug trial protocols often include periodic

measurement of serum drug concentrations. Typically the planned use of these data is vaguely stated, e.g., "to explore relationships between drug concentrations and response variables." When should concentration data be collected, and what should be the goals and uses of these data? To address these goals do we need (or when do we need) different designs, e.g., concentration-controlled designs?

Gordon W. Pledger, Ph.D., Senior Research Fellow, Biostatistics Department, R.W. Johnson Pharmaceutical Research Institute, Route 202, Box 300, Raritan, NJ 08869; (908)704-4530 (phone); (908)725-5087 (fax)

9. What Can be Done About "Poor" Investigator Sites in Multi-Clinic Studies?

The objective is to explore ways in which the analysis of multi-clinic studies can be improved by accounting for differences in the "quality" of investigators.

There are always investigator sites with more missing values, higher drop-out rates, questionable adherence to entrance criteria, inappropriate collection of measures, and more subtle problems. Can the statistical analysis of these studies be modified to account for the "quality" of the investigator? How do we identify "poor" investigators? How do we measure the "quality" of an investigator? The discussion will involve techniques suggested by Tukey, Cochran, Rubin, and others.

David Salsburg, Ph.D., Statistical Consultant, 22 Holly Terrace, New London, CT 06320; (203)443-2577

10. Subset Analysis for Phase II/III Studies

Subset analyses are constantly applied in Phase II/III studies. Most of these analyses are performed for efficacy-related reasons. However, in some cases, the subset analyses are motivated by safety related reasons.

We will discuss the necessity of running subset analyses. For examples, the ITT analysis vs. the Completer analysis and subpopulation analyses grouped by the demographic data. Other topics, including the multiple comparison adjustment and the metaanalysis of subpopulation, will also be discussed.

Naitee Ting, Ph.D., Associate Director, Biostatistics, Pfizer, Inc., Eastern Point Road, Groton, CT 06340; (203)441-4871 (phone); (203)441-3219 (fax)

Announcement!

Biopharmaceutical Section Business Meeting in Orlando

There will be a business meeting of the Biopharmaceutical Section and a reception at 6 pm on Tuesday, August 15, at the ASA Annual Meeting in Orlando

Notes of ASA Biopharmaceutical Section Executive Committee Meeting

March 28, 1995
Birmingham, Alabama

This report is based on Lilliam Kingsbury's notes. Formal minutes will follow.

Attendees:

| | |
|-----------------|-------------------|
| Mike Boyd | Lilliam Kingsbury |
| Bob Davis | Gary Neidert |
| Chuck Davis | Denise Roe |
| Sally Greenberg | Lianng Yuh |
| Joe Heyse | |

Midwest Biopharmaceutical Statistics Workshop

Lianng Yuh reported on the upcoming MBSW, to be held May 22-24 at Ball State University in Muncie, IN. Bradley Efron will have a tutorial on bootstrapping. A highlight is the banquet address by Victor Cohn, a journalist who authored *News & Numbers* and is the first Fellow of the ASA Office of Scientific and Public Affairs; the title of his address is "Statistics for Journalists."

Summer Meeting

Joe Heyse reported that the Summer Meeting will have (a) 95 papers in three invited paper sessions (two allocated and one from competition), three contributed paper sessions, and two FDA invited paper sessions organized by Satya Dubey; (b) 63 contributed papers in 12 sessions; and (c) 6 poster sessions. Three awards will again be given.

Biopharmaceutical Statisticians Workshop

Bob Starbuck and Bruce Rodda presented in Raleigh. There will be 10 round table discussion sessions.

Applied Statistics Conference

This will be held December 11-15 at Merv Griffin's Resort Casino Hotel. There will be four three-hour tutorials and two 1-day short courses.

Budget

A plan for bringing balance to two years' operating budget surplus will be put in place by the end of the year. Our goal is to provide services to the membership by supporting one of more of the following (although additional ideas are welcome):

- Mixer at the Summer Meeting, where there could be an advertisement at the Section table.
- Workshops on specific topics, for which non-industry members could have travel costs covered. Topic suggestions included (a) void created by PMA (possibly invite FDA membership), (b) software usage/SAS procedures (address master's statisticians needs), and (c)

interactive video capability (should this piggyback to a meeting like ENAR?).

- Special lecture at meetings, for which presenters would be paid.
- Speakers at invited paper sections and perhaps special contributed papers. This could extend to other conferences.
- Cover the page charge for members publishing in *Proceedings*.
- Brochure with career opportunities as a means to attract people to the Section.

ASA

A commitment of \$1,500 for an Internet communication network, to be matched by ASA, was approved. Denise Roe will communicate with the Council of Sections. Bob Davis should be informed of requirements.

Ideas for *Amstat News* and the Section newsletter should be communicated to Sally Greenberg.

Book Review

***Randomization and Monte Carlo Methods in Biology.* Bryan F. J. Manly, Chapman and Hall, 1991, x + 281 pp.**

Reviewed by Boris Iglewicz

Director, Biostatistics Research Center, Temple University

Although statistical tests based on randomization concepts have been used during the early stages of statistical development (e.g., R. A. Fisher, *The Design of Experiments*), the present rapid increase in computer power makes it practical to use such tests for a variety of applications. Randomization tests are frequently called permutation tests. Such tests are becoming easier to apply with their incorporation into standard statistical packages. For example, StatXact, from Cytel Software Corporation, contains a variety of permutation tests that are useful in clinical trials. This cycle of development will encourage increased teaching of randomization tests in standard statistics courses, thus leading to further popularity and increased usage.

For a trivial example of randomization tests consider a die that is suspected to favor large values. Three subsequent tosses of the die results in a sum of 17. There are four possible ways that a sum of at least 17 can occur out of 216 total possible realizations. Thus, the null hypothesis of a fair die can be rejected at a one sided P value of .0185. Clearly, the same concept can be used in more complex and far more practical situations. Basically, the P value of a randomization test is obtained by counting all the realizations that lead to at least as extreme a test statistic divided by all possible ways the sample could have been obtained. The advantage of having such tests in common statistical packages is that these can be based on clever combinatorics and methods for reducing the amount of

Corporate Members

The Biopharmaceutical Section has three types of memberships: regular, student, and corporate. The corporate members of the Section, who are also ASA corporate members, provide the Section with \$300 annually in dues. These funds contribute significantly to our annual budget and to our ability to sponsor Section activities.

The current Corporate members of the Biopharmaceutical Section are:

Burroughs Wellcome
 Hoechst-Roussel
 John Wiley & Sons
 Merck
 Pharmaceutical Research Associates
 Proctor & Gamble
 R. W. Johnson PRI
 Sandoz Pharmaceuticals
 The Upjohn Company
 Warner-Lambert/Parke-Davis
 Wyeth-Ayerst Research

The Biopharmaceutical Section gratefully acknowledges its Corporate members and their support.

enumeration. The alternative is always available to estimate the P value from, say, 5,000 random simulations from the sample data. For example, for a two-sample test based on n_1 and n_2 observations, respectively, the $n_1 + n_2$ observations can be randomly subdivided into samples of size n_1 and n_2 . These simulated results can then be used to estimate the P value.

Randomization and Monte Carlo Methods in Biology is a welcome addition to this topic. It reads well and can be easily followed by statisticians and other applied workers with a reasonable statistics background. The writing style is clear and enjoyable. The mathematical level is deliberately held low, with sophistication reserved to the nice selection of references.

There are many interesting practical examples, but these are largely from the paleobiological area. That should not detract from their practical use, as similar medical research applications are apparent.

The book's level and emphasis is not made without sacrifice. As an example, application to regression analysis is discussed without mentioning the H matrix or other aspects of the large collection of diagnostic tools. Similarly, robust regression is barely mentioned. As another illustration, the discussion of the two-sample t test in comparison with the Wilcoxon rank sum test does not mention that for heavier-tailed nonnormal distributions the t test frequently has a low significance level, thus lower power, as compared with the nonparametric Wilcoxon test. Additionally, emphasis on biological examples means that some recommendations need to be carefully reviewed by biopharmaceutical statisticians prior to use. As an example, consider the following quote from page 116 "They quote one example where 1,000, 1,500, 2,000 randomizations give estimates of 4.7, 6.0, and 4.8%, respectively, when the true significance level is 5.1%. The difference between these three estimates are [sic] minimal if the significance level is merely regarded as a measure of the strength of evidence against the null hypothesis." It is not clear if the FDA or government medical research granting agencies would be comfortable with that view. Clearly, use of a larger number of simulated samples is recommended for this situation. In this connection, the reviewer was surprised by the absence of the mention of "P value," although that concept is frequently used.

The coverage includes application to one and two sample tests, analysis of variance, regression, special data, time series, and multivariate methods. The discussion and examples of distance matrices is quite thorough and interesting. Also discussed are other computer-intensive methods, such as the jackknife, bootstrap, and the Monte Carlo method defined by the author as similar to randomization tests, but the simulated samples are generated from a prespecified model. The book provides brief, but fair, descriptions of the latter methods. In addition, a number of FORTRAN programs are included.

Overall, this is a well written and enjoyable introduction to a statistical area that will find far heavier future use. The more sophisticated reader may look for greater depth, but this can be found in the references. The material provides a deeper understanding of the advantages and limitations of standard procedures and a fine introduction to permutation tests and other computer-intensive statistical methods.

ASA Biopharmaceutical Section Executive Committee & Section Representatives

Past-Chair:

Robert R. Starbuck, Ph.D.
 Assistant Vice President, Clinical Biostatistics & Data Management
 Wyeth-Ayerst Research
 145 King of Prussia Road, B-2
 Radnor, PA 19087-4517
 phone: 610-341-2070
 fax: 610-341-2092

Chair:

Lillian Kingsbury, Ph.D.
 Director of Biostatistics
 Bio-Pharm Clinical Services
 512 Township Line Road
 Blue Bell, PA 19422
 phone: 215-283-0770
 fax: 215-283-0733

Chair-Elect:

Gary L. Neidert, Ph.D.
 Director of Clinical Data Management
 The Upjohn Company
 9165-298-139
 Kalamazoo, MI 49001
 phone: 616-329-8591
 fax: 616-329-5579

Secretary and Treasurer:

Robert L. Davis, Ph.D.
 Director of Biostatistics
 Astra/Merck Group
 725 Chesterbrook Blvd.
 Wayne, PA 19087-5677
 phone: 610-695-1070
 fax: 610-889-1288

Program Chair:

Joseph F. Heyse, Ph.D.
 Director, Health Economics Statistics
 Merck Research Laboratories, BL=32
 Sumneytown Pike
 West Point, PA 19486
 phone: 610-397-2334
 fax: 610-397-2931

Program Chair -Elect

Steven M. Snapinn, Ph.D.
 Senior Investigator
 Merck Research Laboratories
 BL3-2
 West Point, PA 19486
 phone: 610-397-2935
 fax: 610-397-2931

Executive Committee Member:

Michael N. Boyd, Ph.D.
 Director, Biostatistical Services
 Pharmaceutical Research Associates
 2400 Old Ivy Road

Charlottesville, VA 22903-4826
 phone: 804-977-2772
 fax: 804-293-7084

Executive Committee Member:

Shein-Chung Chow, Ph.D.
 Director, Biostatistics & Data Management
 Bristol-Myers Squibb Company
 777 Scudders Mill Road
 Plainsboro, NJ 08536
 phone: 609-897-2736
 fax: 609-897-6050

Executive Committee Member:

Charles S. Davis, Ph.D.
 Associate Professor
 Dept. of Preventive Medicine
 University of Iowa
 2837 Steindler Bldg.
 Iowa City, IA 52242-1176
 phone: 319-335-9625
 fax: 319-335-9200

Executive Committee Member:

Sally A. Greenberg, Ph.D.
 Statistical Research Director
 Syntex Development Research
 3401 Hillview Ave., A6-100
 Palo Alto, CA 94304-1397
 phone: 415-813-4833
 fax: 415-813-4844

Executive Committee Member:

C.S. Wayne Weng, Ph.D.
 Statistical Manager
 Schering Laboratories
 2000 Galloping Hill Road
 Kenilworth, NJ 07033
 phone: 908-298-5457
 fax: 908-298-5700

Executive Committee Member:

Lianng Yuh, Ph.D.
 Director of Biometrics
 Department of Biometrics
 Pfizer Central Research
 Eastern Point Road
 Groton, CT 06340
 phone: 203-447-1531
 fax: 203-441-3600

Representative—Applied Stat Conf.:

Norman R. Bohidar, Ph.D.
 Biometrician
 Merck Research Laboratories
 BL3-2
 West Point, PA 19486
 phone: 610-397-2981
 fax: 610-397-2931

Publications Officer:

Janet M. Begun, Ph.D.
Clinical Data Services, Inc.
8124 Brookwood Court
Raleigh, NC 27613
phone: 919-787-2748
fax: 919-787-5457

Chair of CRSFAC:

Vernon M. Chinchilli, Ph.D.
Data Coordinating Center
Center for Biostatistics & Epidemiology
Pennsylvania State University
M.S. Hershey Medical Center
P.O. Box 850
Hershey, PA 17033
phone: 717-531-4262
fax: 717-531-4359

Biopharm Report Co-Editor:

William J. Huster, Ph.D.
Senior Statistician
Lilly Corporate Center, 2233
Indianapolis, IN 46285
phone: 317-276-9802
fax: 317-277-3220

Biopharm Report Co-Editor:

Curtis G. Wiltse, Ph.D.
Research Scientist
Lilly Corporate Center, 2233
Indianapolis, IN 46285
phone: 317-276-5773
fax: 317-277-3220

Council of Sections Rep.:

Harji I. Patel, Ph.D.
Biostatistics & RIS
Berlex Laboratories, Inc.
300 Fairfield Road
Wayne, NJ 07470-7358
phone: 201-305-5386
fax: 201-694-9093

Council of Sections Rep.:

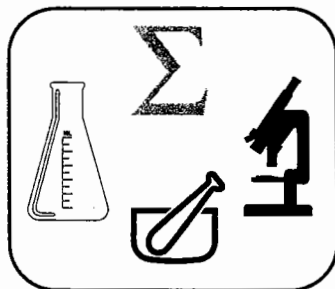
Denise J. Roe, Dr.P.H.
Assistant Professor
University of Arizona
Arizona Cancer Center
1515 North Campbell Avenue
Tucson, AZ 85724
phone: 602-626-2281
fax: 602-626-2284

Let's Hear from You!

If you have any comments or contributions, contact Co-Editors William J. Huster, Eli Lilly and Company, Lilly Corporate Center, 2233, Indianapolis, IN 46285; phone: (317) 276-9802; fax: (317) 277-3220; e-mail: huster@lilly.com or Curt Wiltse, Lilly Corporate Center, 2233, Indianapolis, IN 46285; phone: (317) 276-5773; fax: (317) 277-3220; e-mail: wiltse_curtis_g@lilly.com

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

© 1995 The American Statistical Association

**Biopharmaceutical Report**

c/o American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3402
USA

FIRST-CLASS MAIL
U.S. POSTAGE
PAID
WASHINGTON, D.C.
PERMIT NO. 9959

FIRST CLASS POSTAGE

DR. ROBERT L. DAVIS
215 MANOR ROAD
HARLEYSVILLE PA 19438