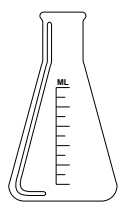


Biopharmaceutical Section



American Statistical Association

Biopharmaceutical Report

Volume 6, No. 2

Summer 1998

Chair: *Ken Koury*

Editors: *Anne Meibohm, Ersen Arseven, and Curtis Wiltse*

Practical Issues in Linear Models Analyses in Multicenter Clinical Trials

Paul P. Gallo

Novartis Pharmaceuticals

1. Introduction

Clinical trials comparing treatment therapies are commonly conducted at a number of different sites. This is done primarily to enroll a sufficient number of patients to investigate a hypothesis in a timely manner. A beneficial consequence is that this can provide evidence that the trial results are not strongly setting-dependent, so that it may be reasonable to interpret that they apply to a broader population of clinical sites or patient populations.

An aspect of multicenter trials with important implications for data analysis is that it is a natural consequence of the manner in which these trials are conducted that the within-center sample sizes can be quite unequal. Trials in which the largest center contributes four or five times as many patients as the smallest are quite common, and larger imbalances are often encountered. To enforce similarity of center sample sizes can substantially increase trial duration (Senn, 1997) and increase potential for subtle selection biases as well.

The methodology considered here is classical linear models analysis, in which an approximately normally distributed response can be reasonably modeled as a linear function of predictors, with a common error variance. There are differences of opinion in the pharmaceutical industry concerning some aspects of linear models analyses in multicenter trials, specifically the following interrelated topics: inclusion of a treatment-by-center interaction term in the statistical model; the weights given to individual centers in test statistics for treatment effects; and the manner of computing ANOVA-framework test statistics, most frequently manifested in choices among different "types" of sums of squares available in PROC GLM in the SAS statistical system.

There has recently been renewed interest and activity reflected in the literature on this topic: recent worthwhile discussions of statistical issues in multicenter trials are provided by Senn (1997), Källén (1997), Snapinn (1998), Jones et al. (1998), and Lin (1999). Here, we consider some specific practical and computational issues and their implications. The impact of center size imbalance on various analysis options (such as those mentioned above) is shown to be quantifiable. A specific issue addressed is that for a commonly employed analysis method ("Type III with interaction"), inequity among center sizes invalidates the method by which sample size is routinely calculated. We illustrate how this knowledge may be used to quantify power loss, modify sample size if necessary, or consider alternative analysis options.

2. Determination of a primary analysis model

In the pharmaceutical industry, it is conventional to pre-specify a primary hypothesis, and the details of a primary analysis, based upon pre-study expectations and assumptions concerning the most appropriate manner of investigating the hypothesis. Though the data themselves may provide valuable clues concerning how they

Contents

FEATURED ARTICLE

Practical Issues in Linear Models Analyses in Multicenter Clinical Trials.....GALLO 1

BIOPHARMACEUTICAL SECTION NEWS

1998 JSM ActivitiesKOURY 9

Call for Special Contributed Sessions and Invited Posters at the 1999 JSMCHUANG-STEIN 10

Deming Conference on Applied Statistics Has a Lot to OfferSTOCK 11

Corporate MembersPICHOTTA 11

1998 ASA Fellows 12

Best Student Paper Award.....ZHANG 12

1997 Treasurer's Report...MEEKER 13

Minutes of the March 31, 1998 Executive Committee Meeting.... 13

see MULTICENTER, page 2

MULTICENTER, *continued from page 1*

should most appropriately be analyzed, and a number of analyses will generally be performed as part of a thorough examination of the data, results of the primary analysis will generally receive the most emphasis in initial consideration and interpretation of trial results.

In determining a primary analysis model, there will often be a number of variables, in addition to treatment, which are candidates for model inclusion based upon their potential for being predictive of response. Baseline values for the response are usually particularly strong candidates; others may reflect measures of patient well-being at the start of the study, past medical history, and sometimes demographic factors or physical characteristics (e.g., age, gender, weight).

While there might be many such quantities which could be considered possibly predictive of response, it is usually felt to be desirable to keep primary models fairly simple, both for ease of interpretation, and for methodological reasons. (Depending on the analysis method used, models saturated with many predictors may have problems with few error degrees of freedom, small cell instability, poor asymptotic approximations, etc.) Thus, primary analysis models often contain a limited number of factors for which

there is the greatest consensus for an expected strong relationship to response. Other factors may be considered "exploratory," with their influence possibly addressed subsequently, but not in the primary analysis model.

Historically, a frequent exception to this process has been factors involving center, such as treatment-by-center interaction. 'Center' is commonly included in models as a main effect since it has often proven to be meaningful, and to focus attention on important aspects of the data, even when the nature of the effect was not anticipated. (Thus, center might be considered an "explanatory," rather than exploratory, effect.) Also, since randomization is often stratified within center, it may be considered good statistical practice to include this term.

Treatment-by-center interaction is nearly always an exploratory factor; we rarely undertake a trial with a clear expectation regarding the nature of different effects expected in different centers. If this were the case, then it would be very important to determine in advance the correct "mix" of centers with different attributes we might desire in order to investigate a treatment hypothesis, since the results could be very dependent upon the actual centers used. If a trial is conducted in 20 centers, for example, we may feel that we've employed a sufficiently broad range of settings for the results to be properly interpreted; we generally don't worry about the types of centers which might have been "missed" (for example, had the trial been run in 30 centers), and how this might have affected the results. Power considerations for treatment-by-center interaction are rarely considered in trial design. (See Lewis, 1995: "The power to look at interactions is ...

small, and apparent trends in the "wrong" direction at individual centers must be quite likely, so that they should generally be disbelieved. Indeed the decision to carry out a multicenter trial requires the *assumption* that the treatment effect is similar from center to center, because we will be unable to test this adequately at the end of the study. Undoubtedly there will be some minor differences in the main treatment effects at the different centers ... but substantial treatment-by-center interactions should be just as unlikely as any other sort of treatment interaction and are not our main interest.")

In addition, almost uniquely among usual candidates for model inclusion, treatment-by-center interaction has potential for impairing the analysis results, specifically by inflating the variance of the estimate on which the treatment test is based when centers are not of equal size. This will be illustrated and discussed below.

Support for the practice of including the interaction comes from an interpretation of the 1988 FDA statistical guidelines, associated with the legitimate concern that consistency or discordance of outcomes across clinical settings can be important in understanding how, and how well, a treatment works, as well as a feeling that this method correctly "treats centers equally," in a sense which will shortly become more clear.

Common industry approaches concerning inclusion of treatment-by-center interaction in primary analysis model specification include the following:

- (a) include this interaction term in the model;
- (b) do not include it (possibly addressing the possibility of meaningful interaction in supplemental analyses);
- (c) a hybrid of approaches (a) and (b), in which the interaction is initially included, but a preliminary significance test for the interaction is performed; if the term is not significant (a level of 0.10 is often used), it is removed prior to performing the test for treatment.

(Other approaches may also be considered; an intriguing alternative in certain situations is mixed models analysis, in which effects involving center are considered to be random. This will not be considered here.)

The descriptions above do not completely identify the analyses, since associated with any particular model, there are different ways to construct hypothesis tests. In the U.S. pharmaceutical industry, a common practice (quite possibly the majority practice) is (a), with the test for treatment constructed using the default method of calculation of sums of squares ("Type III") in the procedure PROC GLM in SAS.

When the interaction appears in the model, lack of statistical significance of this term is generally seen as validating the result of the main effects test for treatment, despite the lack of power considerations for the interaction. When the interaction is significant (there seems not to be an industry standard significance level criterion), an attempt is generally made to investigate and "explain" the interaction. For example, it might be addressed whether this is due to an atypical

The decision to carry out a multi-center trial requires the assumption that the treatment effect is similar from center to center, because we will be unable to test this adequately at the end of the study. Undoubtedly there will be some differences in the main treatment effects at the different centers ... but substantial treatment-by-center interactions should be just as unlikely as any other sort of treatment interaction and are not our main interest.

–J. A. Lewis

center (i.e., the interaction would be insignificant if data from such a center were excluded), particularly a small center in which the apparent difference might more easily be dismissed, or a center in which a reason for its difference can be reasonably postulated. It may be investigated whether the apparent interaction results from confounding of center with important predictors which may not have been included in the model.

Interactions are sometimes characterized as "quantitative" or "qualitative," depending on whether or not within-center treatment effects seem to operate in the same direction, the idea being that quantitative interactions might not be considered as important a concern. In multicenter trials, though, it is often the case that many centers are so small that such a distinction realistically cannot be made formally (as, for example, in Gail and Simon, 1985). Often, this is addressed simply by looking at point estimates of within-center treatment effects and seeing if they have the same sign. The variability of these estimates frequently limits the usefulness of this approach; apparent qualitative interactions in this sense are generally quite likely, even when there is no real interaction (Senn, 1997).

3. Center weighting

For ease of illustration, we restrict consideration to a comparison between two treatment groups using a model containing no terms other than treatment group, center, and possibly the treatment-by-center interaction. The general concepts described below, if not all numerical details, apply in more general models. We let n_{ij} denote the number of patients in center j assigned to treatment group i , $i=1, 2, j=1, \dots, c$; also, $n_j = n_{1j} + n_{2j}$, and $n = \sum n_j$. Let D_j denote the true mean difference in the response variable between treatments 1 and 2 in center j , with d_j denoting an estimate of this quantity; in our simple situation, d_j is the difference of the mean responses in the two treatment groups for patients in center j .

With interaction in the model (in the SAS Type III sense), analysis is equivalent to inference based upon

$$d_U = \sum c^{-1} d_j$$

(equivalent in the sense that a t -test using this estimate and its estimated standard error provides identical inference to the F -test obtained from the analysis of variance).

With the interaction excluded, analysis is based on

$$d_W = \sum w_j d_j$$

where w_j is inversely proportional to the variance of d_j :

$$w_j = (n_{1j}^{-1} + n_{2j}^{-1})^{-1} / \sum (n_{1j}^{-1} + n_{2j}^{-1})^{-1}.$$

so that each within-center estimate is given a weight based upon its precision. Thus, inclusion or exclusion of the interaction term can have much more of an impact than simply explaining variability in the data: it can fundamentally alter the weighting associated with factor levels and possibly very much change the nature of the inference being performed. (Note: in SAS PROC GLM, use of Type II sums of squares provides inference based upon d_W even when the interaction appears in the model, i.e., the Type II sum of squares for treatment, whether or not interaction is in the model, is identical to the Type III sum of squares in a no-interaction model.)

Our focus will be on these most commonly used forms of estimators, rather than on entries in analysis of variance

tables. Henceforth, d_U and d_W will be referred to as the "unweighted" and "weighted" estimators, respectively. This choice of terminology can itself be somewhat misleading, implying that 'center' is the relevant unit for inference rather than 'patient;' weighted estimators tend to equalize weights given to individual patient responses, while unweighted estimators tend to weight patients according to the sizes of their centers. (Note that if each $n_{1j} = n_{2j}$, d_W is simply the difference of the overall treatment group means.)

It is generally understood that the weighted estimator is the more precise of these two types. In this context, Jones et al. (1998) and Lin (1999) have presented simulation results demonstrating that Type II analyses are generally more powerful than Type III when centers are not of equal size. In fact, the increased efficiency of weighted inference can be addressed analytically, and quantified as a function of a measure of center size imbalance.

Without regard to any particular model, and with σ^2 denoting the common variance of each observation, the variances of the estimators d_U and d_W are, respectively,

$$V_U = \sigma^2 \sum \sum n_{ij}^{-1} / c^2$$

$$V_W = \sigma^2 / (\sum (n_{1j}^{-1} + n_{2j}^{-1})^{-1}).$$

V_U is always at least as large as V_W , with equality only when the treatment-center combination sample sizes are identical. For the types of center imbalance encountered in clinical trials, their difference can be substantial, as will be illustrated in the following section.

Variance is, of course, not the sole basis on which the estimators can be compared, since they may estimate different quantities (say, D_U and D_W , defined analogously to d_U and d_W). Advocates of unweighted analyses often claim that it is appropriate to treat centers equally, by giving estimated effects from each center equal weight in the sense achieved by d_U (though Källén, 1997 and Lin, 1999 present some persuasive arguments against this interpretation).

It is not at all clear, though, that in practice the unweighted and weighted approaches are routinely based upon meaningfully different versions of an overall treatment effect. First of all, if interaction does not exist (or, for practical purposes, is minimal), then the two approaches are based upon the same parameter, which the weighted method estimates more precisely.

Furthermore, even if there is a treatment-by-center interaction, but there is nothing systematic about the relationship between center size and within-center effect, then both approaches are based upon estimates of essentially the same parameter. To make this more clear, consider the following scenarios illustrated in Fig. 1, which present the actual center effect D_j relative to center size. In case (i), there is no interaction, so any weighting scheme produces an estimate of the common treatment difference. In (ii), there is an interaction, but the treatment effect is uncorrelated with center size, and the weighted and unweighted effects are again identical. In case (iii), there is an interaction systematically related to center size; in this case, the weighted and unweighted approaches are based upon different parameters, and their difference describes an important aspect of the nature of the interaction.

A main point here is that it may frequently be the case that simple averages and weighted averages are quite close to each other; in the current context, efficiency would thus become a

more important basis for comparison. If treatment-by-center interaction is itself an exploratory effect, the existence of an interaction for which weighted and unweighted treatment effects are substantially different is all the more exploratory. When the weighted and unweighted effects are meaningfully different, it seems difficult to conclude that either is inherently "better" (except from an efficiency standpoint, for which the answer is clear); the difference may be saying something important about the nature of an interaction (e.g., a larger effect in large centers) which hopefully should be determined from a thorough examination of the data. More will be said about comparing unweighted and weighted estimators in Section 7.

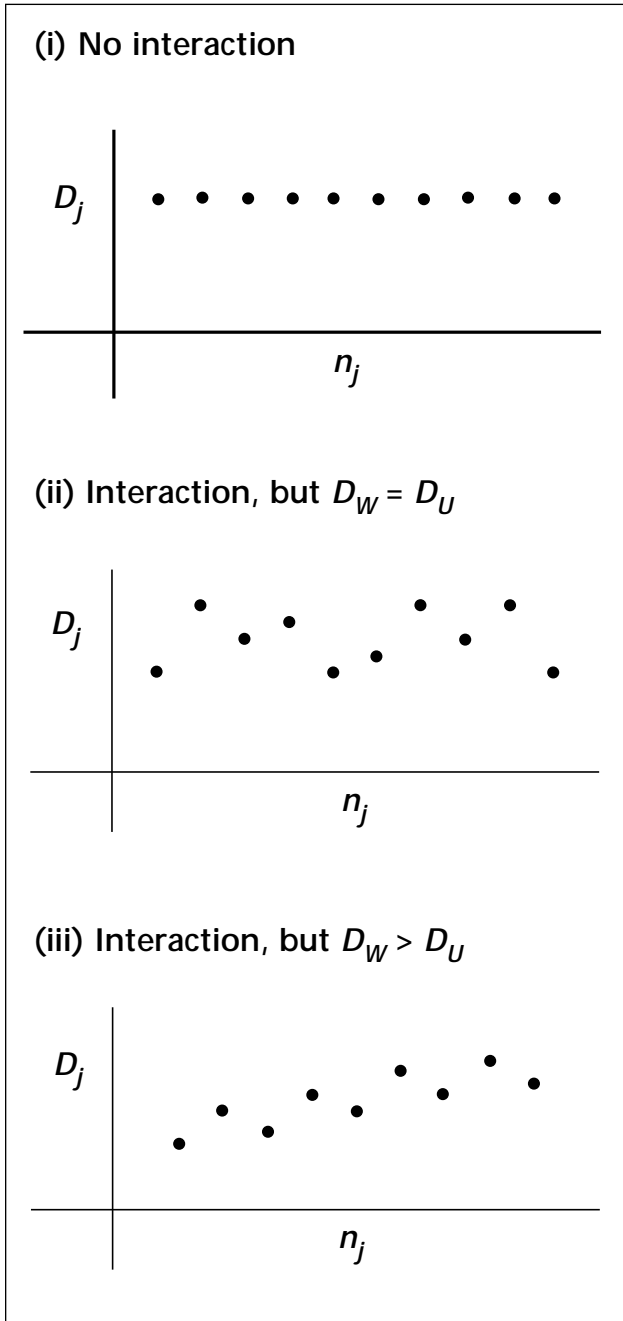


Figure 1. Within-center effect D_j vs center size.

4. Power and sample size

In assessing the impact of center size inequality on the relative efficiencies of the estimators, it will be instructive to do this in parallel with a consideration of how trials are usually designed, with a sample size chosen to achieve specified power considerations. When a linear model is to be employed and there is roughly equal allocation to two treatment groups, the total sample size n is conventionally determined as follows:

$$n = 4 (z_{1-\alpha/2} + z_{1-\beta})^2 / \Delta^2,$$

where z_p is the p th percentile of a standard normal distribution, and Δ is a standardized treatment difference it is desired to detect with power $1-\beta$ using a one-sided level $\alpha/2$ test. (Refinements to this formula can be made, for example, using the t -distribution to account for degrees of freedom dependent on the design and model; here, we consider only the large-sample normal approximation.)

The validity of the sample size calculation depends upon the assumption that the variance of the overall treatment difference estimator is

$$V = 4\sigma^2 / n.$$

In fact, this is the minimum possible variance based upon n patients; the actual variance depends on the distribution of patients across centers and treatment groups. For the weighted estimator, the bound is achieved only if the within-center treatment group sample sizes are equal, i.e., $n_{1j}=n_{2j}$. In clinical trials, though, it is generally the case that treatment allocation is nearly equal within each center because of the manner in which the randomization is carried out, and the variance bound is usually nearly achieved (i.e., V_W is very close to V). Unweighted estimators, on the other hand, do not achieve this bound unless additionally the center sample sizes are the same (in which case the weighted and unweighted estimators are identical).

The following ratio of V_U to V can be interpreted as the inefficiency of the unweighted estimator relative to a number of quantities: to an estimator from a balanced design with the same number of patients; to the weighted estimator (which usually has nearly full efficiency); or to the value assumed in sample size calculations:

$$I_U = n \sum \sum n_{ij}^{-1} / 4c^2.$$

A similar quantity I_W can be computed for the weighted estimator; however, as mentioned above, unless there is substantial treatment imbalance both within and between centers, this quantity will generally be close to 1.

For a particular set of center sizes $\{n_{ij}\}$, I_U thus provides a measure of how much the conventional sample size should have been inflated (with center sizes remaining in the same proportion) in order for an unweighted analysis to achieve the desired power. Furthermore, it is straightforward to show that if the sample size had been calculated to achieve power $1-\beta$, then the actual power for an unweighted analysis with center sizes associated with an inefficiency of I_U is

$$\Phi ((1-\sqrt{I_U}) z_{1-\alpha/2} + z_{1-\beta}) / \sqrt{I_U}$$

where Φ denotes the standard normal distribution function.

Such quantities can be easily calculated in particular situations. To give a general idea of the impact of different forms and levels of imbalance on sample size and power, consider a trial with four centers and 100 patients in each of two treat-

ments. Table 1 describes the inefficiency of unweighted estimation and the actual power if the sample size had been computed in the conventional manner with a desired power of 80%.

Thus, for example, in the situation of the next to last line of Table 1, if one center is one-third as large as each of the other three, the variance of the unweighted treatment effect estimator is 25% larger than in any equal allocation distribution. Based on conventional calculations, the number of patients enrolled should have been increased by 25% in order to yield the desired power of 80%; otherwise the power achieved would have been only about 70%. Focusing further on this example, had the small center not been part of the trial, the efficiency would only have been reduced by about 10% from the optimal efficiency associated with this number of patients, since we would then have an equal allocation (hence, maximally efficient) trial with 10% fewer patients. Inclusion of the extra data impairs efficiency by 25% and is thus substantially worse than ignoring it, at least based upon the precision of the estimated treatment effect. (This is a different version of a paradox pointed out by Senn, 1997.)

Table 1. Inefficiency and power for various patient-center distributions

| Distribution of patients | Inefficiency* | Power |
|--------------------------|---------------|-------|
| 25, 25, 25, 25 | 1 | 80.0% |
| 20, 20, 30, 30 | 1.042 | 78.4% |
| 10, 10, 40, 40 | 1.563 | 61.1% |
| 10, 20, 30, 40 | 1.302 | 69.0% |
| 20, 20, 20, 40 | 1.094 | 76.4% |
| 10, 10, 10, 70 | 1.964 | 51.6% |
| 10, 30, 30, 30 | 1.250 | 70.7% |
| 4, 32, 32, 32 | 2.148 | 48.1% |

* of unweighted inference, relative to a balanced design

If one plans to use an unweighted primary analysis, it would thus seem warranted to account for the anticipated center imbalance when determining sample size. The exact distribution of patients across centers would generally be unknown at this point of course, but there are various ways in which one might determine a reasonable value. For example, in a trial with a large number of centers, if one can anticipate likely 'smallest' and 'largest' center sample sizes (say, a and b), and if it seems reasonable to expect that center sizes should be approximately evenly spaced between those two values (for example, 10, 12, 14, ..., 48, 50), it can be shown that I_U is well approximated by the following, which depends only on the ratio $r=b/a$:

$$(1/2) \log r ((r+1)/(r-1)).$$

(This is obtained by noting the similarity between the form of I_U and the integral of x^{-1} .) Minor deviations from equal spacing of center sizes will not strongly affect this type of calculation. Thus, for example, in designing a trial with many centers, with sizes anticipated to be evenly spread between about six and 30, the conventional sample size should be inflated by about 21% in order to achieve the

desired power. Mid-study sample size modification based on observing an unexpected level of imbalance is another possibility for addressing this issue; this probably would not be viewed as compromising the trial, since the only information required would be center allocation and no unblinding of interim data would be involved.

5. Construction of composite centers

In performing unweighted analyses, a practice of defining artificial "pooled" or "composite" centers is often employed; that is, data from different centers are treated in the analysis as if they came from the same center. A number of small centers may be combined, or one or more small centers may be combined with a larger center. This practice attempts to minimize the large variance inflation and data instability of unweighted analyses when there are very small centers. Composites may be constructed to the extent of eliminating empty cells to ensure that treatment effects are estimable in models containing interaction terms. More commonly, this is done to achieve some minimum cell size felt to appropriately limit the influence of individual observations; values around 5 are often chosen.

A small amount of such construction probably should be expected to have minimal impact on analysis results. Nevertheless, in deciding whether, or to what extent, to use composites, the potential implications should be understood. One immediate concern is that, to the extent any center main effects exist, the error variance will tend to be over-estimated. In implementation of any composite strategy, this effect can be investigated in a reasonable manner by comparing the error mean square from an unaltered analysis to that from one in which the composite centers are used. Any increase can be used, similarly as in the previous section, to quantify loss of efficiency. For example, a 10% inflation in error would correspond to a reduction in power from 80% to 76%. While often this effect may not itself be serious, this may further reduce efficiency already impaired by performance of an unweighted analysis.

Questions of bias introduced by this practice are more subtle and are difficult to summarize briefly, though there would seem to be some potential for bias. This practice certainly changes the weights given to centers in unweighted analyses; inclusion in a composite decreases the weight of a center and its patients. Advocates of an unweighted approach on the philosophical basis that centers must be treated equally must also realize that this can no longer be claimed when composites are utilized. Unlike weighted analyses, it may not even be possible to claim that centers of similar sizes are weighted comparably; for example, one center may be right at the maximum size limit for inclusion in a composite and thus substantially downweighted, while a center with a single additional patient might stand alone and receive full weight.

6. Explanatory value of the interaction term

Previous sections have focused on aspects of unweighted analyses which result in loss of efficiency. Since a straightforward manner in which to perform a weighted analysis is to omit treatment-by-center interaction from the model, it is

worth considering whether this practice might also entail a loss of efficiency, by inflation of the error variance; indeed, proponents of unweighted analysis sometimes cite the negative impact of excluding an interaction with legitimate explanatory value. (Of course, use of Type II sums of squares is a mechanism for performing a weighted analysis while still allowing the interaction to explain variation.)

In a model which includes treatment-by-center interaction, let v_{int} and v_e represent the degrees of freedom associated with that interaction and with error, respectively; and let F denote the F statistic for the significance test of the interaction, that is, the ratio of the interaction mean square to the MSE. The MSE in a reduced model which does not contain the interaction term differs from that in the full model by the following multiplicative factor:

$$M = (v_e + v_{\text{int}}F) / (v_e + v_{\text{int}}).$$

If the F statistic were close to 1, as would be expected if there were no real interaction, then removal of the interaction should have minimal effect on inferences, any slight changes being mainly due to increased error degrees of freedom. For larger values of F , exclusion of the interaction term will increase MSE. (To view this in context, consider that an F ratio of about 2 will roughly correspond to a borderline statistical significance in many clinical trial designs.) Unless the model is quite saturated with additional factors, though, it should be the case for most trials that v_e is much larger than v_{int} . (If not, then perhaps there are so many small centers that it was not advisable to have included the interaction term in the first place.) Barring extreme significance of the interaction, it will generally be the case that M does not greatly exceed 1. In particular cases, this effect can be evaluated directly, and its impact addressed straightforwardly.

This concern is no more or less inherently relevant for the interaction than for any other factor, though a past history of routine inclusion may tend to focus it more on this term. A logical approach would seem to be to select model factors based on expected impact on error, and to apply the same standard to the interaction as to other candidates.

7. Comparing weighted and unweighted estimators

A practice sometimes employed when using unweighted inferences as primary analyses involves performing a supplemental weighted analysis as a sort of consistency or robustness check. If center sizes are very different, though, it should *not* necessarily be expected that the two types of analyses will routinely yield similar inferences when a real effect exists. This could of course occur, but might for example require that the smaller variation of the weighted estimator be offset by a larger estimated treatment effect reflected in the unweighted estimator (perhaps due to a larger observed effect in smaller centers).

When center sizes differ, the estimates themselves, rather than the p -values, provide a more natural basis for comparing the two approaches. If *estimates* (not p -values) from weighted and unweighted analyses are similar, it would seem reasonable to conclude that the analyses are basing inference on essentially the same parameter, and that the weighted analysis results are preferable based upon their greater efficiency. If

the estimates seem very different, this might be due to an easily identifiable aspect of the design and data: perhaps a very large center with somewhat atypical results is dominating a weighted analysis, or perhaps there are patients in very small centers with somewhat extreme response values given strong influence in an unweighted analysis. In the absence of such circumstances, then perhaps, as discussed in Section 3, there is something systematic about the relationship of treatment effect to center size which should be investigated as part of a thorough examination of the data. In this sense, a comparison of weighted and unweighted estimators can serve as a useful diagnostic tool.

This then raises the question of exactly how one might address whether or not the estimates are "similar." This probably should incorporate judgment regarding the sizes of differences which have clinical relevance, but can also be addressed in a formal statistical manner. Since both weighted and unweighted estimators are linear contrasts involving model terms, their difference is as well. In our simple model, for example, it is easily shown that the covariance of d_W and d_U is equal to the variance of d_W , from which it follows that

$$\text{Var}(d_U - d_W) = V_U^2 - V_W^2.$$

A significance test for the difference between D_U and D_W is essentially just a test of a single degree of freedom contrast involving least squares interaction parameter estimates. (In a two-center, two-treatment design, it is equivalent to the interaction test.) It might thus be sensible to consider performing such a test only if the interaction is significant (being cautious against over-interpretation of a negative outcome due to limited power), to see if there is evidence that the interaction is of a type for which the estimators reflect different quantities.

8. Example

The following example illustrates these issues as applied to an actual clinical trial data set. The trial was run in 25 centers, with a total of 522 patients assigned to one of four treatment groups (three doses — high, mid, low — of a test substance, plus a placebo group). Center sizes ranged between 8 and 30; the largest treatment-center combination size was 8 and the smallest was 1. The response variable was continuous, and the protocol-specified analysis model included treatment, center, and their interaction, plus an additional variable which turned out to be very insignificant with minimal impact on analyses, and which for simplicity we will not use here. In anticipation that there would be small centers, an algorithm for constructing composites was pre-specified: centers were combined within geographic regions to ensure treatment-center sizes of at least 5. (Given the nature of the indication, it was considered feasible that response might be related to geographic location.) Furthermore, the pooling algorithm was defined to prevent small cells in some secondary analyses as well as the primary one simultaneously. Thus, more centers were included in composites than would have been the case had the algorithm only involved the primary analysis; also, some centers appeared in composites despite having more patients with primary endpoint data than some "stand-alone" centers, because they had fewer patients for some secondary analyses. Fifteen centers were grouped into three composites; there were 10 "stand-alone" centers, and thus 13 nominal centers to be used in the protocol-specified analysis.

For simplicity, we will focus on the comparison between the placebo and mid-dose groups; the issues were much the same in the analysis of the full data set. Center sample sizes, treatment effect estimates, and the compositing strategy are illustrated in Table 2. Table 3 contains ANOVA output for an analysis using the actual center definitions. Treatment difference estimates and associated inferences were as follows:

| | Estimate | Std. err. | P-value |
|-------|----------|-----------|---------|
| d_U | 1.116 | 0.731 | 0.128 |
| d_W | 1.556 | 0.664 | 0.020 |

Due to the center size imbalance, the inefficiency I_U of the unweighted estimate, as defined in Section 4, was 1.248; I_W was 1.016. This is directly reflected in the 10% larger standard error for d_U . Note that although there seems to be no evidence of interaction, the large efficiency penalty for the unweighted analysis due to its inclusion in the model has already been paid.

The weighted treatment effect estimate is about 40% larger than the unweighted. Reasons for this are suggested in Figure 2, which presents the within-center estimates d_j (the differences of group means) plotted versus the weights given them in construction of the weighted estimator (closely related to their sample size, as described in Section 3): centers with the greatest observed effects are seen to be among the largest centers, and some centers with small effects are among the smallest. In fact, much of the difference in the estimates is due to the outcome in the center denoted in Table 2 as "X"

Table 2. Treatment-center outcomes.

| Pooling | Center | n_{1j} | y_{1j} | n_{2j} | y_{2j} | d_j |
|-------------|--------|----------|----------|----------|----------|-------|
| Unpooled | A | 6 | 9.17 | 6 | 6.77 | 2.40 |
| | B | 6 | 8.54 | 6 | 10.35 | -1.81 |
| | C | 7 | 11.95 | 7 | 10.26 | 1.69 |
| | D | 6 | 10.40 | 6 | 8.98 | 1.42 |
| | E | 7 | 9.62 | 7 | 6.24 | 3.37 |
| | F | 6 | 8.45 | 6 | 8.80 | -0.36 |
| | G | 6 | 9.69 | 6 | 7.25 | 2.44 |
| | H | 6 | 9.07 | 7 | 6.28 | 2.80 |
| | I | 5 | 5.41 | 6 | 9.19 | -3.78 |
| | J | 6 | 12.67 | 6 | 10.22 | 2.44 |
| Composite 1 | K | 6 | 10.57 | 7 | 3.80 | 6.77 |
| | L | 4 | 4.91 | 2 | 5.81 | -0.90 |
| | M | 5 | 10.20 | 6 | 12.58 | -2.38 |
| | N | 5 | 7.33 | 4 | 5.82 | 1.51 |
| | O | 3 | 16.97 | 2 | 11.70 | 5.27 |
| Composite 2 | P | 3 | 7.29 | 2 | 12.58 | -5.28 |
| | Q | 6 | 12.67 | 6 | 6.26 | 6.41 |
| | R | 4 | 13.38 | 4 | 11.54 | 1.84 |
| | S | 6 | 12.19 | 6 | 10.60 | 1.59 |
| | T | 8 | 11.32 | 7 | 10.96 | 0.36 |
| | U | 3 | 6.88 | 3 | 7.71 | -0.83 |
| Composite 3 | V | 7 | 10.99 | 7 | 7.34 | 3.65 |
| | W | 5 | 13.85 | 7 | 10.88 | 2.97 |
| | X | 4 | 10.88 | 1 | 15.99 | -5.11 |
| | Y | 2 | 11.41 | 2 | 9.97 | 1.44 |

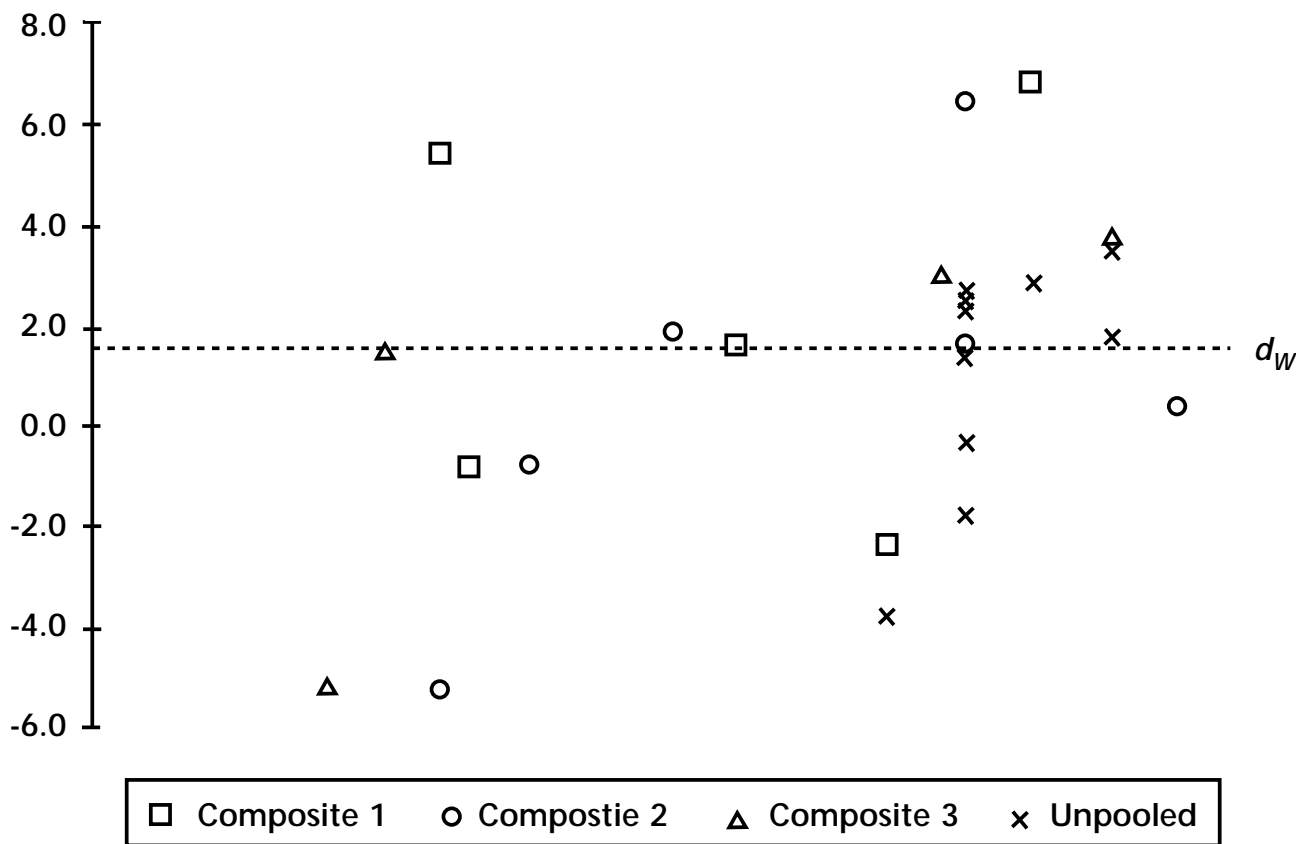


Figure 2. Within-center estimate d_j vs center weight.

Table 3. ANOVA results for mid-dose vs placebo comparison, actual centers.

| Source | d.f. | Mean square | F | P |
|---------------------|------|-------------|------|-------|
| <i>Unweighted</i> | | | | |
| Treatment | 1 | 66.13 | 2.33 | 0.128 |
| Center | 24 | 41.12 | 1.45 | 0.088 |
| <i>Weighted</i> | | | | |
| Treatment | 1 | 155.74 | 5.49 | 0.020 |
| Center | 24 | 42.45 | 1.50 | 0.071 |
| Treatment-by-center | 24 | 20.76 | 0.73 | 0.816 |
| Error | 211 | 28.38 | – | – |

(corresponding to the lower-leftmost point in Figure 2); this center provides the single most highly variable within-center estimate, as there is only a single patient in the mid-dose group. Though the placebo group mean in that center is quite typical of the other centers, and the single value in the mid-dose group is not very extreme compared to the distribution of values throughout the study, this center (in particular, its single mid-dose patient) nevertheless had what would seem to be undue influence on the unweighted analysis results: its exclusion would increase the unweighted estimator by about 25%, explaining much of the difference between the weighted and unweighted results.

Table 4 presents ANOVA output based on use of the composite centers. The weighted results are nearly identical to those based on the actual centers, as expected, since pooling has minimal effect on weighted inference. Decreasing the influence of very small centers on unweighted results is what the use of composites is supposed to accomplish, so one might expect that using the pooled centers would yield unweighted inferences more similar to the weighted results. This is not the case, however: the unweighted estimate based on the pooled centers is only slightly larger than when using the actual centers, and its standard error has actually increased:

| | Estimate | Std. err. | P-value |
|-------|----------|-----------|---------|
| d_U | 1.251 | 0.786 | 0.113 |
| d_W | 1.509 | 0.674 | 0.026 |

Compositing has created a different type of imbalance, not because there are small centers, but because now there are large ones. The quantity I_U is now 1.360, explaining the large standard error of the unweighted estimate; I_W is 1.001. Examining Table 2 and Figure 2 helps explain why the unweighted estimate is smaller than might be expected. Several larger centers have been included in a composite and thus substantially downweighted. The 15 largest actual centers contributed between 12 and 15 patients; among these, 4 of the 5 with the largest observed effects are included in a composite, while only 2 of the other 10 are in a composite. The two centers with the largest estimated effects are easily identified in Figure 2: Center K, with 13 patients in a composite center of 44 patients, whose weight is essentially

reduced by about 70%; and Center Q, with 13 patients in a composite of 58, with weight thus reduced by about 80%. Performing the analysis with either of these two centers "standing alone" (i.e., removed from the composite, in which realistically they didn't need to be anyway) gives a treatment effect estimate nearly identical to that of the unweighted analysis.

The compositing algorithm, though sensibly defined, seems to have played out in a manner which very artificially has tended to obscure an apparent effect, specifically by decreasing the influence of much of the data most strongly indicating that effect. This example has also tended to point out the stability of weighted inference relative to unweighted, i.e., it is much less sensitive, for example, to issues of center size distribution, center definition, extreme data values, model specification, etc.

Table 4. ANOVA results for mid-dose vs placebo comparison, composite centers.

| Source | d.f. | Mean square | F | P |
|---------------------|------|-------------|------|-------|
| <i>Unweighted</i> | | | | |
| Treatment | 1 | 75.10 | 2.54 | 0.113 |
| Center | 12 | 33.30 | 1.12 | 0.341 |
| <i>Weighted</i> | | | | |
| Treatment | 1 | 148.39 | 5.01 | 0.026 |
| Center | 12 | 32.64 | 1.10 | 0.359 |
| Treatment-by-center | 12 | 12.65 | 0.43 | 0.952 |
| Error | 235 | 29.62 | – | – |

9. Discussion/Recent guidelines

The results presented here tend to reflect favorably on weighted analyses; this is generally consistent with the conclusions of several other authors cited above. It should be emphasized that a strategy in which primary analyses give comparable weight to patient responses more than to center estimates, as can be implemented by omitting treatment-by-center interaction from the analysis model, does not necessarily imply that the interaction be ignored. Rather, this approach may be based on the realization that weighted inference will frequently be more likely to provide an accurate answer to the question of whether a treatment effect exists, even if interaction is present, due to its added efficiency and stability, as well as to the general nature and magnitude of many interaction effects. When a weighted analysis suggests that a treatment effect exists, the possible existence and implications of interaction (with center or perhaps with other variables defining important subgroups) can be addressed in subsequent analyses. One has full ability to try to explain how the nature of any interaction (including one which would make differently-weighted estimators substantially different) affects interpretation of the main-effects treatment results.

This issue is not merely one of reducing Type II error and increasing the chance of finding a difference by controlling

the test-statistic variance. We have seen several instances in which unweighted analyses seem to have falsely indicated a treatment effect which did not hold up under a more thorough examination of the data; in these cases, an extremely small number of responses in small centers had large and seemingly undue positive influence on treatment estimates which were not sufficiently offset by the unweighted framework variance inflation.

Even if one strongly felt in a particular situation that the unweighted treatment effect parameter were the correct one on which to base valid inference, it may well be the case that a weighted estimator better estimates this parameter in a squared error loss sense, since the smaller variance may outweigh any perceived "bias." (The pooled-center analysis parameter estimates shown in Section 8 suggest that this phenomenon applies there, for example.)

A strategy in which primary analysis models do not include treatment-by-center interaction, and in which possible interaction is subsequently investigated when an overall treatment effect is indicated, seems much in line with the recently-issued ICH Guideline E9 (1998). Section 3.2, *Multicentre Clinical Trials*, contains the following: "The main treatment effect may be investigated first using a model that allows for center differences, but does not include a term for treatment-by-center interaction ... If positive treatment effects are found in a trial with appreciable numbers of subjects per center, there should generally be an exploration of the heterogeneity of treatment effects across centers, as this may affect the generalizability of the conclusions." Section 5.7, *Subgroups, Interactions, and Covariates*, states: "The primary variable(s) is often systematically related to other influences apart from treatment. ... such as ... the different centers of a multicenter trial. In some instances, an adjustment for ... subgroup effects is an integral part of the planned analysis and hence should be set out in the protocol.... When

the potential value of an adjustment is in doubt, it is often advisable to nominate the unadjusted analysis as the one for primary attention, the adjusted analysis being supportive In most cases, however, subgroup or interaction analyses are exploratory and should clearly be identified as such; they should explore the uniformity of any treatment effects found overall."

References

- Gail, M. and Simon, R. (1985). "Testing for Qualitative Interactions between Treatment Effects and Patient Subsets." *Biometrics* 41, 361-372.
- International Conference on Harmonisation (1998). Guideline E-9: *Statistical Principles for Clinical Trials*.
- Jones, B.; Teather, D.; Wang, J.; Lewis, J. A. (1998). "A Comparison of Various Estimators of a Treatment Difference for a Multi-Centre Clinical Trial." *Statistics in Medicine* 17, 1767-1777.
- Källén, A. (1997). "Treatment-By-Center Interaction: What Is the Issue?" *Drug Information Journal* 31, 927-936.
- Lewis, J. A. (1995). "Statistical Issues in the Regulation of Medicines." *Statistics in Medicine* 14, 127-136.
- Lin, Z. (1999, to appear). "An Issue of Statistical Analysis in Controlled Multicenter Studies: How Shall We Weight the Centers?" *Statistics in Medicine* 18.
- Senn, S. (1997). *Statistical Issues in Drug Development*. Chichester: Wiley.
- Snapinn, S. (1998). "Interpreting Interaction: The Classical Approach." *Drug Information Journal* 32, 433-438.

Biopharmaceutical Section Activities in Dallas

Ken Koury

The recent Joint Statistical Meetings in Dallas provided a showcase for Biopharmaceutical Section activities. The technical program, including invited sessions, special contributed sessions and regular contributed paper sessions, was strong and well-attended, as were the luncheon roundtables and Section-sponsored short courses. Thanks to Tom Capizzi, our 1998 Program Chair, for doing an excellent job in developing and coordinating the various aspects of the overall program.

The Section mixer and business meeting was also highly successful. For the second year in a row, over one-hundred members enjoyed the refreshments and took the opportunity to socialize with colleagues, to listen to highlights of Section activities, and to congratulate our Best Paper and Student Competition Award winners, as well as our newly elected officers and ASA Fellows (see related articles).

Our newly elected officers, who will begin their terms in 1999, are Tom Capizzi (Chair-Elect), Sally Greenberg

(Secretary-Treasurer), Curt Wiltse (Program Chair-Elect), and Nancy Smith (Council of Sections Representative).

Thanks to Sandy Heft, the results of the 1998 Best Paper Competition have already been tabulated, and our winners are:

- **First Place:** Catherine Tangen, University of Washington, *Complementary Nonparametric Covariance Methods to Proportional Hazards Regression in a Randomized Clinical Trial*. Co-authored by Gary Koch, University of North Carolina.
- **Second Place:** Devan Mehrotra, Merck Research Laboratories, *A Minimum Risk Strategy for Comparing Treatment Proportions in Stratified Trials*. Co-authored by Radha Railkar, Temple University.
- **Third Place:** Kristen Meier, U.S. Food and Drug Administration, *Study Design Requirements for Evaluating the Performance of a Diagnostic Test*.

Congratulations to the winners who will receive their awards at our business meeting at the 1999 JSM — see you all in Baltimore! (It looks like the only way we can beat this record-setting time to determine the award winners is to set up exit polls at our sessions next year. For some reason, though, I don't think that we'll be able to get ABC and CBS into a bidding war.)

Call for Special Contributed Sessions and Invited Posters at the 1999 JSM

Christy Chuang-Stein

Special Contributed Sessions

Have you ever wondered how you can get experts on a particular subject together in one room for a good discussion? Does it worry you that despite all the committee efforts, your contributed paper might not be appropriately grouped at the JSM? Have you ever wondered how to organize a Special Contributed Session at the JSM? If your answer to any of the above questions is yes, I encourage you to take the initiative and organize a Special Contributed Session on a theme that is of interest to you.

A Special Contributed Session consists of five contributed papers, each 20 minutes long, and 10 minutes at the end of the session for floor discussion and concluding remarks by the Session Chair. While not strictly forbidden, the use of discussants is strongly discouraged for Special Contributed Sessions. By comparison, a Regular Contributed Session consists of seven papers. So, speakers in a Special Contributed Session get a better deal in terms of presentation time. Another advantage of Special Contributed Sessions is that there is no limit on the number of Special Contributed Sessions that a section such as the Biopharmaceutical Section can sponsor. Also, frequently a Special Contributed Session attracts a larger audience than a Regular Contributed Session because of the theme-oriented presentations.

There is no trick to organizing a Special Contributed Session at the JSM. Your responsibilities as the organizer are to decide on the theme and arrange all the presentations centered around the theme. So, if you are thinking about presenting a contributed paper at the 1999 JSM and can convince other statisticians with similar research/development interests to do the same, please accept the challenge to put a Special Contributed Session together. (However, remember you don't have to be a presenter to organize a Special Contributed Session.) You can start by networking with your friends and colleagues on this possibility right now. It is really easier than you think and can be a lot of fun.

The deadline for the Biopharmaceutical Section to submit Special Contributed Sessions to the 1999 JSM Program Committee is February 1, 1999. This is to allow the Program Committee to identify the abstracts that belong to each Special Contributed Session and to organize them into the proper session. Please let me know of your plan as early as possible so I can coordinate with the Program Committee to keep track of the abstracts as they arrive at the ASA. As for submitting the abstracts, you (if you are a presenter yourself) and your speakers must send an abstract and pre-registration material to the ASA office by the February 1 deadline. Please instruct your speakers to indicate on the abstract form that they are speaking in a Special Contributed Session and to give your name as the session organizer.

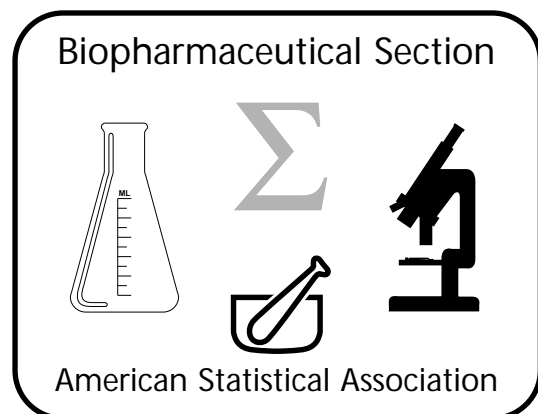
Invited Posters

I know that the previous Joint Statistical Meetings have not been very successful with poster presentations. Unlike other professional societies, we (statisticians) have a tendency to underestimate the power of poster presentations. For 1999, ASA is hoping to organize noticeable, big-fuss, lots of publicity Invited Poster Sessions at the JSM. An Invited Poster Session lasts 120 minutes and consists of 10-20 invited poster participants. The invitation of poster participants is the responsibility of the JSM Program Chair, in consultation with the Program Committee. The 1999 JSM Program Chair has sent out an invitation for invited poster proposals. Invited posters will be selected from the proposals submitted. There are several advantages for presenting an invited poster at the JSM. They include:

- ASA implemented the one person-one talk policy starting with the 1998 JSM. Invited posters, however, do not count against a speaker's allocation. In other words, people presenting an invited poster can still present an invited or a contributed talk or serve as a discussant.
- Invited poster presenters will have access to just about any type of equipment that ASA can arrange.
- The plan is to hold the invited poster sessions late in the afternoon in a nice place, with a cash bar handy, so that attendees can socialize while studying the posters.
- Because of the length of the poster sessions, topics can be discussed extensively and one-on-one contact with the audience is the norm.
- If you are not yet ready for public speech, invited posters can be a good way to prepare you for the ultimate public speech goal. It is a great professional development activity and creates an opportunity for you to meet with individuals with similar research interests.

If you are interesting in organizing a Special Contributed Session and/or submitting an invited poster proposal, please let me know. I would be happy to discuss your ideas with you. I can be reached at christy.j.chuang-stein@am.pnu.com or by phone at (616) 833-0209.

Please let me hear from you.



Deming Conference on Applied Statistics Has a Lot to Offer

David Stock

Each year the Deming Conference brings together outstanding speakers from academia and industry to teach applied statistics in a tutorial format. The speakers are experts in their areas and have extensive experience both using the methods and teaching them to others. This year the conference will be held from December 7 to 11 at the Resorts Casino Hotel in Atlantic City, New Jersey. Complete conference information and registration materials are available by contacting the conference organizers (see below) or through the Web at: nimbus.temple.edu/~kghosh/deming98/.

This year's conference features a dozen three-hour tutorials and a pair of two-day short courses. The tutorials are offered during the first three days of the conference (December 7-9) and cover topics in applied statistics, biostatistics, and clinical trials. Conference participants may register for all three days or any single day. The short courses follow the tutorials. One course is on multiple comparisons, and the other is on data mining. Registration for the short courses is separate from the tutorials.

The tutorial sessions offer solid introductions to areas important to applied statisticians. This year's tutorials are:

- *An Introduction to the Analysis of Repeated Measures Experiments Using Mixed Models*, by Dallas Johnson (Kansas State University);
- *Ideas for Studying Regressions Through Graphics*, by R. Dennis Cook (University of Minnesota);
- *Statistical Issues in Pharmacoeconomic Studies*, by Rukmini Rajagopalan (Glaxo Wellcome Inc.);
- *Bayesian (and Conditional Frequentist) Hypothesis Testing and Model Selection*, by James Berger (Duke University);
- *Applied Regression Analysis, 3rd Edition What's In, What's Out, and Why*, by Norman R. Draper (University of Wisconsin);
- *Design and Analysis of Equivalence Trials*, by Kalyan Ghosh and Ivan S.F. Chan (Merck & Co., Inc.);
- *Co-ordinate Free Log-Linear Models*, by Daniel Zelterman (Yale University);
- *Handling Small Sample Problems in Biostatistics: Applications to Survival Analysis*, by Robert L. Strawderman (University of Michigan) and Martin T. Wells (Cornell University);
- *Connecting Clinical Trial Design to Analysis*, by Vance W. Berger (CBER, FDA);
- *Applying Tree-Based Regression Methods to the Analysis of Designed Experiments*, by Bert Gunter (Merck & Co., Inc.);
- *Statistics in Medical Research: Developments in Clinical Trials*, by Edmund A. Gehan (Georgetown University Medical Center);
- *From Graph Design Principles to Graphical Templates That Facilitate Understanding*, by Daniel B. Carr (George Mason University).

The short-courses feature clear, accessible, expositions of cutting edge methodology that can be used to deal with issues of growing concern to statisticians. These courses assume no prior expertise and contain numerous case studies showing how to apply the methods to practical problems. The first

course, *Multiple Comparisons: Applications and Case Studies*, is being given by Peter Westfall (Texas Tech University) and Dror Rom (Prosoft Software, Inc.) The second course, *Data Mining and Knowledge Discovery: A Practical Introduction*, is being given by Richard DeVeaux (Williams College) and Lyle Ungar (University of Pennsylvania).

Books related to the topics covered at the conference are available to anyone at a significant discount. An order form can be obtained at the conference Web site: nimbus.temple.edu/~kghosh/deming98/.

If you would like further information about the Deming Conference but do not have access to the Web, please contact: Walter R. Young, Wyeth-Ayerst Research, PO Box 42528, Philadelphia, PA 19101, E-mail: youngw@war.wyeth.com, Fax: (610) 989-4553 or David Stock, Bristol-Myers Squibb, Department 716, 5 Research Parkway, Wallingford, CT 06438, (203) 677-3553.

Corporate Members

Philip J. Pichotta

Membership Chair

The Biopharmaceutical Section has three types of memberships: regular, student, and corporate. The corporate members of the Section, who must first be 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:

- **Merck & Co., Inc., West Point, PA**
- **Pfizer, Inc., Groton, CT**
- **Scirex Corporation, Bloomingdale, IL**
- **Trilogy Consulting Corporation, Waukegan, IL**
- **Warner-Lambert Company, Ann Arbor, MI**

The Biopharmaceutical Section gratefully acknowledges its corporate members and their support. The number of corporate members has decreased from 12 to five. We are looking for ways to increase the number of corporate members and to provide better services to them. Currently, we list our corporate members periodically and that is about all they get for their membership. The dues do help support activities that will benefit members in their organizations. If you have any suggestions on how to enhance the services to our corporate members, please contact me by phone, (847) 937-3708, or E-mail, philip.pichotta@abbott.com.

Section News

Congratulations 1998 ASA Fellows!

Congratulations to the members of the Biopharmaceutical Section who were elected as Fellows in 1998! The new fellows and their citations are:

Gregory Campbell, Director, Division of Biostatistics, Center for Devices and Radiologic Health, Food and Drug Administration: For contributions to statistical methodology for evaluation of medical diagnostic tests and image analysis; for leadership in statistical approaches to the evaluation of medical devices.

Christy Chuang-Stein, Director of Biostatistics, Pharmacia and Upjohn: For sustained excellence in applications of statistics to medical research, particularly in the evaluation of pharmaceutical safety issues; for methodological research, service to the profession, and for editorial service.

Marie Davidian, Associate Professor, Department of Statistics, North Carolina State University: For contributions to the theory and practice of nonlinear mixed effects models, especially in pharmacokinetics; for service to the profession.

Robert L. Davis, Executive Director, Astra Pharmaceuticals: For significant contributions in the development of pharmaceutical therapies, outstanding management and development of professional staff, furthering the understanding and appreciation of statistics in the community, and service to the profession.

Stephen W. Looney, Biostatistician and Professor of Family and Community Science, University of Louisville Medical School: For excellence in cooperative statistical research in the social and health sciences, for innovative teaching and stewardship, and for dedicated service to the American Statistical Association.

George W. Williams, Vice President, Biostatistics and Research Data Systems, Merck Research Laboratories: For leadership and administration of major biostatistical programs, for excellence as a biostatistical collaborator, for excellence in the application of statistical methods to clinical trials, and for service to the profession.

Lianng Yuh, Director of Statistics, Department of Clinical Research, Central Research Division, Pfizer, Inc.: For outstanding leadership of a major biopharmaceutical group; for important contributions to biopharmaceutical statistics; for sustained encouragement of interaction between industry and academia; and for leadership and services to the profession and the pharmaceutical industry.

Please note that the identification of Fellows as members of the Biopharmaceutical Section was based on the membership list available on the Section's Web site. We know that there are errors in the list; so if we have omitted any Fellows who are members of the Section, please let Anne Meibohm know.

Additional 1997 Fellows

Two additional Fellows should have been identified as members of the Biopharmaceutical Section in the article on

1997 ASA Fellows in the Spring 1998 issue. Our apologies, Dan and Richard.

Daniel F. Heitjan, Associate Professor of Biostatistics, Columbia University: For important research on statistical methods for missing and coarse data; for excellent collaboration and teaching in medicine; and for editorial service in statistics and medicine.

Richard L. Tweedie, Professor of Statistics, Colorado State University: For outstanding leadership in statistical centers across two continents; for important contributions to the theory of Markov processes; and for the successful application of probability and statistics to a diverse range of practical problems.

Nominations for Fellows

If you know of any members of the Biopharmaceutical Section who you think deserve to be a nominated to fellowship, please nominate them or pass along their name to Larry Gould, the chair of the Section's Fellows Committee (*goulda@merck.com*). The members of the committee will be glad to provide help and answer questions about the process. A package to guide potential nominators as to what is needed for a good nomination package will be distributed in September.

Best Student Paper Award Winners

Ji Zhang

Five students were presented the 1998 Best Student Paper Award at the annual Biopharmaceutical Section Business meeting in Dallas during the JSM. The five winners are:

Charles Tan (Temple University)—*MLE and M-Estimator Based Approaches for Linear Statistical Relationship and Their Applications in Assay Methods Comparison*,

Catherine Tangen (University of Washington)—*Complementary Nonparametric Covariance Methods to Proportional Hazards Regression in a Randomized Clinical Trial*,

Brian Wiens (Temple University)—*Testing Similarity of Three Binomial Proportions*,

Kewei Ming (University of Pennsylvania)—*Substantial Gains in Bias Reduction from Matching with a Variable Number of Controls*,

Anna Legedza (Harvard School of Public Health)—*Prior Elicitation and Computation in Phase I Clinical Trials Using the Continual Reassessment Method*.

The award consists of a certificate and a cash award of \$1,000 sponsored by the Biopharmaceutical Section. This competition, in its fifth year, seeks to encourage the study of statistics and its practice in the biopharmaceutical industry and to increase student participation in the Section's programs and activities at the annual JSM. Five awards have been presented each of the last three years. The student paper reviewers for this year were Naitee Ting, Demissie Alemayehu, Avital Cnaan, and Ji Zhang.

Students are strongly encouraged to participate in future competitions. For more information please contact the Biopharmaceutical Section through www.amstat.org or directly at www.ccnnet.com/~sallyg/asa/biopharm.htm.

1997 Treasurer's Report

Jeff B. Meeker

Secretary/Treasurer

We have now completed a two-year effort to reduce our cash-on-hand, significantly, through a reduction in prices for Section events and services, additional activities, and some one-time expenses thought to be beneficial to the Section. This was felt the fairest way to return our excess cash to Section members. During 1997, we showed a loss of \$6874, which brought us to our end-of-year position of \$49,160, close to our original goal.

Summary of 1997 Revenue and Expenses

| | Income | Expenses |
|----------------------|-----------------|-----------------|
| General | | |
| Dues | \$10,114 | |
| Continuing Education | \$2,127 | |
| Food Functions | | \$2,766 |
| Awards | | \$6,000 |
| Other | \$2,111 | \$4,650 |
| Total | \$14,352 | \$13,416 |
| Newsletter | | \$6,761 |
| Proceedings | \$5,040 | \$6,089 |
| Conference | \$24,550 | \$24,550 |
| Total | \$43,942 | \$50,816 |

Two activities which were initiated to reduce costs are to be continued. The first is the social hour associated with our annual business meeting at JSM. It has been extremely successful in bringing Biopharmaceutical Section members together in a relaxed atmosphere and increasing attendance at our meeting. The second is the workshop in the fall. After two successful workshops, a third is planned for this fall.

The 1998 budget has been finalized and submitted to ASA and is summarized below:

| | Income | Expenses |
|--------------|-----------------|-----------------|
| General | \$14,125 | \$13,235 |
| Newsletter | | \$8,500 |
| Proceedings | \$6,500 | \$5,865 |
| Conference | \$25,000 | \$20,310 |
| Total | \$46,875 | \$49,160 |

We are expecting a slight loss for 1998.

The possibility of an increase in dues, not to the original \$11, but possibly to \$6 or \$7 is likely. The Executive Committee considered one in 1997 and decided not to raise the dues for 1998. However, a dues increase is on the agenda for the August, 1998 meeting of the Executive Committee. Each dollar in dues above the initial dollar that goes to ASA adds approximately \$2000 income. An increase in this range may be needed to cover expenses and is in line with that charged by other sections. The Biopharmaceutical Section, however, is one of the most active, has an annual budget approximately twice that of any other section, and, in terms of services provided to its members, would still be the best bargain in section membership.

Minutes of ASA Biopharmaceutical Section

Executive Committee Meeting

March 31, 1998, Pittsburgh, Pennsylvania

Ken Koury introduced Avital Cnaan, who has been appointed a member of the Executive Committee. He also announced Ersen Arseven has been appointed Associate Editor of the *Biopharmaceutical Report*. Ken distributed 1998 letterhead.

Attendees:

| | | |
|-----------------|---------------|---------------|
| Avital Cnaan | Larry Gould | Anne Meibohm |
| Bob Davis | Ralph Harkins | Phil Pichotta |
| Chuck Davis | Sandy Heft | Bob Small |
| Richard Entsuah | Ken Koury | Steve Snapinn |
| Sally Greenberg | Jeff Meeker | Liang Yuh |

Fellows Committee

Larry Gould reported the section submitted three nominees to the ASA Fellows Committee. He intends to institute several improvements, including: 1) systemize procedures with time tables for each activity; 2) provide reminders and program checks for those putting together the nomination packages; 3) prepare a notice of Biopharmaceutical Section recipients for the *Biopharmaceutical Report* after the JSM meeting, asking for additional nominations; and 4) provide follow-up to those making recommendations.

JSM Meeting Minutes

The minutes of the October 29, 1997, meeting held at the Section Workshop in Bethesda, Maryland, were approved.

Treasurer's Report

Jeff Meeker reviewed the final 1997 financial statement from ASA. He also reviewed the proposed 1998 budget.

The Executive Committee discussed the need for an increase in Section dues. A proposal to increase dues by \$2 effective in 1999 was deferred to the August meeting.

The Executive Committee discussed the price of 1998 *Proceedings of the Biopharmaceutical Section*. It was recommended to raise the prepublication price from \$18 to \$25, consistent with the end of our program to reduce cash-on-hand. Post-publication price for both Section members and non-members will remain the same.

1998 ENAR program

The Section sponsored three sessions at the 1998 ENAR meeting:

- *Impact of Recent Therapeutic Advances for AIDS on Clinical Trial Design and Analysis* (Tony Lachenbruch),
- *The Role of DSMBs and Accumulating Evidence in the Conduct of Clinical Trials* (Desmond Thompson),

- *Exploratory Data Analysis Using Classification Trees: Biomedical Applications* (Frank Shen).

The ENAR Program Chair, Rich Chappell, was very easy to work with. He also avoided scheduling sessions sponsored by the Biopharmaceutical Section during our Executive Committee meeting.

1999 ENAR program

ENAR allocated three invited paper sessions to the Biopharmaceutical Section for 1999. The three proposed sessions, with the organizers are:

- *Applications of Computer Intensive Methodology in Drug Development* (Preclinical and Clinical) (Sandy Heft),
- *Statistics in Genomics and Bioinformatics* (Tom Vidmar),
- *The Use of Meta Analysis in Treatment Evaluation and Drug Development* (Sue Marcus).

1998 Joint Statistical Meeting Program

Ken Koury announced the first choice for the Executive Committee meeting is 7:30 A.M.–NOON on Monday, August 10 and the second choice is Tuesday, August 11 at the same time. The first choice for the reception and business meeting is 6:00–7:30 A.M., Tuesday, August 11, and the second choice is Monday at the same time. It was noted that the Executive Committee meeting should precede the Section business meeting.

Tom Capizzi reported we have four invited paper sessions for JSM:

- *Statistical Issues in Vaccine Clinical Trials* (Brian Wiens and Tony Lachenbruch),
- *Evaluation of Statistical Procedures in the U. S. FDA Guidance for Bioequivalence Studies* (J. P. Liu),
- *Design of Phase I Trials* (Bill Rosenberger),
- *Permutation Tests in Clinical Trials* (Vance Berger).

The Section has four special contributed paper sessions:

- *Statistical Issues in the Evaluation of Health Related Quality of Life*,
- *Statistical Issues in Therapeutic and Diagnostic Devices*,
- *Statistical Methods for Incomplete Data I*,
- *Statistical Methods for Incomplete Data II*.

The Section has eight regular contributed paper sessions:

- *Statistical Methods for Analysis and Modeling of Data from Biomedical Experiments* (Stephen Eckert),
- *Statistical Methods for Handling Multiplicity Issues* (Shirley Lan),
- *Experimental Design Strategies for Fixed Duration and Sequential Clinical Trials* (Christine Clark),
- *Statistical Evaluation of Time to Event and Censored Data* (Robert Smith),
- *Statistical Approaches for Improving Design and Interpretation of Biomedical Studies* (William Myers),
- *Some Sophisticated Methods for Fitting Mixed Effects Linear and Nonlinear Models* (Stephen Gulyas),
- *Statistical Issues in Bioassay and Combination Treatment Studies* (Mithat Gonen),
- *Statistical Methods Useful for Modeling Time-Dependent Data* (James Kenyon).

There is a new rule this year that regular contributed paper sessions should have seven papers. Thus, there will be 56 regular contributed papers. There are only four posters.

The number of regular contributed papers and posters is considerably lower this year than in previous years. This appears to reflect the location and a new rule limiting individuals to just one appearance as a presenter or discussant. Regular contributed paper submissions to JSM were down 25–30% from the three previous years and our decrease is consistent with that.

1998 Short Course

The Section will sponsor a one-day short course, *Design and Analysis of Clinical Trials*, by J. P. Liu and Shein-Chung Chow. The course is tentatively scheduled for Wednesday, August 12. We will also cosponsor the one-half day short course, *Bayesian Methods in Clinical Research*, based on the text *Bayesian Biostatistics* by Darlene Stangl. That course is tentatively scheduled for Monday, August 10.

1998 Round Tables

The Section will sponsor six round table discussions. Summaries of 1996 round table discussions appeared in the December, 1997 issue of *Biopharmaceutical Report*. Summaries of 1997 round table discussions are scheduled to appear during 1998.

1998 Best Student Paper Awards

Ji Zhang will chair the 1998 Best Student Paper Awards. The committee also consists of Demissie Alemayehu and Naitee Ting. A notice appeared in the January, 1998, *Amstat News*, and letters were mailed to universities. To date, 27 abstracts have been received. The deadline for submission of the paper is May 1.

Council of Sections

ASA has indicated they will provide up to 1/4 page of section specific material in a mailing to new members.

Assignment: Phil Pichotta will develop the material.

Sections who have outreach activities to the public are requested to contact ASA.

Assignment: Ken Koury will call the individual at ASA responsible for outreach to get more information as to what is requested.

ASA requested a list of section experts that would be available for media contacts.

The Council of Sections meeting is scheduled 1:00–4:00 P.M. on Sunday, August 9, with the debriefing meeting 9:00–10:00 A.M. on Thursday, August 13. There will also be meetings again for various section officers.

Publications Committee

The following articles have appeared in *Amstat News*: October, 1997: 1997 Joint Statistical Meetings Summary (Denise Roe),

December, 1997: FDA/Industry Interaction Workshop Summary (Christy Chuang-Stein),
 January, 1998: 1997 Highlights (Bob Davis) and Student Paper Competition (Denise Roe),
 February, 1998: Invited Paper Proposals for 1997 ENAR Meeting and JSM (Christy Chuang-Stein) and Membership Benefits (Phil Pichotta),
 April, 1998: Biopharmaceutical Report (William Huster).

The proposed topics for the remainder of 1998 are:

May: Web site/listserv (Sally Greenberg),
 June: 1998 Joint Statistical Meetings Overview (Tom Capizzi),
 July: Reminder of 1998 Joint Statistical Meeting Activities (Tom Capizzi) and Announcement of Fall Workshop (Workshop chair),
 August/September: Winners of 1997 Best Contributed Paper Competition and 1998 Best Student Paper Competition (Award Committee Chairs),
 October: Executive and Business Meeting Summary from Joint Statistical Meeting (Jeff Meeker),
 November: 1999 Student Paper Competition Announcement and Procedures (Student Paper Committee Chair),
 December: Summary of Fall Workshop (Workshop Chair).

Assignment: Bob Small will submit the 1998 Workshop to *Amstat News* for inclusion in the list of upcoming meetings.

The ASA policy for *Amstat News* is that we are allocated one page per issue to communicate with our members. The problem occurs on articles related to workshop and conference information. Feature articles about conferences or workshops which provide dates and locations and lists of speakers are fine. However, if the intent is to provide registration forms and schedules, it becomes advertising. Each section is allowed one free ad each year. For 1998 we used our ad to schedule the Midwest Biopharmaceutical Workshop. We therefore have no space to publish the registration form and schedule for the Fall Workshop. ENAR approached us to see if they could use our ad space, but we had already used it.

The 1997 *Proceedings of the Biopharmaceutical Section* are ready to go to press. ASA has recommended the same press run size (600 copies). As indicated earlier, the pre-publication price will increase from \$18 to \$25, but otherwise the price will remain at \$25 for members and \$38 for non-members. We have agreed to invite speakers from the 1998 Midwest Biopharmaceutical Statistics Workshop to publish in the 1998 *Proceedings of the Biopharmaceutical Section*.

Biopharmaceutical Report

Anne Meibohm reported three issues of the *Biopharmaceutical Report* are planned for 1998. Topics include a summary of the 1997 Workshop, a report on 1997 round table discussions from JSM, a report on newly elected ASA fellows, minutes of the Executive Committee meeting and a financial report. Ersen Arseven was appointed associate editor.

Electronic Mail List

Sally Greenberg indicated there has not been a lot of activity lately on the mail list.

The proposal to allow recruiters to post job listings was deferred to the August Executive Committee meeting.

Web Site

Sally Greenberg indicated she is behind but expects to be caught up by the end of April. She also needs a back-up editor.

Membership Committee

Phil Pichotta indicated two new members have been appointed to the Membership Committee, David Carlin for a two-year term and Mark Munsell for a three-year term. This will allow a rotation of members on the Committee. The chair will represent the Committee on the Executive Committee.

No plans have been developed for recruitment of new members. There was a discussion of corporate members, specifically what the Section offers in return for corporate membership.

Midwest Biopharmaceutical Statistics Workshop

The Midwest Biopharmaceutical Statistics Workshop is May 18–20, in Muncie, Indiana. A preliminary program was published in *Amstat News*. There is a Web site for the conference.

Assignment: Sally Greenberg will provide a link from our Web site.

Applied Statistics Meeting.

The Applied Statistics Meeting will be held in Atlantic City, December 7–11. The meeting will consist of tutorials on December 7–9 and short courses on December 10–11.

Assignment: Sally Greenberg will provide a link from our web site to the one for this conference.

PhRMA

Lianng Yuh reported PhRMA Statistics and Data Management Workshop will be November 9–11 at the Hyatt Regency Hotel in Bethesda, Maryland. The theme of the workshop is "Opportunities and Challenges for Industry and FDA." Topics will include Genomics, Single Trial Strategy, and Data Mining. Lianng is overall Program Chair and Ken Koury is the Clinical Program Chair.

1999 Joint Statistics Meeting.

We were tentatively allocated three invited sessions at the 1999 JSM. Proposals include:

- *Progress of Adverse Events Working Group* (Noel Moeberg),
- *Sample Size Estimation—Special Cases and Software Comparisons*,
- *Reflecting Past Statistical Advancements and Envisioning New Statistical Applications in Biopharmaceutical Science*.

Two special contributed paper sessions are being considered:

- *Dose Proportionality and Linear Pharmacokinetics*,
- *Clinical Trials Simulation*.

1997 Joint Biopharmaceutical Section/FDA Workshop

Bob Small reported the FDA did not receive the funding they were expecting for the Workshop. Also, they lost the proposed site.

A program committee has been established, consisting of Bob Davis, Sandy Heft, Ralph Harkins, Len Oppenheimer, Greg Campbell, Nancy Smith, Tony Lachenbruch, plus other members from the FDA. They expect attendance to be 200–250. The price was suggested to be \$125, with \$65 for government employees.

Section Officers

Bob Davis indicated nominees are required for Chair-elect and Program Chair-elect for 2000 and a Council of Chapters Representative for 2000–2002.

Let's Hear from You!

If you have any comments or contributions, contact Editor Anne Meibohm, Merck Research Laboratories, BL3-2, P.O. Box 4, West Point, PA 19486; Phone: (610) 397-2545; Fax: (610) 397-2931; E-mail: anne_meibohm@merck.com; Associate Editor Ersen Arseven, Arseven Consulting, Inc., 247 South Blvd., Nyack, NY 10960; Phone: (914) 358-1348; Fax: (914) 358-8570; E-mail: r7consult@aol.com; or Past Editor 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.

*© 1998 The American Statistical Association
Printed in the United States of America*

Biopharmaceutical Report

c/o American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3415
USA

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

FIRST CLASS POSTAGE