Stratification Methods in Caries Clinical Trials

A. KINGMAN

National Caries Program, National Institute of Dental Research, Bethesda, Maryland 20205

J Dent Res 63(Spec Iss):773-777, May, 1984

Prognostic factors associated with caries incidence can be useful in both the design and analysis of randomized clinical trials. They can be used as a pre-randomization stratification technique, as is done in randomized block designs. Here, participants are first grouped into homogeneous strata or blocks, and then randomly assigned to treatment groups separately within each stratum. Alternatively, prognostic factors can be introduced as post-stratification variables in the statistical analysis.

The question naturally arises whether pre-randomization stratification techniques should be employed. Statisticians are equivocal on this issue. Arguments voiced for stratification[4,5,20] are usually related to the increase in precision or the elimination of imbalance among groups that is achieved by using stratification. This is particularly relevant for trials having relatively few subjects.[11] The arguments voiced against stratification[17-19] typically reflect pragmatic considerations related to the conduct or administrative aspects of the trial. In any event, the ability to implement and conduct a specific clinical trial successfully almost always requires compromises made between that which is theoretically optimal and that which is possible.

In caries clinical trials whose participants are of a restricted age range, pre-randomization stratification methods are often avoided for several reasons: (1) the lack of prognostic variables which are strongly associated with caries incidence, (2) logistical problems related to the school-based format of conducting these trials, (3) anticipated attrition rates of considerable size, and (4) the use of large numbers of participants per treatment group.

Even though there are no prognostic variables that have consistently been shown to have strong associations with caries incidence, several are known to be moderately associated.[6,9,14] Such variables can be useful as post-randomization stratification variables. Analytically, one incorporates such variables as explanatory factors (as in the analysis of variance) or covariates (as in the analysis of covariance) into the statistical model.

The post-stratification methods are used for several reasons. For caries trials, the more important include:

(1) inducing equivalence among treatment groups that have been generated by randomization. This is of particular importance in caries clinical trials because of the relatively large attrition rates experienced among treatment groups during the study period.

(2) achieving more powerful statistical tests or obtaining more precise estimates by reducing the error variance.

(3) assessing the degree to which the treatment effects are uniform for the subpopulations represented by the various strata. This can be useful for identifying the types of subjects who are most likely to benefit from specific preventive procedures.

The effects of imbalance among treatment groups (partial confounding), relative to a prognostic variable, were investigated by Korts[16] for the no-interaction model. He illustrated the wide range of treatment effects one could observe in a caries trial due solely to the degree of imbalance of subjects within groups while holding the strata marginals fixed. He showed that, by ignoring the prognostic factor, percent reductions in caries due to treatment ranging from −35% (*i.e.*, a 35% increase) to 59% could be observed, even though the differences between groups were constant for each stratum of the stratification variable. In contrast, by including this prognostic factor, identical treatment means – and hence, the same reductions (24%) – were obtained for each trial.

The focus in this paper will be on models with interaction in which imbalance among treatment groups is minor. Issues regarding efficiency and the identification of differential treatment effects for various sub-populations will be discussed.

We will assume a two-way fixed effects model having unequal cell frequencies, a fixed set of treatments as one factor (call it T), and a prognostic factor (call it B) as the stratification factor. I shall also assume that all cells in the data array contain at least one observation (no empty cells).

If we let y_{ijk} represent the observation (caries increment) for the kth individual in the ith treatment group and in the jth stratum, then the general two-way cross-classification model can be written as

$$E(y_{ijk}) = \mu_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij}, \qquad \text{Eq. (1)}$$

$$\text{for } i=1,2,...,I; j=1,2,...,J; k=1,2,...,n_{ij};$$

assuming interaction is present, or as

$$E(y_{ijk}) = \mu_{ij} = \mu + \tau_i + \beta_j, \qquad \text{Eq. (2)}$$

$$\text{for } i=1,2,...,I; j=1,2,...,J; k=1,2,...,n_{ij};$$

where

$$\mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0 \text{ for } i \neq i', j \neq j'$$

assuming the effects are additive. Here

μ_{ij} represents the mean (caries increment) for individuals in the ith treatment group and the jth stratum,
μ represents an overall mean (caries increment),
τ_i represents the effect of the ith treatment group,
β_j represents the effect of the jth stratum, and
$(\tau\beta)_{ij}$ represents the non-additive effect (interaction) of the ith treatment group and the jth stratum.

Also n_{ij} represents the number of individuals in the ith treatment group and in the jth stratum. We also assume the usual notation $n_{i.} = \sum_j n_{ij}$, $n_{.j} = \sum_i n_{ij}$ and $n_{..} = \sum_{i,j} n_{ij}$

In the balanced case, differences between the observed marginal means are unbiased estimates of differences between main effects in model (2), whereas, in the unbalanced case, the corresponding differences are biased, that is, they are contaminated by effects of the other factor.

Furthermore, in the balanced case for either model (1) or (2), the sums of squares for T, B and T x B are additive, they are orthogonal, and, therefore, tests regarding one effect are independent of the other effects. In the unbalanced case, these sums of squares are not additive, the effects are not orthogonal, and when tests for significance of one effect are made, the tests must be adjusted for other appropriate effects in the model.

There are two general approaches that can be taken

when analyzing unbalanced data sets.[2,8] The additive sum of squares approach is one for which the individual sums of squares for main effects and interaction are additive to the sum of squares representing variation among the cell means. This method is approximate because the individual sums of squares do not have chi-square distributions, and hence, ratios of mean squares do not have exact F-distributions. The method of unweighted squares of means is an example of this method. This method provides a good approximation to the least-squares method, provided the ratio of largest to smallest cell sample size is not too large.

The other general approach is that of least-squares. In the unbalanced case, one obtains sums of squares which have exact chi-square distributions, the resulting F-tests are exact, but the sums of squares for the main effects and interaction are not additive. Two specific examples of least-squares methods are known as the method of fitting constants and the method of weighted squares of means. The least-squares methods will be discussed in this presentation.

In each situation, one must decide whether the model given by (1) or (2) is appropriate. When prior information regarding the appropriate model is lacking, a test for the presence of interaction should be performed. The method of fitting constants can be used to derive this test. This test is based upon a comparison of the reduction in sum of squares attributable to model (1), with the corresponding sum of squares attributable to model (2). Equivalently, it can be viewed as the additional sum of squares obtained by fitting model (1) after model (2) has been fitted.[10,23] The outcome of this test determines the procedure to use for testing main effects. Generally, if one obtains a $p > 0.05$ ($p > 0.20$ or $p > 0.25$ if overall p-value is considered important), the additive model (2) is assumed, and main effects are tested by the method of fitting constants. Tests for each main effect are performed, eliminating the other. If both tests are significant, the full additive model (2) is adopted, and treatment comparisons, contrasts, and inferences are based upon this model.

On the other hand, if the test for interaction is significant, one must proceed differently. The optimal procedure to follow here is unclear, and arguments continue to be made by statisticians for[7,8,22,25] and against[1,15] performing statistical tests for main effects in the presence of interaction. For the sake of argument, I will take the position that it makes sense to test for main effects in the presence of interaction if one can assume that the treatments are consistently ordered[22] – that is, if differences between each pair of treatments are of the same sign, although not of similar magnitude, for all strata, separately. Of course, one must also assume that the confounding due to unequal cell frequencies is not too severe.

Since the individual population cell means μ_{ij} are estimable (we assumed that each cell had at least one observation), any linear combination of cell means is estimable. Treatment main effects (also called marginal or average effects by other authors) are defined in terms of linear combinations of these cell means. The ith treatment main effect is defined as $\Sigma\, v_j\, \mu_{ij}$, for known constants v_j, such that $\Sigma\, v_j = 1$. The jth stratum main effect is analogously defined by $\Sigma\, u_i\, \mu_{ij}$, where $\Sigma\, u_i = 1$.

The method recommended for testing equality of treatment main effects is the weighted squares of means analysis proposed by Yates.[25] If one assumes equal weights, as Yates recommended, then the associated sum of squares for testing treatment main effects can be shown to correspond to that obtained by the ANOVA method, using regression model (1), assuming the usual restrictions.[13,24] It is given by

$$Q = \sum_i W_i \hat{\mu}_{i\cdot}^2 - (\sum_i W_i \hat{\mu}_{i\cdot})^2 \,/\, \sum_i W_i$$

where

$$\hat{\mu}_{i\cdot} = 1 \,/\, b \sum_j \bar{y}_{ij}, \qquad W_i^{-1} = 1 \,/\, b^2 \sum_j 1 \,/\, n_{ij}$$

Other sets of weights can also be used in defining treatment main effects, and, in certain instances, may be more appropriate.[7] One such instance occurs when the observed strata frequency distribution of subjects is known to reflect that existing in the target population. The marginal proportions $n_{\cdot j} \,/\, n_{\cdot\cdot}$ would be the weights v_j used to define treatment main effects. The corresponding sum of squares[7,22] would be used to test equality of main effects.

Of course, tests for homogeneity of variance among the individual cells should be performed before one decides whether to assume model (1) or (2). The particular configuration of sample sizes among treatment groups within strata can influence the tests for main effects and interaction. However, the robustness of the analysis of variance procedure relative to unequal cell variances is most notable when cell sample sizes are equal.[3]

To illustrate some of these ideas, I shall use a 36-month clinical trial recently conducted by NIDR, in which a weekly fluoride mouthrinse and a daily fluoride mouthrinse were compared with a placebo rinse.[12] The stratification variable used was the MGSI, which is based upon the initial caries examination.[14] Its value represents the number of distinct regions in a subject's dentition where caries was diagnosed. These regions, originally identified by Grainger, are described in Table 1. The MGSI assumes a value from 0 to 5. For this study, there were only 12 and four subjects who had an MGSI value of 4 or 5, respectively. Accordingly, they were grouped with those scored as 3.

Typical characteristics of subjects in each MGSI class are given by;

MGSI	Description
0	no caries
1	pit and fissure caries only
2	pit and fissure caries & posterior proximal caries
3	pit and fissure caries & posterior proximal caries & anterior caries

Eight hundred and twenty-four subjects in grades 5, 6, and 7 were initially randomly assigned to the three treatment groups. Five hundred and ninety-eight subjects (73%) completed the study. The frequency distributions for subjects initially examined and those who completed the study are presented in Table 2 by MGSI stratum and treatment group. Clearly, adequate balance within the respective strata across treatment groups was realized for subjects who completed the study.

The cell mean increments for each treatment group, together with their standard deviations, are presented in Table 3. It appears that there is interaction between treatment groups and MGSI status, since differences between pairs of treatments increase with MGSI status. The method of fitting constants was used to derive the test for interaction. The F-statistic 2.31 had an associated p-value of 0.03. Thus, interaction was assumed present, and model (1) was considered as appropriate for analyzing these data. For this

TABLE 1
GRAINGER'S ORIGINAL SEVERITY ZONES

Zone 1 – Occlusal pit and fissure surfaces of molars and pre-molars and the buccal pits and lingual grooves of molars

Zone 2 – Approximal surfaces of posterior teeth, including the distal surfaces of canines

Zone 3 – Approximal surfaces of the maxillary anterior teeth, excluding the distal surfaces of canines

Zone 4 – Labial surfaces of anterior teeth

Zone 5 – Approximal surfaces of the mandibular anterior teeth, excluding the distal surfaces of canines

TABLE 2
DISTRIBUTION OF SUBJECTS INITIALLY RANDOMIZED AND SUBJECTS COMPLETING THE STUDY BY GROUP AND MGSI STATUS

	Initially Randomized			Completed Study		
MGSI	Placebo	Weekly	Daily	Placebo	Weekly	Daily
0	39	43	41	25	26	26
1	125	136	128	103	101	98
2	68	71	80	51	51	51
3	36	28	29	25	21	20
Totals	268	278	278	204	199	195

TABLE 3
MEAN CARIES INCREMENTS (STANDARD DEVIATIONS) BY TREATMENT GROUP AND MGSI STATUS

MGSI	Placebo	Weekly	Daily
0	1.80 (2.94)	0.92 (1.55)	0.88 (1.53)
1	2.82 (3.95)	2.41 (3.68)	2.10 (2.60)
2	4.84 (5.05)	3.41 (4.46)	3.18 (4.19)
3	10.00 (8.82)	5.86 (4.21)	4.55 (6.98)

model, the cell sample means are the most efficient estimates of the corresponding population cell means (μ_{ij}).

Although it is impossible to show that the treatments satisfy the consistently ordered property, the assumption seems tenable in light of the pattern observed among the sample cell means. Therefore, main effects (or marginal or average effects) were defined, and tests for their equality were performed by using the weighted squares of means procedure. Equal weights ($v_j = 0.25$) were used in defining main effects. The W's are based on the homogeneity of variance among cells. Although this assumption is not satisfied here (Bartlett's test produced $p = 0.001$), the balance among treatment groups within each stratum renders the homogeneity assumption more of an esoteric than practical consideration for comparing treatment. However, it may greatly influence the other tests, since the smallest frequencies were observed in the cells having the largest variances.

Efficient estimates of treatment main effects under model (1) and their standard errors are presented in Table 4. The results of the analysis of variance are summarized in Table 5.

The comparisons of primary interest in this study include:

(1) placebo rinse *vs.* weekly fluoride rinse,

(2) placebo rinse *vs.* daily fluoride rinse, and

TABLE 4
ESTIMATES (AND STANDARD ERRORS) OF TREATMENT MAIN EFFECTS

	Main Effect	Standard Error
Placebo	4.86	0.34
Weekly	3.15	0.35
Daily	2.68	0.36

TABLE 5
ANALYSIS OF VARIANCE TABLE SUMMARY

Source	df	MS	P
Model	11	181.71	0.001
Interaction	6	40.21	0.033
MGSI Status	3	460.59	0.001
Treatments	2	188.56	0.001
Error	586	17.44	

TABLE 6
SUMMARY OF COMPARISONS AMONG TREATMENTS FOR MGSI STRATA (P-VALUES)

	MGSI Status			
Comparison	0	1	2	3
Placebo *vs.* Weekly	0.454	0.484	0.084	0.001
Placebo *vs.* Daily	0.434	0.226	0.044	0.001
Weekly *vs.* Daily	0.974	0.608	0.776	0.317

(3) weekly fluoride rinse *vs.* daily fluoride rinse.

Treatment contrasts were formulated and tested, for each MGSI status separately, and as main effects. The results for each MGSI status are summarized in Table 6.

The strata-specific treatment comparisons clearly indicate that both the weekly and daily fluoride rinses were effective in preventing caries among subjects in MGSI stratum 3 (those who initially had anterior decay and both types of posterior decay), in spite of the relatively small sample sizes. There appears to be an effect of each fluoride rinse (marginal significance) among subjects in MGSI stratum 2 (both posterior proximal decay and pit and fissure decay). The effect of each fluoride rinse among subjects in MGSI stratum 1 (having only pit and fissure decay) was not significant. Neither was the effect of either fluoride rinse statistically significant for subjects who were initially caries-free. However, there are insufficient numbers of subjects per group to make this comparison adequately. The largest percentage reductions for each fluoride rinse were found for subjects who were initially caries-free (49% and 51% for weekly and daily), and for subjects who initially had all three types of decay (41% and 55%). Differences between the weekly and daily fluoride rinses could not be detected statistically for any MGSI strata.

The tests for main effects are summarized in Table 7. Each of the fluoride rinses was judged effective as a main effect, but the fluoride rinse groups could not be statistically distinguished from each other.

Contrasts among strata main effects showed that the caries increments for the four MGSI strata were all statistically distinguishable, pair-wise. This finding was also obtained when heterogeneity among cell variances was assumed.

TABLE 7
SUMMARY OF COMPARISONS AMONG TREATMENT MAIN EFFECTS (P-VALUES)

Comparison	Difference	P-value
Placebo *vs.* Weekly	1.71	0.001
Placebo *vs.* Daily	2.18	0.001
Weekly *vs.* Daily	0.47	0.351

TABLE 8
SUMMARY OF PRECISION COMPARISON

Method	Mean Square Error	F-statistic (d.f.)
Without Stratification	20.05	7.11 (2,595)
With Stratification	17.44	10.81 (2,586)

The analysis of covariance was also performed on these data. Separate regression lines were assumed for each treatment group. The test for homogeneity of slopes was rejected ($p = 0.008$). This is consistent with the significant interaction detected in the stratification analysis. This re-emphasizes that the focal point for making inferences about treatments ought to be on the types of interaction, *i.e.*, treatment comparisons at specific levels of the MGSI factor, rather than on main effects of treatment. For this study, one might consider comparing treatments at the overall mean of the MGSI variable, or for specific values of MGSI, since the MGSI averages are essentially equal across groups.

The gain in efficiency of making treatment comparisons can be assessed by comparing the magnitudes of the mean square errors and the F-statistics with and without stratification for the no-interaction model. It is much more difficult to assess in the stratification model with interaction. The mean square errors and F-statistics are given in Table 8. There appears to be an increase in precision here, but its magnitude is rather marginal. This is not surprising, since the simple correlations between MGSI status and caries increments were found to be 0.41, 0.32, and 0.26 for the placebo, weekly, and daily fluoride rinse groups, respectively. Thus, the advantage of stratification in this study lies primarily in the delineation of treatment effects for the sub-populations representing the distinct profiles in initial caries experience found among this school-aged population.

In summary, the post-stratification method can be quite useful in analyzing caries clinical trials data. It is a useful method for adjusting treatment main effects for imbalance relative to a prognostic factor, or to increase the precision of the study by eliminating that portion of the error variance due to the prognostic factor. Furthermore, it enables the investigator to assess the effects of treatment separately for subjects who initially presented different types of caries experience.

The required assumptions for the stratification method are much less restrictive than those needed for the analysis of covariance. For example, it can be used when moderate or strong non-linear relationships exist between a covariate and caries increment. It can also be used as an alternative to covariance analysis whenever there is severe non-parallelism in slopes among treatment groups for the covariate. In such instances, this author would prefer partitioning the range of the covariate into a small set of strata, and proceed to analyze the data by the stratification method. Tests for main effects, if essential, can still be performed, using the interaction model. The stratification method has the advantage of being more easily understood, and more amenable to interpretation by the investigator.

REFERENCES

1. APPELBAUM, M.I. and CRAMER, E.M. (1974): Some Problems in the Nonorthogonal Analysis of Variance, *Psych Bull* 81:335-343.
2. BANCROFT, T.A. (1968): **Topics in Intermediate Statistical Methods,** Vol. 1. Ames, Iowa: Iowa State University Press.
3. BOX, G.B. (1954): Some Theorems on Quadratic Forms Applied in the Study of AOV Problems. II. Effects of Inequality of Variance and of Correlation Between Errors in the 2-way Classification, *Ann Math Stat* 25:484-498.
4. BROWN, B.W., Jr. (1980): Designing for Cancer Clinical Trials: Selection of Prognostic Factors, *Cancer Treatment Rep* 64: 499-502.
5. BYAR, D.P.; SIMON, R.M.; FRIEDEWALD, W.; SCHLESSELMAN, J.; DeMETS, D.; ELLENBERG, J.; MITCHELL, G.; and WARE, J. (1976): Randomized Clinical Trials, *N Engl J Med* 295:74-80.
6. DOWNER, M. (1978): Caries Prediction from Epidemiologic Data. In: **Methods of Caries Prediction,** Bibby, B.G. and Shern, R.J., Eds., Washington and London: Information Retrieval, Inc., pp. 37-42.
7. ELSTON, R.C. and BUSH, J. (1964): The Hypotheses that Can Be Tested When There are Interactions in an Analysis of Variance Model, *Biometrics* 20:681-699.
8. GOSSLEE, D.G. and LUCAS, H.L. (1965): Analysis of Variance of Disproportionate Data When Interactions are Present, *Biometrics* 21:115-133.
9. GRAINGER, R.M.; LEHNHOFF, R.W.; BOLLMER, B.W.; and ZACHERL, W.A. (1984): Analysis of Covariance in Dental Caries Clinical Trials, *J Dent Res* 63 ( ):
10. GRAYBILL, F.A. (1976): **Theory and Application of the Linear Model.** Belmont, CA: Wadsworth, Duxbury Press.
11. GRIZZLE, J.E. (1982): A Note on Stratifying Versus Complete Random Assignment in Clinical Trials, *Controlled Clin Trials* 3:365-368.
12. HEIFETZ, S.B.; MEYERS, R.J.; and KINGMAN, A. (1982): A Comparison of the Anti-caries Effectiveness of Daily and Weekly Rinsing with Sodium Fluoride Solutions: Final Results, *Pediatr Dent* 4:300-303.
13. HOCKING, R.R. and SPEED, F.M. (1975): A Full Rank Analysis of Some Linear Model Problems, *J Amer Statist Assoc* 70:706-712.
14. KINGMAN, A. (1979): A Method of Utilizing the Subjects' Initial Caries Experience to Increase Efficiency in Caries Clinical Trials, *Community Dent Oral Epidemiol* 7:87-90.
15. KLEINBAUM, D.G. and KUPPER, L.L. (1978): **Applied Regression Analysis and Other Multivariate Methods.** Belmont, CA: Wadsworth, Duxbury Press.
16. KORTS, D.C. (1977): Analysis of Caries Clinical Trial Data in the Presence of Study Group Imbalance, *Community Dent Oral Epidemiol* 5:231-236.
17. MEIER, P. (1981): Stratification in the Design of a Clinical Trial, *Controlled Clin Trials* 1:355-361.
18. PETO, R.; PIKE, M.C.; ARMITAGE, P.; BRESLOW, N.; COX, D.R.; HOWARD, S.V.; MANTEL, N.; McPHERSON, K.; PETO, J.; and SMITH, P.G. (1976): Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient. I. Introduction and Design, *Br J Cancer* 34: 585-612.
19. PETO, R.; PIKE, M.C.; ARMITAGE, P.; BRESLOW, N.; COX, D.R.; HOWARD, S.V.; MANTEL, N.; McPHERSON, K.; PETO, J.; and SMITH, P.G. (1976): Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient. II. Analysis and Examples, *Br J Cancer* 35:1-39.
20. POCOCK, S.J. and SIMON, R. (1975): Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial, *Biometrics* 31:103-115.

21. Principal Requirements for Controlled Clinical Trials of Caries Preventive Agents and Procedures (1982). Technical Report of the Commission on Oral Health, Research and Epidemiology, *Int Dent J* 32:292-310.
22. SCHEFFE, H. (1959): **The Analysis of Variance.** New York: John Wiley & Sons.
23. SEARLE, S.R. (1971): **Linear Models.** New York: John Wiley & Sons.
24. SEARLE, S.R.; SPEED, F.M.; and HENDERSON, H.V. (1981): Some Computations and Model Equivalences in Analyses of Variance of Unequal-Subclass-Numbers Data, *Amer Statistician* 35:16-33.
25. YATES, F. (1934): The Analysis of Multiple Classifications with Unequal Numbers in the Different Classes, *J Amer Statist Assoc* 29:52-66.

Stratification Methods in Caries Clinical Trials: Discussion of Dr. Kingman's Presentation

H. SMITH, Jr.

Mt. Sinai School of Medicine, New York, New York 10025

J Dent Res 63(Spec Iss):777, May, 1984

Dr. Kingman has written an excellent paper on the use of stratification methods in clinical trials. The emphasis in the paper is on post-stratification methods that can be used with the two-way classification model, with or without the interaction term. His analysis is done using the least-squares techniques: method of fitting constants and the method of weighted squares of means.

The paper needs clarification on several assumptions made for the analyses. I quote: ". . .it makes sense to test for main effects in the presence of interaction if one can assume that the treatments are consistently ordered. . . one must also assume that the confounding due to unequal cell frequencies is not too severe." How severe? In the clinical trial time frame – say, 36 months – how do withdrawals from the trial affect this statement?

The same sort of caution on unequal cell sizes takes place in other places in the paper. It would be helpful if Dr. Kingman gave us some idea of how unequal or how unbalanced the data could be without affecting the analysis. Post-stratification will, in general, produce unequal cell sizes and may be a very important consideration.

As to the example, I can only say that whenever an author produces a real set of data he's asking for trouble. Not to be outdone, I have some real reservations about the trial. For example, the fluoride arms were weekly and daily rinses – what about the placebo? Weekly, daily, no rinse? Why were only 268 people randomly allocated to the placebo arm, while 278 people were randomly allocated to the other two arms? Yet, more people finished the trial on the placebo arm than on the two treated arms. The completion rates are 76% on a base of 268 for Placebo, 72% on a base of 278 for Weekly, and 70% on a base of 278 for Daily. Were the placebo patients handled differently? This is a strange result, in my experience.

Table 2 represents a set of data which looks as if the standard deviation is proportional to the mean, thus requiring a log transformation to stabilize the variance. In particular, the significant interaction alluded to in the paper can be attributed to the placebo group, with an initial MGSI of 3. Otherwise everything looks parallel.

Finally, this is a 36-month trial, and there is no way to indicate when the ultimate trends shown in Table 2 took place. For example, what did the Table look like after 12 and 24 months? These aspects are not shown.

All in all, I found the paper excellent, and I congratulate Dr. Kingman on his fine contribution to the conference.