Effectiveness of Remune

Churdboonchart et al. (2) paint an impressive picture of the effects of Remune on CD4 cell count in HIV-infected subjects. However, as individuals who were closely involved in the study, we believe that the paper presents a misleading account of the study results and a distorted view of the beneficial effects of Remune in this population.

We became involved in this study during its planning stages as a result of a request from Thailand’s AIDS Subcommittee for HIV Vaccine Trials, National Commission for the Prevelopment and Control of AIDS, to provide statistical expertise. We were integrally involved in the study design, contributed to the development of case report forms, provided data management training for the sites and for Dr. Churdboonchart’s staff, set up the study randomization and held the blinded treatment codes during the study, and gave presentations to the study’s Data and Safety Monitoring Board (DSMB) on the role of DSMBs and the planned interim analysis for this study. One of us (S.K.) served as a member of the DSMB. We developed an analysis plan for the final study data that was approved by Dr. Churdboonchart, conducted the interim analysis of the study and presented this to the DSMB, and prepared a final statistical report on the study results.

The prespecified primary analysis of CD4 cell count for this study used the summary statistic approach (3), in which a slope was computed for each subject by fitting a linear regression to his or her log-transformed CD4 measurements at weeks 0, 12, 24, 36, and 40 and where the resulting slopes were then compared between the Remune and placebo (incomplete Freund’s adjuvant) groups using the van der Waerden nonparametric test (4). This prespecified primary analysis of CD4 cell count yielded a P value of 0.34, indicating no significant difference between the Remune and placebo groups.

There were also several prespecified secondary analyses of the CD4 endpoint, all based on computing a single summary statistic for each subject and then comparing the Remune and placebo groups using the van der Waerden nonparametric test (4). These secondary analyses were based on using untransformed (as opposed to log-transformed) CD4 counts, two alternative metrics to the CD4 slope (change between baseline and week 40 and normalized area under the CD4 curve [AUC]), and an alternative method for calculating an individual CD4 count based on the “averaging” method described by Churdboonchart et al. (2). The analyses based on the averaging method were added as secondary analyses at the request of Dr. Churdboonchart at the completion of the study. Table 1 lists the results of the primary analysis of CD4 counts and the nine secondary analyses included in our final report. We have not adjusted any of these P values to control for the inflated false-positive rate that arises when multiple tests are conducted (American Statistical Association [ASA] ethical guidelines for statistical practice [http://www.amstat.org/profession/ethicalstatistics.html]).

Note that the primary analysis and seven of the nine secondary analyses of CD4 cell count fail to demonstrate a statistically significant difference between the Remune and placebo groups. Although some of the secondary analyses suggested a possible difference between the Remune and placebo groups, the multiplicity of tests undertaken, as well as the fact that these were secondary analyses, argues against much emphasis being placed on them (ASA guidelines [see above]). Accordingly, the final statistical analysis of the study that we prepared for Dr. Churdboonchart and the AIDS Subcommittee noted that while some of the secondary analyses were suggestive of a possible association between Remune and CD4 count, the study data overall did not demonstrate a significant difference in CD4 between the Remune and Placebo groups.

In contrast to these results, Churdboonchart et al. present only a single analysis of CD4 count, corresponding to analysis 10 in Table 1 but using the Wilcoxon test (4) instead of the van der Waerden nonparametric test (4). This prespecified primary analysis of CD4 cell count yielded a P value of 0.024, indicating a significant difference between the Remune and placebo groups.

Running multiple tests on the same data set at the same stage of an analysis increases the chances of obtaining at least one invalid result. Selecting the one “significant” result from a multiplicity of parallel tests poses a grave risk of an incorrect conclusion. Failure to disclose the full extent of tests and their results in such a case would be highly misleading.

In our opinion, the paper by Churdboonchart et al. gives a distorted account of the clinical trial by virtue of its incomplete and selective reporting of the CD4 cell count results.

REFERENCES

1. Bailar, J. C., III, and F. Mosteller (ed.), 1992. Medical uses of statistics, 2nd ed. New England Journal of Medicine Books, Boston, Mass.

TABLE 1. Statistical analyses of CD4 count

| Method for calculating CD4 count | Log transformation | Metric       | P*  |
|---------------------------------|--------------------|--------------|-----|
| 1. original                     | Yes                | Slope        | 0.34* |
| 2. recalculation                | Yes                | Slope        | 0.36 |
| 3. original                     | Yes                | Change by wk 40 | 0.34 |
| 4. recalculation                | Yes                | Change by wk 40 | 0.20 |
| 5. original                     | No                 | Change by wk 40 | 0.13 |
| 6. recalculation                | No                 | Change by wk 40 | 0.07 |
| 7. original                     | Yes                | AUC          | 0.11 |
| 8. recalculation                | Yes                | AUC          | 0.07 |
| 9. original                     | No                 | AUC          | 0.044 |
| 10. recalculation               | No                 | AUC          | 0.024 |

* Not adjusted for multiple comparisons.
* Prespecified primary analysis.
Kim and Lagakos state “Had AUCMB been used as the primary analysis for this study lacks a valid scientific rationale. We believe that Dr. Lagakos’ preference for the slope metric as the primary analysis but received no reply. We had also written to Dr. Lagakos regarding this preference for the slope metric as the primary analysis but received no reply. We believe that Dr. Lagakos’ preference for the slope metric as the primary analysis for this study lacks a valid scientific rationale. Indeed, in their final analysis report (21 October 1999) Drs. Kim and Lagakos state “Had AUCMB been used as the primary analysis it would have been declared significant.”

In summary, our article reports absolute changes in CD4 cell counts, which we and other clinicians in Thailand have deemed clinically significant and important from a public health perspective. We have utilized AUC, one of the most utilized metrics for surrogate markers in AIDS clinical trials. It is important to realize that one of the first AIDS trials to utilize AUC was a study comparing dideoxyinosine (ddI) to zidovudine (AZT), which revealed a small (15-cell maximum) difference for one of the two doses of ddI and AZT, which was not significant for the mean or median counts. However, comparison of the AUCs showed a significant difference (P < 0.03) between each of the ddI arms and AZT. It was this analysis which contributed to the approval of ddI (6). We believe the insistence on the slope metric without a valid scientific rationale is therefore unjustified. Dr. Tukey, a leader of modern statistical theory, warned that statistics should not be “sanctified” to impede scientific progress (1). We are indeed pleased that both the Technical and the Ethical committees of the Thailand Ministry of Health have reviewed all of the information from this trial and have approved further clinical development of Remune in Thailand.