Controlled Clinical Trials

Tukey (*1*), in discussing multiple analyses as data accumulate during a controlled clinical trial, leaves the reader with the impression that repeated "looks at the data" invariably require that a larger level of significance be assigned to the overall procedure. In making this point, he cites the work of Armitage. McPherson, and Rowe (*2*) in which only procedures with a fixed upper bound on the sample size are considered.

Tukey fails to mention that there is a general method that can be applied to such sampling procedures which permits continual testing of the data without affecting the overall level of significance. The method in question is based on the idea that, as soon as sufficient data have been accumulated so that the outcome of the test to be performed on the completed data set is certain, then the inference provided by the test can be made immediately at the nominal level of significance [see (*3*), pg. 719]. Applications of this (early stopping) idea have been described in the binomial case (*4*), the Wilcoxon two-sample test (*3*, *5*), and in several other tests (*6*).

The early stopping idea can be applied in principle to a number of statistical tests currently employed in clinical trials; however, the complex design and actual conduct of some trials may make it difficult at the outset to define appropriate stopping rules of any kind. When this idea can be applied the resulting savings in observations (and time) will vary with the statistical test and the pattern of entry of patients into study. Savings may be quite substantial for the Wilcoxon test but are only modest in the binomial case; in general, savings tend to decrease as the length of time during which patients enter study increases. Use of the early stopping idea ensures that the inference made at the time of stopping will be the same as that which would be made if the trial continued to its planned conclusion, that is, the inference based on the outcome in all the patients. In this respect the procedure differs from the type suggested by Tukey (*7*) for which it is possible for the two inferences in question to be conflicting.

Tukey asserts "Once our clinical trial has accumulated favorable evidence for an innovation up to whatever level of significance . . . physicians judge appropriate for action, we cannot, ethically, continue the trial . . . *just to measure the improvement* with greater precision." We think this statement might be broadened to assert that a clinical trial should be stopped as soon as the final inference is a foregone conclusion. Application of the early stopping idea is a method of continually testing this possibility.

DAVID W. ALLING
MAX HALPERIN, JAMES H. WARE
National Institutes of Health, Bethesda, Maryland 20014

References and Notes

1. J. W. Tukey, *Science* **198**, 679 (1977).
2. P. Armitage, C. K. McPherson, B. C. Rowe, *J. R. Stat. Soc. A* **132**, 235 (1969).
3. D. W. Alling, *J. Am. Stat. Assoc.* **58**, 713 (1963).
4. ———, *Biometrika* **53**, 73 (1966).
5. M. Halperin and J. H. Ware, *J. Am. Stat. Assoc.* **69**, 414 (1974).
6. D. Chun, *ibid.* **60**, 859 (1965); N. Breslow, *ibid.* **65**, 639 (1970); R. G. Miller, Jr., *ibid.*, p. 1554.
7. See Tukey (*1*) in the last paragraph of the section entitled "Not-very-sequential designs."

19 January 1978; revised 5 September 1978

I do not disagree with Alling, Halperin, and Ware. I should be glad to see such "curtailed sampling" applied. "Stopping when you know what the answer will be" is very different from "stopping when the answer looks good."

JOHN W. TUKEY
Bell Laboratories, Murray Hill, New Jersey 07974

6 October 1978

Models for Carcinogenic Risk Assessment

Cornfield (*1*) presents a simple kinetic model for carcinogensis that could lead to a threshold. Cornfield shows that such a threshold depends on the existence of at least one irreversible completely protective reaction in the carcinogenic process. He then asserts that this model has as much theoretical justification as others and therefore should be used as part of the safety evaluation procedures for governmental regulatory agencies. We do not argue against the possible existence of such thresholds, nor do we disagree with the potential utility of models based on pharmacokinetics; however, we wish to point out inadequacies in Cornfield's derivation and express our concern with the implications of this article on the assessment of low-exposure carcinogenic risk.

Cornfield's threshold model implicitly requires either instantaneous deactivation of the toxic substance at the target site or complete deactivation before reaching the target. His derivation of this model is based on a steady-state solution to a process in which an irreversible deactivation reaction takes place in vivo. We agree that in the presence of a single exposure, when the amount of the toxic substance is less than the amount of the deactivator, then all of the toxic substance will eventually be converted to the deactivated substance. However, this process will take time to reach equilibrium and, under Cornfield's assumption of proportionality between dose and the probability of a carcinogenic response, if any activated complex ever reaches the target site, a threshold will not exist. We also call attention to the unrealistic assumption of a single exposure implicit in Cornfield's derivation. The primary concern with environmental carcinogens is with situations of either continuous exposure, such as agents in the air we breathe, or with repeated exposures from food or water additives or contaminants. A more realistic model would require continuous production and degradation of deactivator and substrate as well as continuous or repeated exposure to the toxic substance.

Cornfield's model assumes the existence of a single threshold applicable to each member of the exposed population. However, thresholds may vary over time within an individual as well as varying among individuals in the population. This variability of thresholds is most important from a regulatory point of view since all members of a heterogeneous population must be protected at all times. His establishment of a single "population threshold" is of little value to a regulatory agency that must consider the lowest threshold for an individual over his exposure period, as well as the lowest thresholds in the entire population.

Answers to the questions raised by Cornfield are likely to be obtained through basic research in carcinogenic pathways and pharmacokinetics and not by examining limited animal bioassay experiments. With limitations in the current knowledge of carcinogenesis, we see no alternatives for the prudent regulator but to base his decisions on the assumption of no threshold.

CHARLES C. BROWN
THOMAS R. FEARS, MITCHELL H. GAIL
MARVIN A. SCHNEIDERMAN
ROBERT E. TARONE
National Cancer Institute, Bethesda, Maryland 20014
NATHAN MANTEL
Biostatistics Center, George Washington University, Bethesda, Maryland 20014

References

1. J. Cornfield, *Science* **198**, 693 (1977).

6 March 1978

0036-8075/78/1208-1105$00.50/0

Models for Carcinogenic Risk Assessment

CHARLES C. BROWN, THOMAS R. FEARS, MITCHELL H. GAIL, MARVIN A. SCHNEIDERMAN, ROBERT E. TARONE and NATHAN MANTEL

Science **202** (4372), 1105.
DOI: 10.1126/science.202.4372.1105-b