3. Doblar DD, Hinkle JC, Fay ML, Condon BF: Air embolism associated with pulmonary artery catheter introducer kit. ANESTHESIOLOGY 56:307–309, 1982
4. Rao TLK, Gorski DW, Laughlin S, El-Etr AA: Safety of pulmonary artery catheterization. ANESTHESIOLOGY 57:A116, 1982
5. Sise MJ, Hollingsworth P, Brimm JE, Peters RM, Virgilio RW, Shacksfurd SR: Complications of the flow-directed pulmonary artery catheter. Crit Care Med 9:315–318, 1981
6. Hardy J-F, Morissette M, Taillefer J, Vauclair R: The pathophysiology of pulmonary artery ruptures by pulmonary artery balloon tipped catheters. ANESTHESIOLOGY 59:A127, 1983

(Accepted for publication August 21, 1984.)

Anesthesiology
62:97–98, 1985

Multiple Comparison Procedures in the Analysis of Designed Experiments

To the Editor: — An editorial[1] and recent correspondence[2] discussed problems in the analyses of experimental data from more than two groups reported in the medical literature. It was pointed out that use of multiple *t* tests (or Fisher's unmodified LSD test), in which the Type 1 error (alpha, or the probability of false-positive results) is controlled at some fixed value, commonly 0.05, for each comparison leads to a Type 1 error greater than alpha for the set of tests as a whole.[1] Thus, the probability can be quite high that some of the differences found to be significant actually arose by chance. On the other hand, use of multiple comparison procedures, such as Tukey's test or Scheffe's test, designed to control Type 1 error for the set of comparisons as a whole, results in increased Type 2 error (beta, or probability of false negatives).[2] These articles will be most useful to readers who must interpret published results of research; however, little practical guidance has been given to the researcher who must choose an analysis suited to his particular data.

The relative merits of the two approaches to Type 1 error have been described clearly,[1,2] but nothing has been said about the type of questions being asked by the researcher or the aims of his study. These latter considerations are of vital importance in choosing the correct test.[3,4] The types of hypotheses that a researcher may wish to test can be divided into two categories.

In the medical sciences, a researcher who designs and carries out an experiment almost invariably has specific hypotheses or predictions that he aims to test. These hypotheses have been formulated before the experiment was conducted and are reflected in its design. Usually, the number of comparisons required to test such *a priori* hypotheses will be small when compared with the total number of comparisons that potentially could be made. These *a priori* hypotheses will, of course, always be tested in the analysis. When the *a priori* comparisons form or are part of an orthogonal set (*i.e.*, the outcome of each comparison is independent of the outcomes of all other comparisons), multiple comparison procedures are not required and the hypotheses should be tested individually.[4] In practice, the *a priori* hypotheses often will not be mutually independent. Even in this case, Winer[4] states that, when the number of *a priori* hypotheses is small, they should be tested individually at the chosen level of significance.

It also may happen that, having tested his *a priori* hypotheses (or, rarely, having none), and wishing to make the most of the resources invested in the experiment, the researcher deliberately sets out to sift through his data to see if any unexpected but possibly interesting effects have been uncovered. Or, perhaps, the results of the experiment suggest that some interesting effect may exist that had not been anticipated. In any complex experiment, there are a large number of such *a posteriori* comparisons (it should be noted that a deliberate selection of the two most extreme means implies the comparison of all possible pairs of means), and consequently some large differences are likely to arise by chance. To disallow such *post hoc* data snooping would do nothing to aid the advance of science; indeed, unexpected findings sometimes may be the most important. However, these *a posteriori* comparisons are of a clearly different nature to those made to test *a priori* hypotheses. Readers of the scientific literature should be protected from the deluge of spuriously significant results that would appear if this distinction was not made. Therefore, most researchers will wish to control the Type 1 error for the set of *a posteriori* comparisons as a whole, and so should use an appropriate multiple comparison procedure.

If the approach outlined above is followed, then the argument that use of multiple comparison procedures leads to requirements for larger sample sizes and greater investment in experiments[2] clearly does not apply to those hypotheses that the experiment was designed to test. What does become clear is the crucial importance of identifying and specifying the hypothe-

ses to be tested *before* the experiment is carried out. If this is done, there will be no doubt as to which hypotheses may be tested individually and which require use of multiple comparison procedures.

There is a further error in making analytic comparisons that commonly is seen in medical literature. This is the failure to recognize when pairwise comparison of means is inappropriate. Often, the treatments represent steps along some ordered scale, *e.g.*, different doses of a drug. The ordered nature of the treatments often is ignored in the analysis, and comparisons of pairs of means are made using *t* tests or other procedures.[5] Usually the hypotheses to be tested will relate to the existence or form of the dose–response relationship, rather than which particular dose level is different from which other dose level or from placebo. Response curve analysis is the appropriate technique in this case.[5]

In recent years, the cost and ethical implications of improper statistical analysis of medical research data have been a problem of growing concern. The editors of ANESTHESIOLOGY are to be commended for their efforts to improve the quality of data analysis and reporting in the journal.

JOHN L. PLUMMER, PH.D.
NH & MRC Senior Research Officer
Department of Anaesthesia and Intensive Care
Flinders Medical Centre
Bedford Park, South Australia 5042

REFERENCES

1. Longnecker DE: Support versus illumination: Trends in medical statistics. ANESTHESIOLOGY 57:73–74, 1982
2. Schaeffer CF: Regarding the misuse of t-tests. ANESTHESIOLOGY 60:505, 1984
3. O'Brien PC: The appropriateness of analysis of variance and multiple comparison procedures. Biometrics 39:787–788, 1983
4. Winer BJ: Statistical principles in experimental design, 2nd edition. New York, McGraw-Hill, 1971
5. Dawkins HC: Multiple comparisons misused: Why so frequently in response curve studies? Biometrics 39:789–790, 1983

(Accepted for publication August 21, 1984.)

Anesthesiology
62:98–99, 1985

A Simple Nomogram for Determining Drug Infusion Rates

To the Editor:—The infusion of many vasoactive drugs has become commonplace in the practice of anesthesia during the last decade. Some methods for simplifying the calculation of doses for infusion were reported in the Journal recently.[1,2] However, it would be preferable if you could tell the rate of infusion just by

TABLE 1. Dilution Nomogram: Dilute 100 mg of a Drug to X ml, and the Set Pump Rate (ml/h) Is Equal to Drug Infusion Rate in $\mu g \cdot kg^{-1} \cdot min^{-1}$

BW (kg)	X (ml)	BW (kg)	X (ml)	BW (kg)	X (ml)	BW (kg)	X (ml)
1	1666.7	21	79.4	41	40.7	61	27.3
2	833.3	22	75.8	42	39.7	62	26.9
3	555.6	23	72.5	43	38.8	63	26.5
4	416.7	24	69.4	44	37.9	64	26.0
5	333.3	25	66.7	45	37.0	65	25.6
6	277.8	26	64.1	46	36.2	66	25.3
7	238.1	27	61.7	47	35.5	67	24.9
8	208.3	28	59.5	48	34.7	68	24.5
9	185.2	29	57.5	49	34.0	69	24.2
10	166.7	30	55.6	50	33.3	70	23.8
11	151.5	31	53.8	51	32.7	71	23.5
12	138.9	32	52.1	52	32.1	72	23.1
13	128.2	33	50.5	53	31.4	73	22.8
14	119.0	34	49.0	54	30.9	74	22.5
15	111.1	35	47.6	55	30.3	75	22.2
16	104.2	36	46.3	56	29.8	76	21.9
17	98.0	37	45.0	57	29.2	77	21.6
18	92.6	38	43.9	58	28.7	78	21.4
19	87.7	39	42.7	59	28.2	79	21.1
20	83.3	40	41.7	60	27.8	80	20.8