Reports and Letters

Bias in the Allocation of Treatments by Random Numbers

When an investigator wishes to test the significance of differences between the effects of two experimental treatments, he needs an unbiased method of allocating the treatments to the various units of his material (animals, human beings, test tubes). The purpose of this paper is to make better known a simple method of allocation and to compare it with a popular but potentially hazardous method (*1*).

Random method of allocation. The recommended method of allocation may best be illustrated by an example. Suppose it is desired to assign 20 subjects to treatments A and B, respectively. The 20 subjects are represented by serially numbered index cards. The random numbers table is entered at an arbitrary point, and two-digit random numbers are read, moving along horizontally. (In the case of Fisher and Yates' tables (*2*), the random numbers may also be used in the vertical direction.) Index card No. 1 receives the first two-digit number, index card No. 2 the second, and so on. The cards are then sorted in ascending order of random number (00 regarded as 100). If the subjects are to be equally divided between treatments A and B, those represented by the first 10 cards may receive treatment A and the remaining subjects treatment B (or vice versa). Duplicates, triplicates, and so forth are ignored unless they fall at the boundary between treatments. For instance, the tenth and the eleventh cards may bear the same random number. Then the choice regarding which of the two should go into the treatment-A sample can be made by drawing two numbers, one for each of the cards, from another block of random digits after stipulating that the higher number shall correspond to, say, treatment A. (In case of triplicates, three additional random numbers must be drawn.) If one wishes to allocate subjects to more than two treatments, one simply breaks the array of subject cards, which are sorted in ascending order of the affixed random numbers, into as many equal parts as there are treatments, and assigns a different treatment to each part. The method can also be applied when the parts are unequal but of assigned size.

Biased method of allocation. If two treatments are to be compared, but there is no stipulation regarding the equality (or other ratio) of numbers of subjects, it is perfectly legitimate, although possibly inefficient, to decide by coin tossing which treatment shall be applied to each subject; or, for greater convenience, one may use odd and even random numbers. Frequently, however, this method has been applied in equal-number experiments, by allocating randomly until half the subjects have been assigned to one of the treatments and then assigning the remainder automatically to the other treatment. Under those conditions, the tossing of a coin with equal head-tail probability, or the equivalent "odd-even" random numbers method, yields nonrandom allocations such that significance tests tend to be misleading, because what is believed to be, say, the 5-percent level of significance (that is, 5 percent of wrong verdicts when chance alone is operating) may be something quite different in reality.

Illustration of bias. Table 1 shows the percentage of wrong verdicts "significant" in 200 samples allocated by the odd-even method and in 300 samples allocated by the random method for sample size 20, with equal numbers of subjects on imaginary treatments A and B. The theoretical values for the odd-even method are found by probability techniques (*3*) that will be submitted elsewhere for publication. To test results in the case of Fisher's "exact" contingency test (*4*), which with small samples is preferred to the chi square test, we have considered four types of heterogeneity in the experimental material. Irrespective of treatment, certain specified subjects (for example, 1 to 5) are destined to die while the rest (6 to 20) live. Apparent effects of treatment are measured by death or survival. The expected percentages of wrong verdicts for the contingency tests in the case of the random method were taken from Mainland (*5*). Agreement between experimental sampling and expected values for the random allocation is quite satisfactory for experiments of this size. To show results in the use of the *t* test, we have chosen experimental material that contains a simple linear trend such that each serially numbered subject differs from the preceding one by plus one unit of measurement.

The proportions of wrong verdicts for the contingency tests indicate that the odd-even method produces the worst results when the end part of the allocation contains a number of individuals with characteristics different from those allocated earlier, as when subjects 16 to 20 die. (This is because the odd-even method assigns longer runs at the end of a sequence than does the truly random method.) While this bunching of like individuals may appear extreme, its occurrence in practice is, nevertheless, entirely possible when the characteristics of the subjects change with time. If a few odd individuals are scattered haphazardly through the experimental material, the odd-even method will yield results that are hardly disinguishable from those obtained by the random method. On the other hand, it is seen that, when two relevant characteristics alternate in the sample, the odd-even method actually produces fewer wrong verdicts "significant" when chance alone is operating;

Table 1. Percentage of wrong verdicts in testing the difference between two "treatments" applied to ten subjects each, using random and biased methods of allocation by means of random numbers.

Condition of material	Biased method ("odd and even")		Random method	
	Theoretical	Exptl. (200 samples)	Expected	Exptl. (300 samples)
"Exact" contingency test				
Subjects 1 to 5 die	6.25	8.00	3.26	3.67
Subjects 1 to 10 die	10.94	13.00	2.30	2.00
Subjects 16 to 20 die	30.18	29.50	3.26	1.33
Subjects 1, 3, 5, . . . , 17, 19 die	1.01		2.30	
Student's t *test*				
Subjects 1 to 20 have linear trend in measurements	Not computed	25.50	Appr. 5.00	5.00

however, this is accompanied by an increase in the wrong verdicts "not significant" when there is a real difference between the effects of the treatments.

L. HERRERA

Department of Medical Statistics, New York University College of Medicine, New York

References and Notes

1. I am greatly indebted to Donald Mainland for suggesting the method, for his constructive criticisms and suggestions, and for his interest and encouragement. I am indebted to both Mainland and Ruth M. Smith for carrying out the sampling experiments. For random allocation methods also see A. L. Edwards, *Experimental Design in Psychological Research* (Rinehart, New York, 1951), p. 23.
2. R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver and Boyd, London, ed. 3, 1948), p. 104.
3. W. Feller, *An Introduction to Probability Theory and its Applications* (Wiley, New York, 1950), pp. 78–81 and 217–218.
4. R. A. Fisher, *Statistical Methods for Research Workers* (Hafner, New York, ed. 11, 1950), p. 96.
5. D. Mainland, *Can. J. Research* E26, 1 (1948).

11 July 1955

Excretion Patterns of Rats Following Total-Body Exposure to X-radiation

Nitrogen and sulfur metabolism change radically following total-body exposure to ionizing radiation, with excretion increasing within 1 day and remaining high in nonsurvivors but decreasing after a period in survivors (*1–5*). Detection of early metabolic changes could be of vital importance in a national emergency (*6*).

Twelve young (180 ± 5g) male Sherman-strain rats, after a 3-hour fast, were irradiated in lucite chambers with 550 r (LD_{42}^{30}) delivered from a deep-therapy G.E. Maximar X-ray machine (*7*). Animals were maintained at 25°C and were fed Purina Lab Chow and water *ad libidum*. Complete individual 24-hour fasting (water supplied) urine samples were collected before and at intervals after irradiation. Thirteen compounds per sample were determined in triplicate using paper chromatographic and colorimetric methods (Table 1). The individual means were averaged by groups surviving 30 (*5*), 14 (*2*), and 11 (*3*) days, groups I, II, and III, respectively. Each animal served as its own control owing to the wide individual variation, and all percentage changes were related back to the preirradiation-sample values of each group.

Urine volumes and the excretion of most compounds became elevated in all groups. Glycine was subnormal throughout the postirradiation period, while taurine, valine, and aspartic acid were, for the most part, subnormal after the first day. Urea and alanine were elevated during the entire period. One-day postirradiation phosphate, taurine, and alanine increased, while histidine and aspartic acid decreased progressively from groups I to II to III, and the uric acids of the nonsurvivors were significantly greater than those of group I. On day 5, alanine and glycine trends remained unchanged, while group III diverged further from groups I and II by maintaining an elevated phosphate, uric acid, and creatinine and a depressed histidine. Glutamic acid and aspartic acid dropped in all.

During the acute phase, when animals were dying (days 10 to 14), urine volumes, phosphate, urea, and alanine were elevated, while taurine and glycine were depressed. The divergence between groups I and II and group III was marked on day 9 with respect to everything except glycine, valine, and aspartic acid. Animals dying on day 14 (group II) were markedly different from those in group I in everything except phosphate, glycine, alanine, and taurine, which were now at about the same levels as the premortal values of group III. Creatinine and aspartic acid were depressed in group II, while group I had returned to normal levels. Uric acid increased in both. Histidine dropped in group II to the premortal levels of group III, while it steadily increased in group I. Values of groups II and III just prior to death were nearly the same for some compounds but were significantly different for others. It is apparent that some of these changes are common precursors of impending death,

Table 1. Relative percentage changes in excretion patterns of rats exposed to a total-body dose of 550 r x-radiation

Compound determined	Survival group	Post irradiation days							
		−3	1	5	9	13	17	23	29
Phosphate	I	100	163†	83	144	201†	150	121	136
	II	100	209	64	141	208			
	III	100*§	212*†	160	214*†‡				
Creatinine	I	100	112	84	93	115	128†	107	130
	II	100	101	88	119	72*			
	III	100*§	197*†	134*	135*‡				
Urea	I	100	127	101	111	140†	149†	113	142†
	II	100	224*	139	143	161			
	III	100*	114*	144	119*‡				
Uric acid	I	100	242†	85	72	115	130	124	126
	II	100	289*†	88	63	102			
	III	100*§	288*†	164	237†				
Taurine	I	100	118†	92	60†	67†	67†	76†	96
	II	100	147*†	97	81†	69			
	III	100	162*†	80*†	66*†				
Glycine	I	100	84	67†	57†	69†	66†	71†	70†
	II	100	71†	68†	58†	71†			
	III	100	81	68	63†‡				
Valine	I	100	119†	95†	86	86†	86	86	90
	II	100	124*†	118	106	100			
	III	100	122†	94	83				
Alanine	I	100	127	155†	127	109	118	109	145†
	II	100	138	200	107	100			
	III	100	150	200	100*‡				
Aspartic acid	I	100	106	62†	62†	94	75	81	94
	II	100	94	56	62	56			
	III	100	58*†	45*†	64				
Glutamic acid	I	100	77†	74†	97†	116†	139	149	129
	II	100	86	79	90	103			
	III	100§	55*†	60*†	84*				
Histidine	I	100	97	105	122	132†	136†	226†	239†
	II	100	71	84	96	79			
	III	100	58	65	70				
Urine volumes	I	100	201†	156	123	132	123	114	103
	II	100	378†	219†	186†	186†			
	III	100*§	221†	352	182*†‡				
Body weight	I	96‖	95	85†	79†	81†	82†	87†	92
	II	96‖	95	86†	78†	78†			
	III	96‖	95	84†	78†				

Significantly different from the following at $p < 0.05$ (*9*):
* From group I on the same day. † From day 0 within the group.
‡ Final day of group III from final day of group II.
§ From group II on the same day, calculated for day 0 only.
‖ Due to the 24-hour fast.