

# Design and Statistical Procedures for the Evaluation of an Automatic Gamma-Ray Point-Source Calibrator

S. B. Garfinkel, W. B. Mann,  
and  
W. J. Youden

Institute for Basic Standards, National Bureau of Standards, Washington, D.C.

(December 14, 1965)

A description is given of the mechanical design and operation of an automatic gamma-ray point-source calibrator.

The use of statistical design in experiments for evaluating performance factors, such as interchangeability of stations and run differences using the same data obtained in comparisons of the sources, is described in detail.

Key Words: Statistical experiment design, testing equipment, routine testing, radioactivity standardization measurements, gamma-ray point sources

## 1. Design and Performance of an Automatic Gamma-Ray Point-Source Calibrator

Recently, in response to a need for standards for workers in the field of gamma-ray spectrometry, a gamma-ray "kit" for point-source radioactivity standards has been developed [Hutchinson, 1960]. These sources are prepared from solutions which are standardized either by coincidence counting or, as in the case of cesium-barium-137, by measurements using the NBS calibrated  $4\pi\gamma$ -ionization chamber.

The sources are prepared by depositing either 0.05 or 0.1 ml of the calibrated radioactive solution onto mounts consisting of a 0.006-centimeter-thick polyester tape which is supported by an aluminum annulus (3.8 cm I.D., and 5.4 cm O.D.), as shown in figure 1. As it is desirable for all of these sources to be nominally the same strength and the same size, the solution is dispensed with an ultramicroburet [NBS Circ. 594, Mann and Seliger, 1958]. After drying, the sources are covered with another layer of the same kind of polyester tape. The sources are then intercompared with several accurately standardized sources, for the purpose of individual calibration.

For several years these calibrations were performed manually; that is, the sources were placed, one at a time, in a jig which was held in a fixed position relative to a scintillation counter, and the count rates were intercompared. As part of the program to increase the accuracy of the standards, it was decided to design and construct an automatic sample changer with the goal of attaining source intercomparisons with a precision of the order of 0.1 percent.

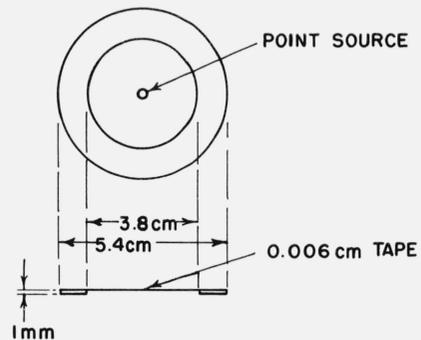


FIGURE 1. Source mount.

The changer is a round turn-table of 1/4-in.-thick aluminum alloy having a diameter of 24 in., with source positions spaced at  $18^\circ$  intervals on the circumference of a circle 20 in. in diameter (fig. 2). These positions have 1-in.-diameter holes in which rigid plastic sample carriers rest. The gamma-ray point sources are held firmly in place on top of the carriers by the pressure of phosphor-bronze springs. There are 20 indexing holes equally spaced around the table as shown in figure 2, the center of each one radially in line with the center of a sample carrier and the center of the table, and 3/8-in. in from the edge of the table. These holes, in conjunction with a solenoid-plunger pin, are used for positioning the sources above the detector.

A shaft which is affixed to the underside center of the table, rests on a steel ball bearing which lies in a conical depression inside a supporting cylinder.

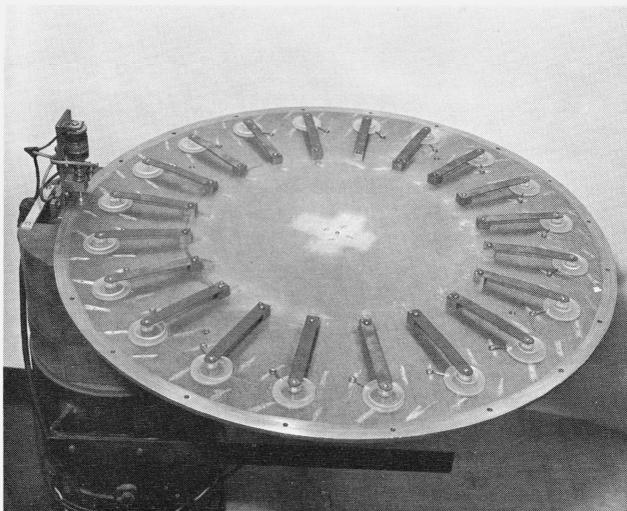


FIGURE 2. Sample changer.

The table is rotated by a 1/100-HP motor and two gears, one of which is fixed on the motor shaft, and coupled to the other gear which is mounted on a concentric spring-loaded friction clutch on the table shaft.

The motor is turned on and off by a miniature switch ( $S_1$ ), which is actuated by the plunger of a solenoid, in the following manner:

At the conclusion of a measurement, while the data are being printed out onto a paper tape, a relay,  $K_1$  (fig. 3) in the recording system is held closed. Capacitor  $C_1$ , which had been charged up during the measurement period now discharges through the coil or relay  $K_2$ , thereby closing it for about 1.5 sec, thus energizing the solenoid. The solenoid-operated plunger is lifted from the indexing hole in the table for this brief period, and mechanically closes the miniature switch ( $S_1$ ), thereby starting the motor, and the table starts to rotate. As it takes about 5 sec for the table to rotate  $18^\circ$ , relay  $K_2$  opens before the next source position is reached, the solenoid is de-energized and the plunger falls back and rests on the surface of the turn table,

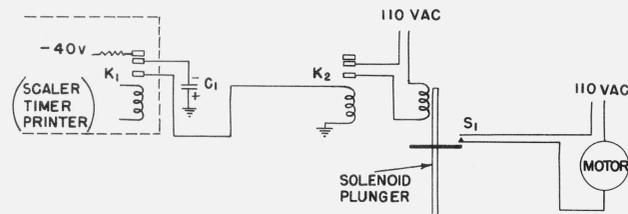


FIGURE 3. Diagram of motor-control circuit.

but as this is not far enough to allow switch ( $S_1$ ) to open, the motor continues to rotate the table. When the next source "arrives" into the counting position, the solenoid plunger falls into the indexing hole, thus stopping table rotation and opening the motor circuit. The purpose of the friction clutch is to allow the motor to slow down gradually after the table has stopped. The time for the sample changing is about 5.0 sec, while the printout takes 20 sec. Thus all changing operations (including the stopping of motor) stop at least 10 sec before the next measurement starts.

Originally, in order to obtain reproducible source-to-detector distance, the table was supported underneath the plunger pin by a roller bearing, and it was assumed that the combination of the spring-loaded plunger pin and the slightly loose fit of the table shaft would ensure this. However, after several series of measurements, it became apparent that sources on some positions of the table were yielding consistently erroneous values. The final design eliminated the effects of any defects in the table which would contribute to errors as a function of vertical displacement.

A lucite block with ramps at each end was affixed to the top of the lead shield, and its dimensions are such that when a source and carrier come into position, they "ride" up the ramp approximately 1.5 mm, so that the carrier is actually free of the table insofar as vertical positioning is concerned (fig. 4). The plastic sample carriers are 0.425-in. thick with a tolerance of  $\pm 0.002$  in. Thus, the source-to-detector distance is inde-

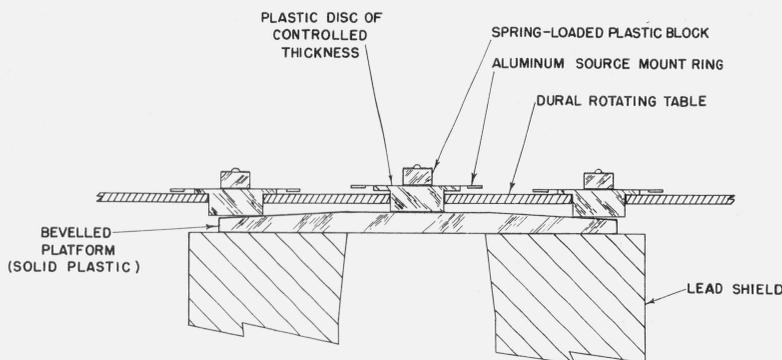


FIGURE 4. Ramp detail.

pendent of variations in the table thickness, and any deviations of flatness of the table. The only function of the table is to bring the sources into position above the detector, the vertical positioning being determined by the phosphor-bronze spring holding the source firmly against its carrier and the latter against the ramp. To get some idea of the reproducibility required in positioning, it should be pointed out that the source is approximately 6 in. from the detector; thus, a change in vertical position of 0.006-in. produces a change of 0.2 percent in the count rate  $\left(n \propto \frac{1}{d^2}, \Delta n \propto \frac{-2 \Delta d}{d}\right)$ .

## 2. Description of Auxiliary Instrumentation

The gamma-ray detector consists of a 3-in. by 3-in. thallium-activated sodium iodide crystal, coupled to a 3-in. electron-multiplier phototube. The associated electronics consist of an amplifier, and gain-stabilization circuit [DeWaard, 1955], which compensates for shift in gain in either the phototube, amplifier, or high voltage supply (this latter being part of the stabilizer). The detector is situated in a lead pig, with walls 1½-in. thick (fig. 5). The aperture at the top of the shield was made small to lessen detection of unscattered gamma radiation from sources adjacent to the source being measured but large enough so that when the table rotates, the detector never "loses direct sight" of a source. Thus, the photopeak is always "present" for continuous operation of the gain-stabilizing circuit. The output from a single-channel

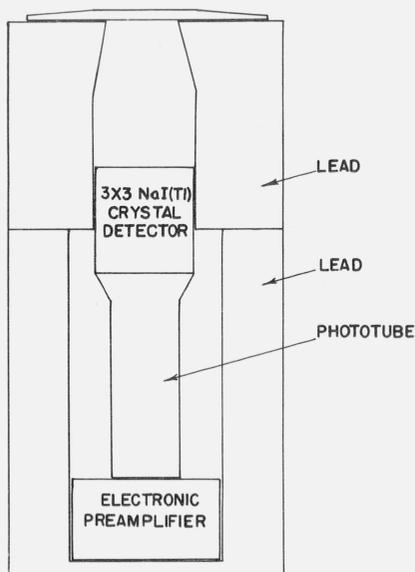


FIGURE 5. Lead pig, showing ramp and detector assembly.

analyzer (which is also part of the stabilizer system) whose window is set on the photopeak is fed into a commercial automatic scaler-timer-printer system. At the end of each source measurement, and after the data are printed, the scaler and timer are automatically reset, and started for the next measurement.

## 3. Background Considerations

The activity of these sources is of the order of  $5 \times 10^4$  disintegrations per second, and they are measured at a distance of about 6 in. from the 3×3-in. detector. No correction is made for the cosmic-ray background, which is of the order of 0.1 percent (or less), as variations in the background affect the ratios of nearly equal sources negligibly.

In the case of the 662-keV gamma-ray of barium-137 m, there is, for example, a relatively large background contribution ( $\sim 4\%$ ) to the photopeak count rate arising from the detection of unscattered gamma rays from the other 19 sources. If, then, there were 19 identical sources, and the twentieth were, say, 1 percent high or low, then, the relative activity of this odd one would be in error by 0.04 percent, if, as the case is, no background corrections are made.

## 4. Performance

In order to assess the stability and reproducibility of the system, two experiments were performed. A cesium-137 source was put onto one of the sample carriers, and over 100 five-minute consecutive readings were taken (with no table rotation), each one consisting of some 200,000 counts. The distribution of the results fitted the expected distribution quite well.

The second investigation involved the placement of 20 sources on the table and determining (a) the relative gamma-ray emission rates of these 20 sources, as well as the bias, if any, of the 20 positions of the table. The statistical design and analyses of these experimental results are given in considerable detail. The interest centers not so much in this particular apparatus as in this type of equipment. There is increasing use of automatic equipment in the routine comparison of specimens.

## 5. Statistical Analysis

Industrial control laboratories and laboratories doing clinical tests are turning increasingly to mechanization of the routine operations involved in the test procedure. Sometimes these operations require the addition of reagents, mixing, and the transfer of material. The last step consists in bringing the prepared material before a testing point where a suitable device evaluates the color, pH, or other property of the specimen. Generally this last stage consists of a device with a number of stations which successively present their specimens to the test point.

For many tests the equivalence of the various stations is clearly satisfactory, provided only that the mechanical clearances are adequate. Should the position of the specimen, as determined by the station, be at all critical it will be necessary to demonstrate that the stations are in fact interchangeable. That is, the particular station occupied by a specimen should not contribute materially to the error in the evaluation of the specimen. Satisfactory interchangeability is desirable—the alternative being to determine suitable corrective factors for the individual stations.

There are three ways to explore experimentally the performance of the individual stations.

One procedure is to transfer the *same* specimen to every station in turn and record the reading for each station. This procedure will run into difficulty if the specimen has to be evaluated immediately, e.g., a color might fade. If the time spent at each station is fairly long, the problem of keeping the evaluating apparatus free from drift has also to be considered.

A second procedure requires the availability of as many *identical* specimens as there are stations, or of specimens which are accurately related to each other.

The above two procedures are classical and straightforward. The third procedure has the interesting feature that the stations can be evaluated while evaluating the regular sequence of specimens encountered in the work of the laboratory. The major requirement is that the specimens be stable. In brief, each specimen is evaluated at a limited number of stations, as few as three or even two stations. Each station will have been occupied by two or three or more different specimens. The values recorded will reflect the net result of the specimen plus the station characteristic. In order to obtain both the specimen values and the station corrections, there must be at least as many observations as the total of specimens and stations. Each observation can be expressed as a function of the unknown values for the specimen and station and the set of equations can be solved. Usually additional observations are made and a least-squares solution obtained. The surplus equations afford an estimate of the experimental error in the observations. This makes it possible to test whether or not the observed differences between the stations exceed experimental errors and to attach an appropriate error to the values calculated for the specimens.

If a special symmetry is used in the assignment of specimens to stations, then improved precision and ease in solving the equations results. There are certain advantageous numbers of stations to place on a wheel because of the combinatorial properties of numbers. A simple case of a wheel with seven stations and seven specimens, A through G, will illustrate the principle involved:

	Station Number						
	1	2	3	4	5	6	7
Run No. a	A	B	C	D	E	F	G
Run No. b	B	C	D	E	F	G	A
Run No. c	D	E	F	G	A	B	C

Thus station 1 is occupied in turn by specimens A, B, and D and specimen A occupies stations 1, 7, and 5 in turn. Inspection shows certain relations have been achieved. The three stations that are occupied by A also encounter the six other members of the complete set B, D, C, G, E, F, of the other specimens. Thus A can always be compared with any other specimen occupying the same station.

Similarly station 1 which encounters the specimens A, B, and D can, by means of these specimens, be directly compared with all the other six stations. Specimen A permits station 1 to be compared with 5 and 7; specimen B compared station 1 with 2 and 6; and specimen D compared station 1 with 3 and 4.

Suppose we wish to evaluate station 1 in terms of the average performance of all seven stations. Let  $A_{1a}$ ,  $B_{1b}$ , etc., represent the observation made on the specimen, A, B, C, etc., in the designated stations and runs. Consider the three observations on specimen A. These observations permit the comparison of station 1 with the average of stations 5 and 7. It is more convenient to multiply by 2 and write:

$$2A_{1a} - A_{7b} - A_{5c} = \Delta_{1, 7, 5}.$$

Similarly  $2B_{1b} - B_{2a} - B_{6c} = \Delta_{1, 2, 6}.$

and  $2D_{1c} - D_{3b} - D_{4a} = \Delta_{1, 3, 4}.$

Each equation is free of any specimen contribution. What about run effects? The run effects, if present, are designated by the letters a, b, and c. Observe that the sum of these three equations involves the subscripts a, b, and c each twice with a negative sign and twice with a positive sign. That is the run effects, if any, neatly cancel out, provided that conditions in each run are constant. We may, therefore, drop the a, b, and c subscripts and treat the differences as differences *between stations*, i.e.,

$$6[1] - [2] - [3] - [4] - [5] - [6] - [7] = \Sigma \Delta$$

where the station numbers are given in the brackets. We may add to this equation the equation

$$[1] - [1] = 0$$

which simply says that station 1 is equal to station 1 (with no error of measurement).

$$7[1] - \{[1] + [2] + \dots + [7]\} = \Sigma \Delta.$$

Dividing by 7

$$[1] - \text{mean of all stations} = \Sigma \Delta / 7$$

$$[1] = \text{mean of all stations} + \Sigma \Delta / 7.$$

Customarily the "mean of all stations" is a number which is the average of all 21 observations. This gives equal weight to every station, every specimen and each run. The  $\Delta$ 's are obtained directly from the observations so that it is a simple matter to calculate a value for each station. These values are completely comparable because the specimen and run effects have been neatly removed making use of the special properties associated with the above triads of letters.

An exactly parallel procedure leads to estimates for each of the seven specimens, estimates that are corrected for any station differences. The simple sum of the seven observations for each run contains the contributions of all specimens and all stations so these sums may be compared directly to detect differences between runs.

If this procedure shows the stations to be satisfactorily equivalent there will be no need to follow any

particular schedule in assigning specimens to stations and no need to make any adjustments. If there are important differences among the stations there is a choice of getting a better wheel or following a suitable scheme of specimen placement that will permit adjustment for station differences.

Clearly, if there are as many specimens as stations, making two runs leads to a unique solution for the differences, but without providing an estimate of the experimental error. In most instances it will be desired to hold the number of runs to three or four because the specimens have to be moved to new stations after each run. Several possible schemes using 3 or 4 runs are listed in table 1. An extensive collection of designs is available in a Bulletin [Bose, Clatworthy, and Shrikhande, 1954].

The example with seven stations just discussed is particularly simple in that any given specimen is

TABLE 1. *Examples of designs useful for intercomparing positions in apparatus*

8 Stations	9 Stations
R5, p. 185 A B C D E F G H B C D E F G H A D E F G H A B C	SR12, p. 143 A B C D E F G H I E C D B I H F G A F A E I G D B C H
10 Stations	13 Stations
T6, p. 231 A B C D E F G H I J B H J A F C E I D G E C D G H A I J B F	C1, p. 250 A B C D E F G H I J K L M C D E F G H I J K L M A B I J K L M A B C D E F G H
15 Stations	16 Stations
T28, p. 237 A B C D E F G H I J K L M N O J O K G F N E L D H A M I B C O G I L M D K G J E N A B H F	LS 14, p. 245 A B C D E F G H I J K L M N O P O P M N K L I J F G H C D A B C L I J K B C D A O P M N F G H E
19 Stations	
S1.1, p. 218 A B C D E F G H I J K L M N O P Q R S C N I A K L F J Q S B M G P H E R O D B Q L E R N A I G F H K P O D J S C M	
12 Stations	14 Stations
R15, p. 188 A B C D E F G H I J K L B C D E F G H I J K L A E F G H I J K L A B C D G H I J K L A B C D E F	R24, p. 192 A B C D E F G H I J K L M N L M N H I J K E F G A B C D J K L M N H I C D E F G A B I J K L M N H B C D E F G A

paired just once with all the other specimens. By "paired" is meant "meets on the same station." This is not true for all the other designs listed in table 1. The arithmetical procedure for computing the estimates for specimens and stations for these designs is given in the above mentioned Bulletin. Above each design in table 1 is given the identification number and page reference where the design is listed in the Bulletin.

Certain of the designs show a simple cyclic displacement of the specimens for the successive runs. The order of the columns (stations) in the designs may be randomized and the rows run in any order without changing the properties of the design.

The apparatus described in this paper uses a wheel with 20 stations. We might use the design for 19 stations and leave one station on the wheel unfilled. An alternative was chosen by using a design for 10 stations and using this design twice. In effect this means two separate and independent sets of data and it was necessary to achieve some way to tie together all 20 stations, which was accomplished by interlacing the stations. First a pair from one design, then a pair from the second design and so on. This spread the two designs evenly over the whole wheel. The assumption was made that the 10 stations assigned to one design would have very closely the same average as the 10 stations assigned to the other design. When each station is rated as a ratio to the average for the set to which it belongs, the 20 ratios would fairly reflect the differences among all the stations.

A wheel with 25 stations could be filled with designs for 10 and 15 stations. By combining designs a wheel of any given number of stations may be accommodated.

The general availability of computers will probably mean that the matrix of equations will be solved with their help. The particular merit of these designs is that the solution can be obtained by inspection; thus consider the design for 10 stations given below:

Station Number									
1	2	3	4	5	6	7	8	9	10
A	B	C	D	E	F	G	H	I	J
B	H	J	A	F	C	E	I	D	G
E	C	D	G	H	A	I	J	B	F

The assignment of the specimens to stations makes it possible to intercompare the specimens without introducing the differences between stations should these be present. Consider specimen A which appears in stations 1, 4, and 6 along with specimens B, E, D, G, F, C. Direct comparisons of A with these six specimens (two at a time) is therefore possible staying within a station. Three other specimens, H, I, and J never share a station with specimen A. The object is to effect comparisons of A with H, I, and J without introducing station differences. We

observe that stations 2 and 5 permit the comparison of H with B, C; E and F. Similarly stations 7 and 9 are used to compare I with B, D; E and G. Finally stations 3 and 10 provide the comparison for J with C, D; F and G. We may combine these three sets of comparisons and obtain the result that H, I, and J as a group may be contrasted with B, C, D, E, F, and G as a group.

It was shown above that stations 1, 4, and 6 provided the station-free comparison of A with B, C, D, E, F, and G as a group. We also have just obtained the station-free contrast of B, C, D, E, F, and G as a group with the group H, I, and J.

Therefore A can be compared with H, I, and J using the group, B, C, D, E, F, G, as an intermediary. Evidently A may be compared with all other specimens using only comparisons made within stations.

We have, therefore, the following comparisons:

$$2A - B - E$$

$$2A - D - G$$

$$2A - C - F$$

and

$$B + C - 2H$$

$$E + F - 2H$$

$$G + E - 2I$$

$$B + D - 2I$$

$$C + D - 2J$$

$$F + G - 2J$$

Note that by multiplying the first three comparisons by 6 and then summing them with the last six comparisons, we have as a result

$$36A - 4(B + C + D + E + F + G + H + I + J).$$

Adding and subtracting 4A gives

$$40A - 4 \text{ (total of all sources).}$$

Dividing by 40 gives A—(average of all 10 sources) in terms of the differences. These operations are shown, for both sources and stations using actual counts, in tables 3 and 4.

Imagine for a moment a perfect wheel, all stations identical, also identical specimens, and identical runs. The 30 observations would then be identical except for experimental error. In an actual experiment each observation may be regarded as undergoing three displacements. The specimen, the station, and the run all combine to effect a net displacement.

The preceding paragraph indicates how to obtain the displacement contributed by specimen A. Using these predicted quantities, i.e., the least square estimates, a matching set of predicted expected values can be obtained for comparison with the actual observations. In fact, the sum of the squares of the 30 discrepancies between observed and predicted values is a measure of the experimental error.

The sum of the squares of the deviations must be divided by (30-1-9-9-2) or 9 to obtain the mean square error. The deductions from 30 refer to the mean, the nine independent specimen constants, the nine independent station constants and two independent run constants. The standard deviation of a single observation is obtained by taking the square root of the mean square error.

In the present experiment a wheel with 20 stations was being used to intercompare sources used as radioactivity standards. There is no suitable standard design for 20 stations with a limited number of interchanges for the sources. Consequently the design for 10 items with three interchanges was used twice. The 20 stations were interlaced by assigning stations 1, 2, 5, 6, 9, 10, 13, 14, 17, 18 to one design and the remaining 10 stations to the other design. This assumes that the averages for the two sets of 10 stations will each be representative of the wheel as a whole. This assumption can be verified when the data become available. All 20 stations can be put on a comparable footing by expressing each station as a percent of the average for the group of 10 to which it belongs; this assumes that the averages of the two groups of 10 stations are the same.

TABLE 2. Counts minus one million for each of the three stations occupied by each source

Station number	Run I		Run II		Run III	
	Source	Count <sup>a</sup>	Source	Count <sup>a</sup>	Source	Count <sup>a</sup>
1	K	42558	B	35323	O	42911
2	L	50654	R	40375	E	42384
5	O	42711	A	37296	L	49092
6	P	37720	K	44580	F	39822
9	A	40622	Q	35730	K	43096
10	B	36471	F	40506	R	41525
13	E	41432	P	40623	A	39876
14	F	39051	E	42361	Q	36443
17	Q	36856	L	49803	B	38637
18	R	41535	O	46438	P	41311
Total exp't 1		409610		413035		415097
3	M	38417	D	37985	T	36523
4	N	35271	M	38203	J	35817
7	T	36910	N	38137	C	38110
8	U	43440	T	39107	G	37974
11	C	38773	J	37635	D	40225
12	D	37316	U	42288	H	37859
15	G	35733	I	36176	N	38121
16	H	37663	G	37996	M	35500
19	I	35491	C	38815	U	42263
20	J	37916	H	40813	I	34940
Total exp't 2		376930		387155		377332

<sup>a</sup> Actual counts diminished by one million.

Twenty sources, identified by letters, were assigned to the 20 stations as shown in table 2. Once the sources were assigned to the stations for the first run, the wheel was started and 5 min counts made at each station giving a count somewhat over 200,000.

Five revolutions of the wheel were made without disturbing the sources. The five revolutions with short stops makes for a more equitable sampling of the background and machine performance during the time required for a run.

At the conclusion of the first run, the sources were transferred to new assigned stations and another five revolutions made. The sources were again shifted for the third run. The station assignments are such as to make possible the intercomparison of any station with the other nine stations in its group without introducing differences between the sources. Counts were recorded for each 5 min period. The five counts were summed and diminished by one million and the remainders entered in table 2. These coded values are all that is needed because the calculations involve differences between the entries in table 2. Naturally the raw data reflect the combined effects of sources and positions. Thus the simple average of the three A counts involves any effects associated with station 9, 5, and 13. Similarly the average of the three counts recorded for station 6 depends on the values for sources P, K, and F. The merit of the design rests in the ease with which the effects associated with individual station and sources can be disentangled.

Tables 3 and 4 show specimen computations for source A and station No. 6 in the first group of 10. The adjustment for a source is made up of quantities obtained by taking differences between sources within the same station. Station effects are therefore not present. Similarly, stations are evaluated by taking differences between stations using the same source, and source effects are thereby eliminated. As a datum, or reference point, the average of all 30 counts is used. The computed adjustments are added or subtracted from this grand average. This gives, on the one hand, adjusted estimates for sources as though there were no differences between wheel stations; and equally adjusted values for stations as though 10 identical sources had been available to compare the stations.

TABLE 3. Calculation of adjustments to observed values for source, using source A as an example

Station	Sources
9	2A - Q - K = 81244 - 35730 - 43096 = 2418
5	2A - L - O = 74592 - 49092 - 42711 = -17211
13	2A - E - P = 79752 - 41432 - 40623 = -2303
	Total = -17096
	-----
	Multiply total by six <sup>a</sup> = -102576
1	K + O - 2B = 42558 + 42911 - 70646 = 14823
17	Q + L - 2B = 36856 + 49803 - 77274 = 9385
14	E + Q - 2F = 42361 + 36443 - 78102 = 702
6	P + K - 2F = 37720 + 44580 - 79644 = 2656
18	O + P - 2R = 46438 + 41311 - 83070 = 4679
2	L + E - 2R = 50654 + 42384 - 80750 = 12288
	-----
	Total below double line = -58043
	Divide by 40 = Adjustment = -1451
	Add grand average of 30 counts = 41258
	Adjusted value for A = 39807

<sup>a</sup> The factor "six" is obtained by inspection to insure that each letter occurs equally often with a minus sign when the summation is made.

TABLE 4. Calculation of adjustments to observed values for stations using station No. 6 as an example

Source	Stations
P	$2[6] - [13] - [18] = 75440 - 40623 - 41311 = -6494$
K	$2[6] - [1] - [9] = 89160 - 42558 - 43096 = 3506$
F	$2[6] - [14] - [10] = 79644 - 39051 - 40506 = 87$
	Total = -2901
	Multiply total by six = -17406
R	$[18] + [10] - 2[2] = 41535 + 41525 - 80750 = 2310$
E	$[13] + [14] - 2[2] = 41432 + 42361 - 84768 = -975$
O	$[18] + [1] - 2[5] = 46438 + 42911 - 85422 = 3927$
A	$[9] + [13] - 2[5] = 40622 + 39876 - 74592 = 5906$
Q	$[9] + [14] - 2[17] = 35730 + 36443 - 73712 = -1539$
B	$[10] + [1] - 2[17] = 36471 + 35323 - 77274 = -5480$
	Total below double line = -13257
	Divide by 40 = Adjustment = -331
	Add grand average of 30 counts = 41258
	Adjusted value for station No. 6 = 40927

TABLE 5. Adjustments to station and source values and comparison with unadjusted values

Station number	Station adjustment	Adjusted value	Unadj. <sup>a</sup> value	Source number	Source adjustment	Adjusted value	Unadj. <sup>b</sup> value
Experiment I							
1	-1608	39650	40264	A	-1451	39807 <sup>d</sup>	39265
2	+461	41719	44471	B	-4341	36917	36810
5	-1774	39484	43033	E	+499	41756	42059
6	-331	40927 <sup>c</sup>	40707	F	-1570	39688	39793
9	-208	41050	39816	K	+2869	44127	43411
10	+555	41813	39501	L	+8785	50043	49850
13	+355	41613	40644	O	+3314	44572	44020
14	+91	41349	39285	P	-1957	39301	39885
17	+733	41991	41765	Q	-5120	36138	36343
18	+1727	42985	43095	R	-1027	40231	41145
Averages		41258	41258			41258	41258
Experiment II							
3	-371	37676	37642	C	+33	38080	38566
4	-1104	36943	36430	D	+557	38604	38509
7	+142	38189	37719	G	-860	37187	37234
8	+1228	39275	40174	H	+1161	39208	38778
11	+1024	39071	38878	I	-2906	35141	35536
12	-940	37107	39154	J	-1169	36878	37123
15	+77	38124	36677	M	+205	38252	37373
16	-1163	36884	37053	N	-576	37471	37176
19	+293	38340	38856	T	-867	37180	37513
20	+814	38861	37890	U	+4423	42470	42664
Averages		38047	38047			38047	38047

<sup>a</sup> The unadjusted value is the average of the three observed counts (table 2) for the station.

<sup>b</sup> The unadjusted value is the average of the three observed counts (table 2) on the source.

<sup>c</sup> Taken from table 4.

<sup>d</sup> Taken from table 3.

No adjustments are required for the run totals because the effects of all 10 sources and all 10 stations are present in every run. Unavoidably every one of the 30 counts is subject to the counting error and any unequalized drifts in background or counting electronics. The adjusted values shown in table 5 are the best estimates of source and station characteristics. We can use these adjusted values, together with the run averages, to compute an ideal table. In table 6 every actual count is replaced by an "ideal" value.

TABLE 6. "Ideal" values calculated using best estimates for stations and sources

Station No.	Run number			Station No.	Run number		
	I	II	III		I	II	III
1	42222	35355	43216	3	37527	38902	36495
2	50207	40738	42469	4	36013	37817	35460
5	42501	38079	48521	7	36968	38282	37908
6	38673	43842	39609	8	43344	39077	38101
9	39302	35976	44171	11	38750	38571	39314
10	37175	40289	41038	12	37310	42199	37954
13	41814	39702	40414	15	36910	35887	37234
14	39482	41893	36481	16	37691	36693	36775
17	36574	50822	37902	19	35080	39042	42449
18	41661	46345	41280	20	37338	40691	35641

The "ideal" values are obtained by combining the calculated adjustment for the station, the source and the run and adding the result to the grand average. The "ideal" value for the count obtained for source K in station 1 in run 1 is obtained by taking from table 5 the station adjustment (-1608); the source adjustment (+2869); the run adjustment (-297). The run adjustment is the difference between the grand average (table 5) and the run 1 average (table 2). The net adjustment, (2869-1608-297) or 964 when added to the grand average, 41258, gives the "ideal" value of 42222 for this observation. The discrepancies between the actual counts and these "ideal" values computed from the best estimates are a measure of the errors involved.

Table 7 lists the differences between the observed counts and the "ideal" values computed using the best estimates for sources, stations and runs. These best estimates impose 21 constraints on the data leaving nine degrees of freedom available for the estimation of error. The two error variances should be compared with the error variances listed in table 8 which were obtained by the computer using unrounded numerical values. The average count is about 1 040 000. Assuming the Poisson distribution the error variance should equal the mean count. Both estimates of error slightly exceed theory but are well within the limits that can be expected for estimates based on just nine degrees of freedom. Evidently the plan of work and equipment gave data which were close to the theoretical Poisson error.

The mean squares shown in table 8 provide the means for judging whether the data provide convincing evidence of differences among the wheel stations. The ratio of the mean square for adjusted positions to the error mean square is the familiar statistic F. This ratio is 2.24 for experiment 1 and 1.44 for experiment 2. Both ratios are less than the 90 percent value (2.44) tables for nine degrees of freedom for both numerator and denominator. The fact that both mean squares do exceed the error mean square does suggest there may be small differences among the stations too small to be conclusively detected in these experiments. If these possible station differences are ignored, there would result some small increase in the error variance associated with the source averages.

TABLE 7. Differences between observed counts and calculated values shown in table 6

Station No.	Experiment 1			Station No.	Experiment 2		
	Run number				Run number		
	I	II	III		I	II	III
1	-336	32	305	3	-890	917	-28
2	-447	363	85	4	742	-386	-357
5	-210	783	-571	7	58	145	-202
6	953	-738	-213	8	-96	-30	127
9	-1320	246	1075	11	-23	936	-911
10	704	-217	-487	12	-6	-89	95
13	382	-921	538	15	1177	-289	-887
14	431	-468	38	16	28	-1303	1275
17	-282	1019	-735	19	-411	227	186
18	126	-93	-31	20	-578	-122	701
Sum of squared differences	10 139 049			10 951 000			
Divide by 9 Error variance	1 126 561			1 216 778			

TABLE 8. Mean squares from analysis of variance

Variance source	Degrees of freedom	Mean square	
		Experiment 1	Experiment 2
Runs	2	768 160	3 353 392
Unadj. stations	9	9 277 905	4 290 006
Adjct. stations	9	2 526 719	1 754 738
Unadj. sources	9	46 227 069	10 549 753
Adjust. sources	9	39 475 883	8 014 484
Error variance	9	1 126 560	1 216 778

TABLE 9. Analysis of variance ignoring stations

Item	Degrees freedom	Mean square	
		Experiment I	Experiment II
Runs	2	768 160	3 352 392
Sources	9	46 227 069	10 549 753
Error	18	1 826 640	1 485 758

In fact if it be assumed that the sources were assigned at random to the stations, the analysis of variance would appear as shown in table 9. The small increase in the error variance results from not correcting for the very small differences between stations.

Another way to make clear the minor contribution to error made by stations is to look at the amount by which the adjusted count for a station differs from the average count for all stations. The "adjusted" counts are adjusted to allow for the fact that different sources were usually in different stations. The differences are shown as percentages in table 10 and plotted in figure 6. The differences are of the order of one tenth of a percent which is quite reasonable for the counts available. The graph gives just a hint of a region of high values and a region of low values.

TABLE 10. Percent by which stations differ from average station

Station and percent				Station and percent			
1	0.154	10	-0.053	3	0.035	12	0.091
2	-.044	13	-.034	4	.106	15	-.007
5	.170	14	-.009	7	-.014	16	-.112
6	.031	17	-.070	8	-.118	19	-.028
9	.020	18	-.166	11	-.099	20	-.078

Further study of the mean squares in table 8 reveals a much larger mean square for sources in experiment I than in experiment II. Source L, which is 0.844 percent above the average of all sources is largely responsible. No other source differs as much as half a percent from the average source. The three largest deviations in experiment I are 0.844, 0.492, and 0.417. In experiment II the three largest deviations are 0.426, 0.279, and 0.113. Apparently experiment I happened to get the sources that deviated most from the average, whereas experiment II got sources that, on the whole, gave somewhat lower counts than those forming experiment I. This state of affairs is plainly revealed in figure 7. This is not to imply great variation among the sources. All but one of the 20 sources fell in the range of 1 035 000 to 1 045 000 for their counts. The unadjusted counts are very similar to the adjusted counts because there was so little difference among the stations. In no case is the difference between observed and adjusted count as much as 1000.

There remains a remark about the mean squares found for runs. If the total exposure time remained the same for each run and the counting apparatus maintained performance, then the mean square for runs should approximate the mean square for error.

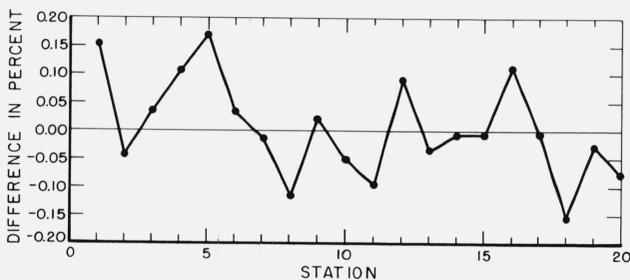


FIGURE 6. Difference of each wheel station from wheel average, expressed in percent.

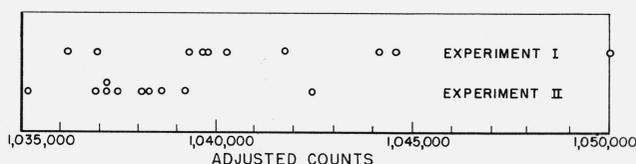


FIGURE 7. Adjusted counts for sources.

The somewhat larger mean square for runs in experiment II is without significance. The mean square would have to exceed the error mean square by a factor larger than four to suggest a real difference between runs.

The use of these "incomplete block" designs is not without a certain price. The original application of these designs was in agricultural field trials. If a large number of varieties of wheat are under comparison it is clear that a block of 20 plots requires a large area of ground. Some of the plots will be at considerable distances from each other and may encounter substantial differences in the soil. Experience showed that comparisons between widely separated plots are subject to greater errors than comparisons between nearby plots. The basic idea back of the incomplete block scheme was to take advantage of the very substantial reduction in experimental error that came from using small blocks. The reduction in error far outweighed the additional mathematics. The indirect comparisons are not as effective as direct comparisons, and therefore result in a lower efficiency. The efficiency of the design used in this work is approximately 70 percent. This may be translated into the following terms. The standard error for the average of three counts with the block design is about that which would be associated with the average of two counts without this design.

In agriculture the sizable reductions in error which resulted from using small blocks outweighed the loss in efficiency. The present experiment affords an interesting example where the reduction in error achieved by eliminating position contributions is relatively slight. On the basis of the error variances given in tables 8 and 9 the variance is increased from 1.17 to 1.66 million when the position effect is left in. Dividing 1.17 by two and 1.66 by three gives  $0.586 \times 10^6$  and  $0.552 \times 10^6$ , respectively, as the variance for the source averages. All this effort would appear to have been to no avail.

One important consequence did come from the use of the design. The design made it possible to evaluate the station effects using the *same* data that were collected to calibrate the sources. Evidence was obtained that the wheel stations are very closely identical. Actually there is no need to take account of wheel stations unless considerably greater counts are taken. In that event the contribution arising from station differences will be relatively more important. It should be pointed out that if the stations had differed by about as much as the source, the precision gained by correcting for station effects would have been impressive. Obviously if stations differed as much as sources, discrimination between the sources becomes impossible. In this event the adjustment for source effects would save the day *provided* a design was used that makes such an adjustment possible.

An exacting test was made of the effectiveness of the numerical adjustments by purposely introducing substantial biases into the wheel stations. Single cardboard shims were placed under the sources (fig. 4)

on five of the 20 stations, so as to increase the source-to-detector distance. Two shims were placed on five other stations, three shims on still another five stations and the remaining five stations were left without shims. The stations were picked at random in allocating the shims. The shims stayed on the stations throughout the experiment.

Twenty sources were placed on the wheel and the same procedure used as before. In this case three revolutions of the wheel constituted a run. The average count per source (and station) *per run* (sources remaining in their stations) was 318391. The average total count per source (and station) for three runs was three times 318391, or 955173.

TABLE 11. Comparison of sources using biased wheel  
Each source and station expressed as a ratio to the average source and station.

Source	Section A				Station No.	Section B		
	2	Biased stations		Diff. Percent		Biased stations		Diff. Percent
		Exp't I	Exp't II			Exp't I	Exp't II	
K	1.0006	1.0000	1.0024	-0.24	1	1.0198	1.0219	-0.21
L	1.0016	0.9998	1.0030	-0.32	2	0.9827	0.9810	.17
O	1.0005	1.0002	0.9964	.38	5	1.0081	1.0087	-0.06
P	0.9997	1.0012	.9986	.26	6	1.0087	1.0055	.32
A	1.0029	1.0073	1.0065	.08	9	0.9795	0.9818	-.23
B	0.9989	1.0000	0.9987	.13	10	1.0064	1.0084	-.20
E	1.0015	0.9996	1.0020	-.24	13	0.9928	0.9905	.23
F	0.9980	.9972	0.9992	-.20	14	1.0064	1.0070	-0.06
Q	1.0005	.9984	.9981	.03	17	1.0210	1.0190	.20
R	1.0014	1.0014	1.0011	.03	18	0.9797	0.9821	-.24
M	0.9987	0.9990	1.0005	-0.15	3	1.0182	1.0173	0.09
N	.9986	.9995	0.9987	.08	4	0.9923	0.9927	-.04
T	1.0029	1.0059	1.0043	.16	7	.9806	.9812	-0.06
U	0.9966	0.9970	0.9966	.04	8	1.0184	1.0165	.19
C	.9964	.9977	.9983	-0.06	11	0.9952	0.9942	.10
D	1.0011	1.0016	.9978	.38	12	.9919	.9924	-.05
G	0.9992	1.0007	.9993	.14	15	1.0173	1.0154	.19
H	.9968	0.9956	.9965	-.09	16	1.0055	1.0076	-.21
I	1.0041	1.0020	1.0035	-.15	19	0.9815	0.9848	-.33
J	1.0000	0.9957	0.9985	-.28	20	.9940	.9920	.20

The above experiment was repeated and the relative values of sources and stations computed. Table 11 lists the results of these computations. The entries in section A of the table show each source as a ratio to the average source and in section B show each station as a ratio to the average station. The difference between the stations with no shims and those with three shims is nearly 4 percent. In spite of these biases introduced into the wheel the adjusted values of the sources (col. 3 and 4) agree with the ratios obtained in another trial using the wheel without shims (col. 2). No adjustments were made for the ratios in column 2, the wheel stations being assumed to be without bias. In fact very slight biases do exist as shown in the preceding study.

The average magnitude of the twenty differences between the paired estimates for the sources is 0.172 percent and for the stations is 0.169 percent. Each estimate is based on about 950 000 counts. As stated earlier, the price of using the experimental design that makes possible the adjustment for the effect of stations, is a certain loss in efficiency. In this case the efficiency is about 70 percent so that the effective

count is  $950\,000 \times 0.70$  or  $665\,000$ . The square root of  $665\,000$  is  $816$ , therefore the *expected* standard deviation of an estimate of a source is  $816/665\,000$  or  $0.123$  percent. The expected *average difference* between two measurements each with standard deviation  $0.123$  is obtained by multiplying by  $2/\sqrt{\pi}$  or  $1.128$ . The theoretical average difference,  $0.123 \times 1.128 = 0.14$  is only slightly less than the experimental average difference.

The good concordance between experiments I and II confirms the error as *calculated* from the statistical analysis on the separate experiments. These errors were  $0.15$  and  $0.14$  percent, respectively. The evaluation of the sources is confirmed by the two experiments and the evaluation of the experimental error is also confirmed by the paired comparisons.

Because sources are compared by taking ratios of counts, the whole statistical analysis was repeated using the logarithms of the observed counts. The analysis of variance and the adjustments in the first

analysis were made using differences rather than ratios, because of the near identities of both sources and stations. The analysis using logarithms did not alter any of the conclusions. Fortunately the counts were large and varied over a very small range. Over this range the logarithms are acceptably proportional to the counts so that the effect of using logarithms was just that of changing units.

## 6. References

- Bose, R. C., Clatworthy, W. H., and Shrikhande, S. S. (1954), Tables of partially balanced designs with two associate classes, North Carolina Agricultural Experiment Station Technical Bulletin No. 107.
- DeWaard, H. (1955), Stabilizing scintillation spectrometers with counting-rate-difference feedback, *Nucleonics* **13**, July, p. 36.
- Hutchinson, J. M. R. H. (1960), Calibration of five gamma-emitting nuclides for emission rate, NBS Tech. Note 71.
- Mann, W. B., and Seliger, H. H. (1958), Preparation, maintenance and application of standards of radioactivity, NBS Circ. 594.

(Paper 70C2-219)