CONF-760533-2

BNL-21212

Nonparametric Analysis of an International Neutron Dosimetry Intercomparison*

Leon J. Goodman
Radiological Research Laboratory, Columbia University
Radiological Research Accelerator Facility
Brookhaven National Laboratory
Upton, New York, 11973, U.S.A.

Introduction

An International Neutron Dosimetry Intercomparison (INDI) was carried out in 1973 at the Radiological Research Accelerator Facility (RARAF) of Brookhaven National Laboratory under the sponsorship of the International Commission on Radiation Units and Measurements (ICRU). Briefly, fourteen groups of scientists, from six countries, performed measurements to determine the separate neutron and photon (x- and gamma-ray) tissue kermas in free air for four energies of monoenergetic neutrons, 15.1, 5.5, 2.1, and 0.67 MeV, and for a source of fission neutrons, ^{252}Cf. For the two highest energies measurements were also made to determine the separate absorbed doses in tissue of neutrons and of photons at three depths in a large water phantom.

Goodman and Colvett (1973, 1974) have described the physical arrangements and the monitoring measurements performed during all measurements. Caswell et al. (1975) and Goodman et al. (1975) have also described INDI and have reported the results obtained. An ICRU report is in preparation.

Independent analyses by Müller (1976) and Goodman (1976) showed that the differences between the reported values could not be ascribed solely to differing factors used to convert instrument response to neutron kerma or absorbed dose, i.e., $\bar{W}$ ratios, kerma ratios, and stopping power ratios.

Nonparametric analyses of tissue-equivalent ionization chamber responses

It is possible to calculate relative responses for most of the dosimeters employed by using data supplied by the participants. Thus, for the same mixed field, the quotients of the responses of the dosimeters by their sensitivities to the gamma rays used for calibration, R'_T and R'_U, respectively, are given by (ICRU, 1976)

$$R'_T = k_T D_N + h_T D_G \qquad (1)$$

$$R'_U = k_U D_N + h_U D_G \qquad (2)$$

*Energy Research and Development Administration Contracts E(11-1)-3243 to the Radiological Research Laboratory, Columbia University and E(30-1)-16 to the Medical Division, Brookhaven National Laboratory, and National Cancer Institute, DHEW Grant No. CA 12536 awarded to the Radiological Research Laboratory, Columbia University supported this investigation.

where D_N and D_G are the kermas or absorbed doses in tissue of neutrons and of photons in the mixed field, k_T and k_U are the ratios of the sensitivities of each dosimeter to neutrons to its sensitivity to the gamma rays used for calibration, and h_T and h_U are the ratios of the sensitivities of each dosimeter to the photons in the mixed field to its sensitivity to the gamma rays used for calibration, respectively. The subscript T refers to the dosimeter having approximately the same sensitivity to neutrons and to photons, and the subscript U refers to the dosimeter having a lower sensitivity to neutrons than to photons.

For homogeneous ionization chambers

$$k_T = \frac{\bar{W}_C}{\bar{W}_N} \frac{(s_{m,g})_C}{(s_{m,g})_N} \frac{K_C}{K_N} \qquad (3)$$

where $\bar{W}$ is the average energy expended in the gas to create an ion pair, $s_{m,g}$ is the effective mass stopping power of the wall relative to the gas, K is the quotient of kerma by fluence for the specified tissue relative to that for the dosimeter material, and the subscripts N and c refer to the neutron field and the gamma-ray calibration field, respectively. For each kerma and absorbed dose the values of k_U, K_N, $\bar{W}_C/\bar{W}_N$, and $(s_{m,g})_N$ were supplied by the participant. To compute the relative responses R_T' and R_U' it was assumed that h_T, h_U, K_c, and $(s_{m,g})_c$ were unity. In contrast to the kermas and absorbed doses, these responses have the important advantage that they do not include variations due to employing differing factors to convert dosimeter response to neutron kerma or absorbed dose.

Calculated relative responses were analyzed by the non-parametric statistical tests described by Youden (1963). This is a method for scoring the performance of laboratories participating in round-robin tests. For each measurement condition the participant with the highest numerical result is given the rank of one, the participant with the next highest result is given the rank of two, and so on until the lowest result is given the lowest rank, P, equal to the number of participants reporting values for the measurement condition. If there are M measurement conditions, then a table of ranks can be constructed having M columns and P rows. This procedure is illustrated as follows:

The relative responses of the tissue-equivalent ionization chambers were calculated from equations (1) and (3) and the values are tabulated in Table 1. These responses were used to assign the ranks shown in Table 2. The row of means gives the mean rank for each column, (P+1)/2. The actual score for each participant is the sum of the ranks received for each measurement condition. When no measurement was reported a dash is entered and disregarded in calculating the score. The mean score is the score that the participant would have received if he had received the mean rank for each measurement condition. The mean score is thus the sum of the mean ranks of each column in which an entry was made for the participant.

If only random differences were involved, the rank a participant received for each condition would be simply a matter of chance, the order of the participants should not persist from condition to condition, and there should be no concordance whatever. Complete concordance exists if each participant always gets the same rank for each of his numerical results. There is then an obvious, systematic order to the results. For finite sets of data it is usually not possible to establish with certainty that concordance does not exist, since even for a set of data chosen completely at random there is always a nonzero probability that some degree of concordance will be manifest. The statistical test establishes the probability that a particular array of ranks could have occurred by chance; however, it is somewhat arbitrary at what level of probability one is willing to reject a set of data on the basis of concordance.

The ranks in Table 2 can be analyzed in two ways. In the first analysis attention is focused on the scores of individual participants. Since the actual score received by a participant will generally differ from the mean score, the question to be answered is: What is the probability that a score occurred by chance rather than due to a pronounced systematic bias? Youden (1963) calculated the approximate upper and lower 5 percent probability limits for ranking scores, as shown in Table 3. The limits are approximate because the scores go by units and it is not possible to have them correspond to the exact 5 percent probability limit. The probability limit refers to the chance of obtaining a round robin with the indicated extreme score. A 5 percent probability limit is reasonably generous since it includes deviant scores for which there is as little as one chance in twenty that the score occurred randomly.

Some judgment must be exercised in applying Table 3 to the scores in Table 2 since all participants did not make measurements for all conditions. The scores for participants D, E, and J are underlined in Table 2 to indicate that they are close to or outside of the approximate 5 percent probability limits. For D and E the comparison to the limit depends on whether nine or ten participants are considered to have reported measurements for the eleven conditions. Participant J reported measurements for only eight conditions, and even for nine participants their score is low.

The scores of D, E, and J may, therefore, be said to display concordance, i.e., their ranks display a systematic difference at about the 5 percent probability limits. This concordance can be corroborated by a second type of analysis involving all participants. The sum, S, of the squared deviations of the actual scores from their corresponding mean scores is calculated, as shown in Table 4. If the ranks depend only on chance, the expected sum of squares, S', is given generally by

$$S' = \sum_i M_i (P_i^3 - P_i)/12 \qquad (4)$$

where P_i participants have been ranked for M_i measurement conditions and the summation includes all measurement conditions. For example, only one condition was measured by all eleven participants, three conditions were measured by ten participants, etc., as shown in the following computation for Table 2:

$$S' = \frac{1[(11)^3-11] + 3[(10)^3-10] + 6[(9)^3-9] + 1[(8)^3-8]}{12}$$

$$S' = 759.5$$

Note that S is greater than S'. Systematic differences spread the actual scores and give a larger sum of squares than S', but even for randomly selected ranks S may exceed S'. The question to be answered now is: What is the probability that the ratio S/S' will exceed some given value by chance rather than due to pronounced systematic differences (concordance)? Table 5 (Youden, 1963) gives the approximate limiting ratios of S/S' for several probability limits. Ratios in excess of the tabulated limits indicate that systematic differences are producing among the rankings some undesired concordance at the indicated probability.

For the ranks of Table 2 we have S/S' = 4.01. Table 5 indicates that even if only nine participants are assumed, there is only a 0.1 percent probability that S/S' will exceed 3.27. That is, there is only one chance in one thousand that the ratio 4.01 occurred by chance. A high degree of concordance is thereby evident.

Inspection of Table 4 shows that participants D, E, and J each contribute from 18 to 20 percent to the sum of the squared deviations. Deleting these participants from the ranking, a second table can be constructed and analyzed. This procedure can be iterated until a table is obtained which shows no concordance at an acceptable level of probability, say 5 percent. This led to Tables 6 and 7 containing six of the original eleven participants. For the ranks in

Table 6, S' = 102 so that S/S' ≈ 1.86. Referring to Table 5 it can be seen that there is no concordance at the 5 percent probability level. Even at the 10 percent level there is no concordance if the number of participants is considered to be less than six due to the absence of 15 out of the 66 possible entries in this array.

There remains the possibility that the five participants deleted from the original table can be arranged in one or more groups which evidence no concordance at an acceptable level of probability. Further analysis showed that the ranks of D and E taken as a group and the ranks of H and K taken as a second group did not show concordance at a probability level greater than 25 percent. The results of participant J cannot be included in any of the previous nonconcordant groups without producing unacceptable concordance, i.e., less than 5 percent probability.

Iterative application of the nonparametric analysis has indicated that four groups of participants could be formed, as shown in Table 8, for which no concordance was evident at an acceptable level of probability. Since acceptable randomness existed within each of these groups, it is reasonable to compute from Table 1 the mean calculated relative responses of each group for each of the eleven measuring conditions. The results are tabulated in Table 9. When these means are ranked, as shown in Table 10, a high level of concordance is indicated by the following observations:

1. Either group 1 or group 2 always received rank 1.
2. Group 4 received the lowest rank for nine of the eleven measuring conditions.

The concordance of the calculated relative tissue-equivalent ionization chamber responses, comparing either single participants or the selected groups, implies that significant systematic differences exist in these responses. The chronology of the measurements (Goodman, 1976) and comparisons of the gamma-ray calibration source ratios (Goodman and Colvett, 1975) show that the groupings in Table 8 cannot be correlated either to a particular measurement session nor to systematic differences between the calibration sources. It is more likely that concordance was due to neglect of or improper application of techniques required to reliably measure the relative responses of some of the ionization chambers. These include

1. Ionization chamber design: correct knowledge of wall and gas composition and dimensions, presence of excessive insulator surface or other inhomogeneity of materials in the gas cavity, a massive central collector and/or excessive

scatter from massive appurtenances for kerma determinations, inadequate circulation of flowing cavity gas, and insufficient electrical guarding of the collector.

2. Irradiation geometry: distance measurements and measurements and corrections for inverse square and chamber wall attenuation, subtended angle, effective chamber center, and cavity displacement in the phantom.

3. Ionization charge: measurements and corrections for collecting efficiency and polarity effects, electrometer system stability and linearity, and stability of collecting potential.

4. Gas density: Measurements and corrections for cavity gas temperature, pressure and flow rate.

Apparently, some of these factors are generally not sufficiently controlled nor uniformly applied to preclude significant systematic differences in determinations of ionization chamber responses.

Nonparametric analyses of other dosimeter responses

Nonparametric analyses were also carried out for the other dosimeter types utilized by at least two participants for at least several measurement conditions. The results are summarized as follows:

Polyethylene-ethylene ionization chambers were employed by only two participants. The ranks of the calculated relative responses showed that there was greater than a 10 percent chance that the values differed randomly, i.e., there was no concordance at an acceptable level of probability.

Graphite-carbon dioxide ionization chambers were employed by four participants. The ranks of the calculated relative responses showed that there was greater than a 25 percent chance that the values differed randomly, i.e., there was no concordance at an acceptable level of probability.

Magnesium-argon ionization chambers were employed by three participants. The ranks of the calculated relative responses showed that the differences could not be distinguished from those of a random population, i.e., no concordance was evident.

Geiger-Müller counters were employed by three participants. The ranks of the calculated relative responses showed that there was greater than a 10 percent chance that the values differed randomly, i.e., there was no concordance at an acceptable level of probability.

Conclusions

Nonparametric analyses of the calculated relative responses of the tissue equivalent ionization chambers showed significant concordance. Apparently, some measurements of and adjustments for ionization chamber design, irradiation geometry, ionization charge and gas density produced systematic differences in the reported values of neutron kerma and absorbed dose that tended to overshadow inconsistencies in the $\bar{W}$ ratios, kerma ratios, and stopping power ratios employed. Therefore, more emphasis is needed on improving the uniformity of the procedures and techniques employed for measuring chamber responses and for applying corrections appropriate to the radiation field.

With regard to the method of analysis used, Youden (1963) points out four advantages of this statistical test:

1. The variance of the scores obtained by ranking results having only random differences is known _a priori_ (Eq. 4).

2. Variations in precision from one participant to another will not affect the test since there is compensation for high and low individual ranks due to poor precision.

3. With ranks, the rejection level can be set in advance of seeing any data.

4. The ranking criterion is intuitively meaningful quite apart from any knowledge of advanced statistical techniques.

References

Caswell, R. S., Goodman, L. J. and Colvett, R. D. (1975). "International intercomparison of neutron dosimetry," page 523 in _Radiation Research: Biomedical, Chemical, and Physical Perspectives_, Proc. Vth Int. Congress of Radiation Research, Seattle, Washington, U.S.A., Nygaard, O. F., Adler, H. I. and Sinclair, W. K., Eds. (Academic Press).

Goodman, L. J. and Colvett, R. D. (1973). "International neutron dosimetry intercomparison," page 23 in _Annual Report on Research Project_, USAEC Report COO-3243-2 (National Technical Information Service, U.S. Department of Commerce, Springfield, Virginia).

Goodman, L. J. and Colvett, R. D. (1974). "International neutron dosimetry intercomparison," page 24 in _Annual Report on Research Project_, USAEC Report COO-3243-3 (National Technical Information Service, U.S. Department of Commerce, Springfield, Virginia).

Goodman, L. J. and Colvett, R. D. (1975). "International neutron dosimetry intercomparison," page 18 in _Annual Report on Research Project_, USERDA Report COO-3243-4 (National Technical Information Service, U.S. Department of Commerce, Springfield, Virginia).

Goodman, L. J., Colvett, R. D. and Caswell, R. S. (1975). "An international neutron dosimetry intercomparison," page 627 in Proc. Second Symp. Neutron Dosimetry in Biology and Medicine, Neuherberg, Germany, EUR-5273, Burger, G. and Ebert, H. G., Eds. (Commission of the European Communities, Luxembourg).

Goodman, L. J. (1976). Private communication

ICRU (1976). Neutron Dosimetry for Biology and Medicine, International Commission on Radiation Units and Measurements, in press.

Müller, J. W. (1976). Private communication.

Youden, W. J. (1963). "Ranking laboratories by round-robin tests," Materials Research and Standards, 3, 1, 9-13. Reprinted in Precision Measurement and Calibration: Statistical Concepts and Procedures, U.S. Department of Commerce, National Bureau of Standards Special Publication 300, 1, Ku, H. H., Ed., 1969.

Table 1. Calculated relative responses of tissue-equivalent ionization chambers

Participant	Measurement Condition										
	1	2	3	4	5	6	7	8	9	10	11
A	4.014	5.896	3.438	1.227	-	-	-	-	-	-	-
B	3.872	5.758	3.383	1.210	4.404	5.927	2.916	0.678	4.912	7.579	5.473
C	3.967	5.893	3.487	1.221	4.465	6.078	2.952	0.700	4.681	7.410	5.322
D	3.942	5.651	3.316	1.172	4.456	6.101	2.960	0.702	4.446	7.389	5.247
E	3.922	5.353	3.163	1.161	4.674	5.926	2.973	0.824	4.487	6.872	5.056
F	4.057	5.724	3.333	1.193	4.588	6.321	3.095	0.747	4.840	7.876	5.258
G	3.909	5.685	3.364	1.188	-	-	-	-	-	-	-
H	4.225	5.887	3.492	1.360	4.751	5.922	2.934	0.734	5.041	8.149	5.739
I	3.950	5.694	3.326	1.183	4.506	6.047	2.982	0.727	4.782	8.005	-
J	4.246	-	-	-	4.902	6.853	3.482	0.945	4.870	7.665	5.351
K	3.995	5.898	3.499	1.225	4.397	6.327	3.106	0.760	4.872	8.048	5.203

Table 2. Ranks of calculated tissue-equivalent ionization chamber responses

Participant	Measurement Condition											Score	
	1	2	3	4	5	6	7	8	9	10	11	Actual	Mean
A	4	2	4	2	-	-	-	-	-	-	-	12	22.5
B	11	5	5	5	8	7	9	9	2	6	2	69	57
C	6	3	3	4	6	5	7	8	7	7	4	60	57
D	8	9	9	9	7	4	6	7	9	8	6	82	57
E	9	10	10	10	3	8	5	2	8	9	8	82	57
F	3	6	7	6	4	3	3	4	5	4	5	50	57
G	10	8	6	7	-	-	-	-	-	-	-	31	22.5
H	2	4	2	1	2	9	8	5	1	1	1	36	57
I	7	7	8	8	5	6	4	6	6	3	-	60	52.5
J	1	-	-	-	1	1	1	1	4	5	3	17	40.5
K	5	1	1	3	9	2	2	3	3	2	7	38	57
Mean	6	5.5	5.5	5.5	5	5	5	5	5	5	4.5		

Table 4. Calculation of the sum of the squared deviations of the actual scores from the mean scores of Table 2

Participant	Actual score-mean score	Squared deviation
A	-10.5	110.25
B	12	144
C	3	9
D	25	625
E	25	625
F	- 7	49
G	8.5	72.25
H	-21	441
I	7.5	56.25
J	-23.5	552.25
K	-19	361
		S=3045

Table 3. Approximate 5 percent probability limits for ranking scores (Youden, 1963)

Number of Participants		Number of Measurement Conditions: 3	4	5	6	7	8	9	10	11	12	13	14	15
3		-	4	5	7	8	10	12	13	15	17	19	20	22
		-	12	15	17	20	22	24	27	29	31	33	36	38
4		-	4	6	8	10	12	14	16	18	20	22	24	26
		-	16	19	22	25	28	31	34	37	40	43	46	49
5		-	5	7	9	11	13	16	18	21	23	26	28	31
		-	19	23	27	31	35	38	42	45	49	52	56	59
6		3	5	7	10	12	15	18	21	23	26	29	32	35
		18	23	28	32	37	41	45	49	54	58	62	66	70
7		3	5	8	11	14	17	20	23	26	29	32	36	39
		21	27	32	37	42	47	52	57	62	67	72	76	81
8		3	6	9	12	15	18	22	25	29	32	36	39	43
		24	30	36	42	48	54	59	65	70	76	81	87	92
9		3	6	9	13	16	20	24	27	31	35	39	43	47
		27	34	41	47	54	60	66	73	79	85	91	97	103
10		4	7	10	14	17	21	26	30	34	38	43	47	51
		29	37	45	52	60	67	73	80	87	94	100	107	114
11		4	7	11	15	19	23	27	32	36	41	46	51	55
		32	41	49	57	65	73	81	88	96	103	110	117	125
12		4	7	11	15	20	24	29	34	39	44	49	54	59
		35	45	54	63	71	80	88	96	104	112	120	128	136
13		4	8	12	16	21	26	31	36	42	47	52	58	63
		38	48	58	68	77	86	95	104	112	121	130	138	147
14		4	8	12	17	22	27	33	38	44	50	56	61	67
		41	52	63	73	83	93	102	112	121	130	139	149	158
15		4	8	13	18	23	29	35	41	47	53	59	65	71
		44	56	67	78	89	99	109	119	129	139	149	159	169

Table 5. Limiting ratios of the calculated sum of squares, S to the expected sum of squares, S', for several probabilities (expanded from a table by Youden, 19

Number of participants	Limiting ratio of S/S' for the stated probability 25%	10%	5%	1%	0.1%
2	1.32	2.71	3.84	6.63	10.83
3	1.39	2.30	3.00	4.61	6.91
4	1.37	2.08	2.60	3.78	5.41
5	1.35	1.94	2.37	3.32	4.62
6	1.33	1.85	2.21	3.02	4.10
7	1.31	1.77	2.10	2.80	3.74
8	1.29	1.72	2.01	2.64	3.47
9	1.28	1.67	1.94	2.51	3.27
10	1.27	1.63	1.88	2.41	3.10
11	1.26	1.60	1.83	2.32	2.96
12	1.25	1.57	1.79	2.25	2.84
13	1.24	1.55	1.75	2.18	2.74
14	1.23	1.52	1.72	2.13	2.66
15	1.22	1.50	1.69	2.08	2.58
16	1.22	1.49	1.67	2.04	2.51
17	1.21	1.47	1.64	2.00	2.45
18	1.21	1.46	1.62	1.97	2.40
19	1.20	1.44	1.60	1.93	2.35
20	1.20	1.43	1.59	1.90	2.31
22	1.19	1.41	1.56	1.85	2.23
25	1.18	1.38	1.52	1.79	2.13
27	1.17	1.37	1.50	1.76	2.08
30	1.16	1.35	1.47	1.71	2.01
41	1.14	1.30	1.39	1.59	1.84
51	1.13	1.26	1.35	1.52	1.73
61	1.12	1.24	1.32	1.47	1.66
71	1.11	1.22	1.29	1.43	1.60
81	1.10	1.21	1.27	1.40	1.56
91	1.10	1.20	1.26	1.38	1.52
101	1.09	1.18	1.24	1.36	1.49

Table 6. Ranks of calculated tissue-equivalent ionization chamber responses after deletion of five selected participants (see text)

Participant	Measurement Condition											Score	
	1	2	3	4	5	6	7	8	9	10	11	Actual	Mean
A	2	1	2	1	-	-	-	-	-	-	-	6	14
B	6	3	3	3	4	4	4	4	1	3	1	36	31
C	3	2	1	2	3	2	3	3	4	4	2	29	31
F	1	4	5	4	1	1	1	1	2	2	3	25	31
G	5	6	4	5	-	-	-	-	-	-	-	20	14
I	4	5	6	6	2	3	2	2	3	1	-	34	29
Mean	3.5	3.5	3.5	3.5	2.5	2.5	2.5	2.5	2.5	2.5	2		

Table 7. Calculation of the sum of the squared deviations of the actual scores from the mean scores of Table 6

Participant	Actual score-mean score	Squared deviation
A	-8	64
B	5	25
C	-2	4
F	-6	36
G	6	36
I	5	25
		S=190

Table 8. Participant groups for which calculated relative ionization chamber responses showed no concordance at an acceptable level of probability

Group	Participants	Probability level at which no concordance was evident
1	J	--
2	H,K	>25 percent
3	A,B,C,F,G,I	10 percent
4	D,E	>25 percent

Table 9. Mean calculated relative ionization chamber responses for participant groups which showed no concordance at an acceptable level of probability

	Measurement condition										
Group	1	2	3	4	5	6	7	8	9	10	11
1	4.246	-	-	-	4.902	6.853	3.482	0.945	4.870	7.665	5.351
2	4.110	5.892	3.496	1.292	4.574	6.124	3.020	0.747	4.956	8.098	5.471
3	3.962	5.775	3.388	1.204	4.491	6.093	2.986	0.713	4.804	7.718	5.351
4	3.932	5.502	3.240	1.166	4.565	6.014	2.966	0.763	4.466	7.130	5.152

Table 10. Ranks of mean calculated relative ionization chamber responses for participant groups which showed no concordance at an acceptable level of probability

	Measurement condition										
Group	1	2	3	4	5	6	7	8	9	10	11
1	1	-	-	-	1	1	1	1	2	3	2.5*
2	2	1	1	1	2	2	2	3	1	1	1
3	3	2	2	2	4	3	3	4	3	2	2.5*
4	4	3	3	3	3	4	4	2	4	4	4

*Due to a tie for rank 2, the rank 2.5 was assigned to maintain the total for this condition at 10.