$$H_1 : \bar{x}_1 > \bar{x}_2$$

$$H_2 : \bar{x}_3 > \bar{x}_4$$

$$H_3 : \bar{x}_{1+2} > \bar{x}_{3+4}$$

Since it was theorized that the position effect (if one existed) may be dependent upon the quality of the product, data were analyzed separately for the control (cc), treated (tt) and the mixed sequence (ct or tc). These are the breakdowns, then, in Table 5. The orthogonal comparisons are shown here for positions 1 vs 2, 3 vs 4 and 1+2 vs 3+4, relating to the mean scores of Table 4.

Probabilistic results from these similar experiments of ham, pork, chicken (white), and chicken (dark) were combined[3] in order to strengthen evidence concerning an effect of position.

Combined evidence for the control (cc) type pairing demonstrated a decrease in preference from position 3 to 4 ($P = .06$) and from the first two positions to the second two positions ($P = .05$). Evidence was not conclusive concerning the decrease in preference exibited from position 1 to 2 ($P = .12$). In the latter case the decrease was 0.22 scale points and was in the hypothesized direction.

Combined evidence for the treated (tt) type pairing did not demonstrate significant positional effects. In the all treated plan the mean for the first two positions was 6.74 contrasted with 6.70 for the last two positions which is reflected by chi-square[6] probability of 0.48. Also the differences between mean preferences, regarding positions 1 vs 2 and 3 vs 4, were not statistically significant.

In the mixed pairing (ct) position 1 was shown to be significantly higher than position 2 ($P = .03$) and the first two positions significantly higher than the last two positions ($P = .03$). Little difference was evidenced between means for positions 3 and 4, however.

A taste testing experiment was reported previously[7] which studied the effect of fatigue over a series of eight samples presented in one sitting. The two foods considered were canned sauerkraut and canned bread with margarine. In a comparison of serving positions 1 or 3 with 5 or 8,

there was found to be no significant difference in preference rating due to the position of the test food, whether it appeared early or late in the eight sample series. These results on the treated samples (no significant difference between positions) would seem to bear out those found on sauerkraut and margarine, since these food items are all relatively low preference items, particularly, the margarine of ten years ago.

Credulence is lent to the theory, then, that while a position effect does appear to exist, the quality of the sample determines whether or not there is a decline in preference with the sequence of presentation.

MAGNITUDE OF ERROR TERMS. A comparison of the magnitude of error terms for the different plans is presented in Table 6. In an analysis of Variance of these variances followed by a Duncan Multiple Range[2] test of means it was determined that the variance for the all control plan was significantly ($P = .05$) smaller than the three plans having both control and treated samples. Also, the error variance of the all treated plan was significantly ($P = .05$) smaller than for the conventional 4! and the plan with two control samples followed by two treated samples (cctt).

These results are in line with anticipations, however, since the ranges of ratings of individuals within the all treated and all control plans are less than for the other plans. Hence, the magnitude of disagreement in preference would be expected to be less for these plans. The differences in magnitude of these variances, though, do point out the necessity for analyzing differences between means, such as those for order, individually, for each plan.

SUMMARY. The preference rating scale is normally used for comparative purposes between samples. However, one can see the effect that might occur on the magnitude of scores of experimental samples, depending upon the quality of standards or controls with which they are compared.

Also the sequence in which relatively good and poor samples were presented made a considerable difference in what conclusions would be drawn. Evidence concerning the effect of the position of presentation was given. For the higher preference samples it was demonstrated that a fatigue effect or sensitivity effect was present causing a decline in the rating of subsequent samples (although not so conclusive between positions 1 and 2 where $P = .12$). For the lower preference samples there was not determined to be a decline in ratings in subsequent positions.

If one desired to minimize Type II statistical error (acceptance of the null hypothesis when it is false), then, one would wish to present test samples after the standards. If the test samples were poorer, the difference would be emphasized by a contrast effect pointed out earlier. The statistical soundness of such a procedure might be questioned, however, since the magnitude of the contrast effect would be expected to be greater when the difference between samples was greater. This would affect the probability statements concerning a "true" difference in the population to a corresponding unknown degree.

REFERENCES

(1) Cochran, G. C. and Cox, G. M. Experimental Designs, John Wiley and Sons, New York, 1950, pp. 103-112.

(2) Duncan, David B. Multiple Range and Multiple F Tests, Biometrics, March 1955, pp. 1-42.

(3) Jones, L. V. and Fiske, D. W. Models for Testing the Significance of Combined Results, The Psychological Bulletin, Vol. 50, No. 5, pp. 375-382.

(4) Kamenetsky, Joe, Contrast and Convergence Effects in Ratings of Foods, Journal of Applied Psychology, Vol. 43, No. 1, 1959.

(5) Peryam, David R., et. al. Food Preferences of Men in the U. S. Armed Forces. Department of the Army, Quartermaster Food and Container Institute for the Armed Forces, January 1960.

(6) Snedcor, G. W. Statistical Methods, The Iowa State College Press, Ames, Iowa, 1946.

(7) Symposium sponsored by the Quartermaster Food and Container Institute for the Armed Forces, Food Acceptance Testing Methodology, October 1953, pp. 92-99.

Figure 1. Preference Rating Card Used for Sensory Evaluations

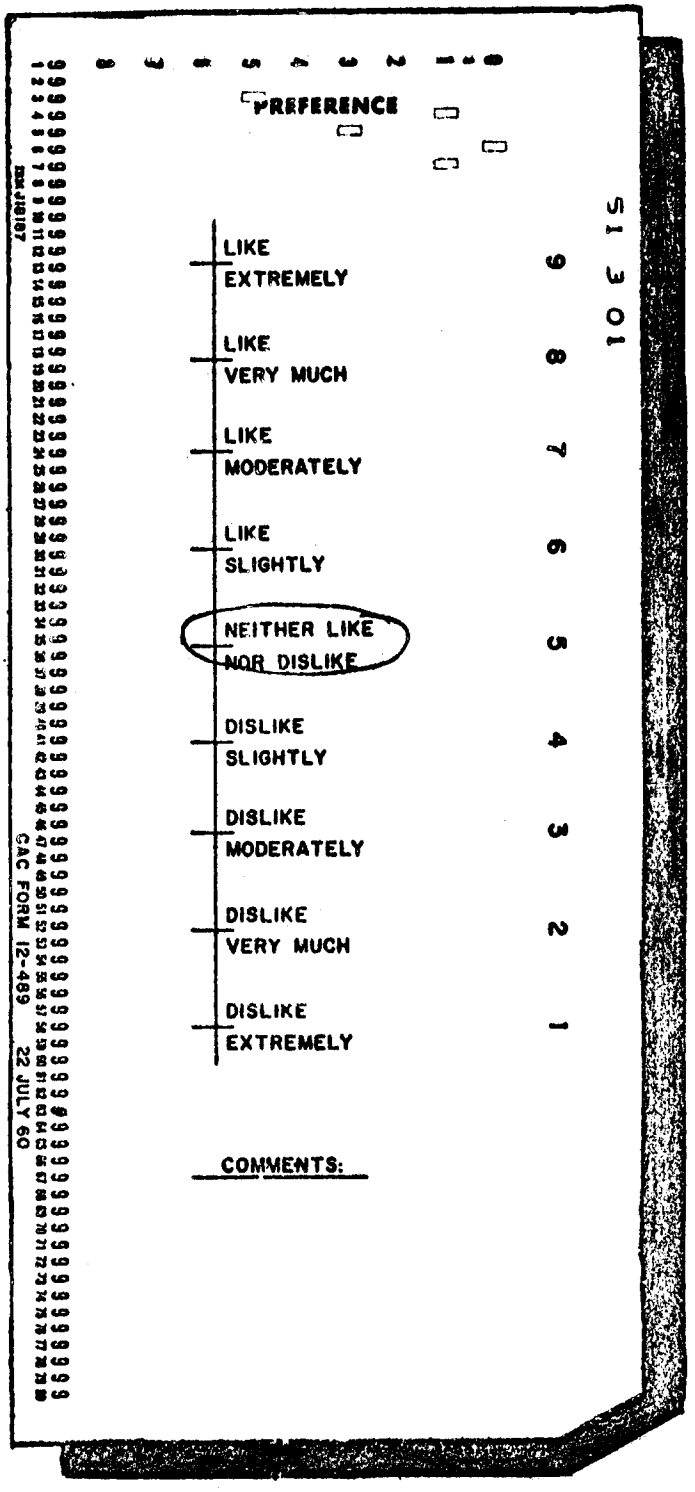


FIGURE 2. Allocation of Samples for the 4! Experimental Plan Utilizing a Latin Square Design

Subject	Order			
	1	2	3	4
1, 5, 9, 13, 17, 21	A	B	C	D
2, 6, 10, 14, 18, 22	D	A	B	C
3, 7, 11, 15, 19, 23	C	D	A	B
4, 8, 12, 16, 20, 24	B	C	D	A

Analysis of Variance Componenets

Source of Variation	Degrees of Freedom
Subjects	23
Orders	3
Treatments	3
Error	66
Total	95

TABLE 1. Mean Preference Ratings of Four Meat Samples in All Possible Serving Orders, Termed a 4! Plan*

Food	Control	Treated	Control - Treated Diff.	Signif. Level for Diff.
Ham	7.17	6.46	0.71	.04
Pork	6.56	5.77	0.79	.01
Chicken, W.	7.33	6.77	0.56	.02
Chicken, D.	6.75	6.06	0.69	.06
Combined			0.69	

TABLE 2. Mean Preference Ratings of Four Meat Samples Where Two Control Samples Were Presented, Followed By Two Treated Samples, Termed a cctt Plan*

Food	1st & 2nd Position Control	3rd & 4th Position Treated	Control - Treated Diff.	Signif. Level for Diff.
Ham	7.40	5.75	1.65	.001
Pork	6.52	5.88	0.64	.02
Chicken, W.	7.14	6.35	0.79	.005
Chicken, D.	6.71	5.81	0.90	.005
Combined			1.00	

TABLE 3. Mean Preference Ratings of Four Meat Samples Where Two Treated Samples Were Presented, Followed By Two Control Samples, Termed an ttcc Plan*

Food	1st & 2nd Position Treated	3rd & 4th Position Control	Control - Treated Diff.	Signif. Level for Diff.
Ham	6.85	6.60	-0.25	n.s.
Pork	5.96	5.92	-0.04	n.s.
Chicken, W.	6.90	7.04	0.14	n.s.
Chicken, D.	5.90	5.83	-0.07	n.s.
Combined			-0.06	

*Individual Means represent ratings of 2 samples by 24 subjects.

TABLE 4. Mean Preference Ratings for Each Serving Order of Four Meat Products for Five Experimental Plans. Individual Means Represent 24 Subjects.

Food	Plan/	Order of Presentation 1st	2nd	3rd	4th
Ham	4!	7.71	6.87	6.21	6.46
	cccc	7.46	7.42	7.37	7.21
	tttt	7.12	7.21	6.83	6.96
	cctt	7.71	7.08	5.50	6.00
	ttcc	6.54	7.17	6.62	6.58
Pork	4!	6.25	5.88	6.38	6.17
	cccc	6.62	6.54	6.79	6.88
	tttt	6.50	6.25	6.21	6.50
	cctt	6.58	6.47	5.75	6.00
	ttcc	5.75	6.17	6.17	5.67
Chicken, W.	4!	7.20	7.04	6.71	7.25
	cccc	7.58	7.46	7.29	6.92
	tttt	6.88	7.04	7.00	6.92
	cctt	7.42	6.88	6.38	6.33
	ttcc	6.83	6.96	7.17	6.92
Chicken, D.	4!	7.12	6.12	6.38	6.00
	cccc	6.83	6.50	6.71	6.33
	tttt	6.50	6.46	6.83	6.33
	cctt	6.67	6.75	5.58	6.04
	ttcc	6.04	5.75	6.08	5.58
Combined	4!	7.07	6.48	6.42	6.47
	cccc	7.12	6.98	7.04	6.84
	tttt	6.75	6.74	6.72	6.68
	cctt	7.10	6.80	5.80	6.09
	ttcc	6.29	6.51	6.51	6.19

TABLE 5. Probability Values* for Orthogonal Comparisons of Position Effect of Mean Preference Ratings in Which Four Meat Samples Were Presented to a Subject

		Pairing	cc		tt		ct
Position	Meat /	Plan	4	2	5	3	1
1 vs 2	Ham		.11	.44	.95	.60	.19
	Pork		.38	.39	.82	.21	.18
	Chicken, W.		.07	.32	.66	.73	.29
	Chicken, D.		.58	.12	.21	.45	.02
	χ^2 Comb'd P		.13	.27	.81	.61	.03
			(.12)		(.82)		
		Plan	5	2	4	3	1
3 vs 4	Ham		.44	.28	.83	.64	.70
	Pork		.14	.62	.75	.82	.25
	Chicken, W.		.22	.08	.44	.38	.96
	Chicken, D.		.09	.09	.87	.08	.22
	χ^2 Comb'd P		.10	.10	.93	.41	.58
			(.05)		(.79)		
		Plan		2		3	1
1+2 vs 3+4	Ham			.22		.15	.008
	Pork			.88		.45	.77
	Chicken, W.			.013		.50	.26
	Chicken, D.			.22		.66	.11
	χ^2 Comb'd P			.06		.48	.03

*Probability Values Presented in This Table Reflect the Test of Significance Regarding the Hypotheses of $\bar{x}_1 > \bar{x}_2$, $\bar{x}_3 > \bar{x}_4$ and $\bar{x}_{1+2} > \bar{x}_{3+4}$.

Combined Results Were Obtained by the Chi-Square Method of Combining Results of Similar Experiments[3]. Individual Probabilities in the Table are Based on an N of 24.

TABLE 6. Error Terms from Five Experimental Plans and Four Meat Products

Plan /	Meat Products					*
	Ham	Pork	Chicken, W.	Chicken, D.	Composite	
41	2.7386	1.9697	1.1203	2.9246	2.1883	
cctt	3.0850	1.6721	1.4882	1.9670	2.0531	
ttcc	1.6356	2.5553	1.1933	1.5490	1.7333	
tttt	1.6360	1.1497	0.8085	1.4420	1.2590	
cccc	0.9134	1.0659	0.7993	0.9171	0.9239	

*Bracketed numbers indicate the variances which are not significantly different at the probability level of 0.05.

AN EVALUATION OF RADIATION-PROCESSED FOODS FOR MILITARY RATIONS*

Donald M. Boyd
Research Analysis Corporation
Bethesda, Maryland

In the fall of 1960 a research group from our organization (then known as ORO) was asked to investigate the possible operational, logistical, and economic advantages to the armed forces of employing radiation-processed foods in the military feeding system and to provide a basis to assist the Army in making decisions on the irradiated-food research program.

The preservation of food by sterilization with ionizing radiation is a relatively new concept and at that time had not been attempted commercially. Experimentally, many foods have been irradiated to determine the value, safety, and efficacy of such processing. Various radiation doses have been employed under different conditions of exposure and of associated treatment techniques. The ultimate goal is the attainment of a process that would safely, and at a reasonable cost, preserve foods so that they could be stored in a fresh-like and wholesome condition for long periods of time without refrigeration. Because meats are the highest-valued items in military rations, special research efforts have been placed on the development of radiation-processed meat items for ration components.

Considerable progress has been made since The Quartermaster General's extensive research program on irradiated foods began in 1953. However, their plans for the construction and operation of a developmental pilot plant were indefinitely suspended on the recommendation of the Director of Research and Development of the Army, in 1959. This action was taken partly because of the uncertainty of the wholesomeness of the foods, but mainly on the basis of need for adequate reliable information concerning the operational, logistical, and economic advantages that would justify the use of irradiated foods in military rations and the construction and operation of a pilot plant. In March 1960, a revised Army program on radiation preservation of foods was approved for wholesomeness studies and fundamental research toward development of

---

*The data presented in this paper have been published in greater detail as ORO-SP-174, "Radiation-Processed Foods as a Component of the Armed Forces Feeding Systems," August 1961.

end items. The Operations Research study, undertaken by RAC, was included in the plan of that new program.

Our investigation began with an analysis of the technological status of radiation-preservation of foods and the additional research effort needed to attain fully acceptable products for incorporation into military rations during the 1965-1975 time frame. Attention was given to various concepts of military feeding and the tactical operational requirements of various armed-forces units for ration support. Then particular consideration was given to the logistical implications involved in the integration of irradiated components in military rations, including savings in manpower, storage, equipment and supplies. The costs of processing, storing and transporting irradiated foods were concurrently studied and our estimates compared with estimated costs of the freeze-dehydration process, and the costs of the commercial canning and freezing processes.

An examination was also made of the different types of radiation sources that could be used for processing foods, and of the availability, costs, and efficiency of these radiation sources.

As a last step, we investigated the feasibility of establishing a mobilization base containing irradiated meats as ration components, and estimated the number and cost of accelerators needed for a production base.

Our findings were published in August of last year.

In order to be approved by the Food and Drug Administration, irradiated foods for public consumption must not show radioactivity levels that are distinguishable from background. Reports have shown that foods can be processed by gamma radiation from Co^{60} or by electron accelerators below energy levels of 10 Mev without inducing measurable radioactivity in the foods. Radiation preservation of meats requires an exposure dose of about 4.5 megarads preceded by short heating to an internal temperature of about 160° F. Reports showed that beef and pork processed this way had remained acceptable for at least 25 months at 70°F storage temperature and for 16 months at 100°F. Bacon and ham had been stored about one year and chicken for one and one-half years with good acceptability.

We found that numerous, extensive studies to determine the wholesomeness of irradiated foods had been conducted, both in-house and

under contract to the Army Quartermaster Corps and the Surgeon General.

Objective analyses of the results of these studies, including long-term animal feeding tests, showed no harmful effects attributable to radiation processing beyond correctable vitamin loss. However, prudently cautious recommendations by the Surgeon General included a few more years of research for completion of the wholesomeness study program.

In the future feeding concept shown in Fig. 1, single meal modules would be employed as follows:

(a) The 25-in-1 uncooked meal might be used behind the contact area. This ration will contain canned, dehydrated, and irradiated foods, and will be served by trained food service personnel in a unit-mess type of feeding.

(b) From the reserve area forward into the contact area the 25-in-1 precooked quick-serve meal would be used where the tactical situation precludes the preparation and serving of the 25-in-1 uncooked meal. This ration will contain precooked freeze-dehydrated foods as the major component, which will be prepared by one or two individuals by pouring hot water directly into the food packages and serving it on disposable trays packed in the carton with the meal. Trained food service personnel would not ordinarily be involved. Under some circumstances precooked irradiated meats might be used in place of freeze-dehydrated meats.

(c) In situations where small groups are dispersed from their units for long periods, the 6-in-1 precooked ready-to-serve meal would be used. It will contain the same type of foods as the 25-in-1 meal.

(d) In the contact area especially and under certain conditions to the rear, the tactical situation will often require the use of an individual ready-to-eat meal. This will contain precooked irradiated foods that will normally be eaten cold but could be warmed by the individual when his situation will allow it. Flexible packaging will add to the value of this ration.

(e) Individuals would also be issued an individual combat food packet for emergency use. This will be a small, compressed, high-caloric-

content food item totalling about 1000 calories. This is not intended to replace a meal when other rations can be provided but is to be capable of sustaining a man for as long as 2 to 10 days under emergency situations without appreciable loss of efficiency and without irreversible physiological damage.

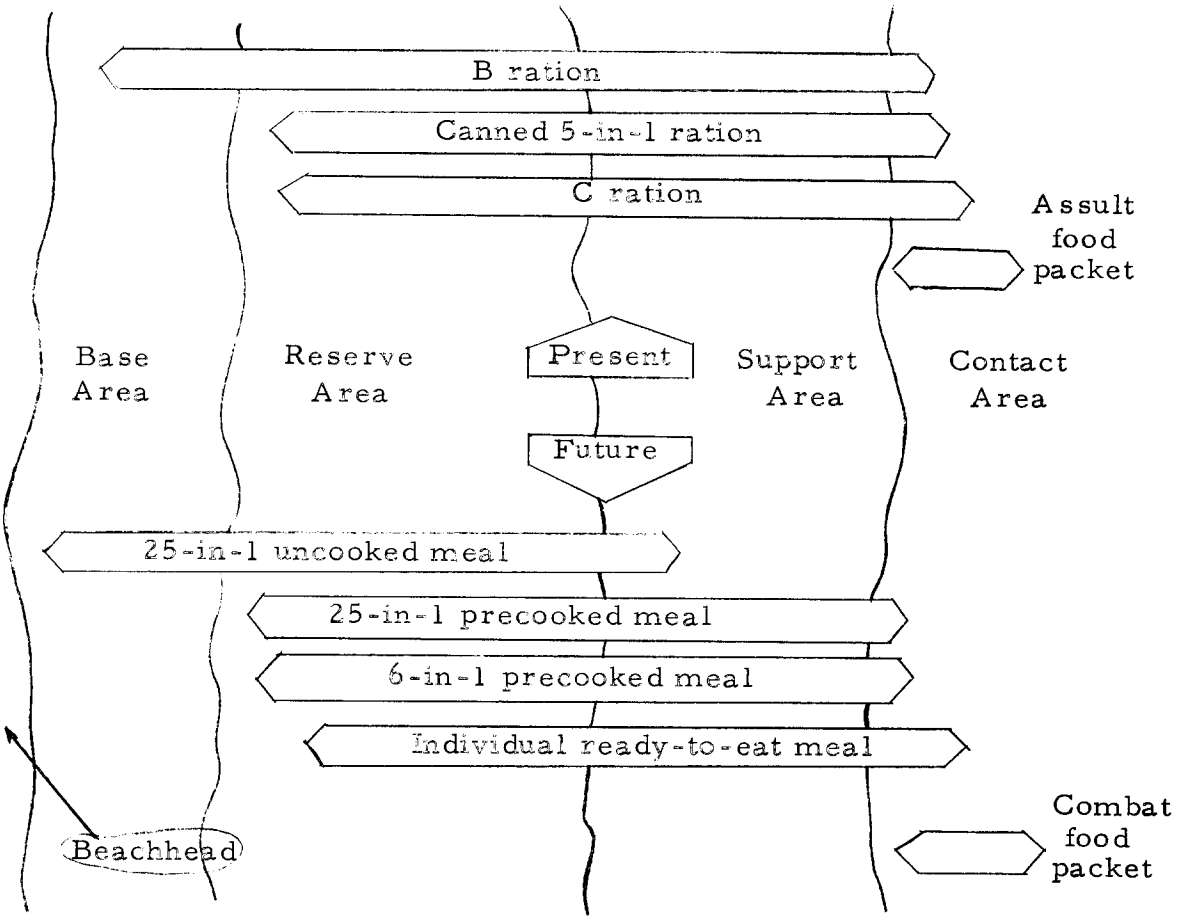


Fig. 1 -- Present and Future Feeding Concepts

With the current emphasis on mobility of US forces vs enemy mass a continuing requirement exists for improved logistical operations to support such mobility. This requirement pervades all classes of supply. Although class I supply involves only a small fraction of the tonnage of class III fuels, simplified rations do contribute to reducing fuel consumption as well as equipment and manpower requirements, and to improving mobility.

In seeking a logistical advantage of employing irradiated rations, we estimated the potential savings these rations would permit if the logistic burden of field kitchens could be eliminated from forward units. An Army of 2 million men was considered, using a distribution of men in a theater based on data from FM 101-10, as shown in Tables I and II.

TABLE I

COMPOSITION OF THEATER SLICE BY ASSIGNMENT

Assignment	Troops
Basic division	13,961
Nondivision	18,540
Theater overhead	24,750
Army	(10,750)
Air Force	(14,000)a
Total	57,251

[a]Two Air Force wings including Army support.

TABLE II

COMPOSITION OF AN OVERSEAS ARMY SLICE

Item	Division	Corps[a]	Army (rounded)[b]
Men			
Armor	14,600	14,600	---
Infantry	13,748	41,244	---
Corps, nondivision	---	74,160[c]	---
Theater overhead	---	99,000[d]	---
Total		229,004	687,000
Field Kitchens			
Armor	98	98	---
Infantry	70	237	---
Total		335	1,005
Bakery companies (mobile)	---	---	5
Refrigeration companies (mobile)	---	---	1

[a]Corps has three infantry divisions and one armored division plus nondivision troops.

[b]Army has three corps.

[c]18,540 x 4

[d]24,750 x 4.

In Table III can be seen the requirements for the operation of field kitchens for 2 million men in a theater of operations. Of particular importance is the fuel consumption rate of over 65,000 tons per year.

TABLE III

REQUIREMENTS FOR FIELD KITCHENS
(For 2 million men overseas)

Item	Quantity
No. of field kitchens	2,925[a]
Men	14,625
Trucks	2,925
Water, millions of gal/year	
Equipment	213[c]
Men	80[d]
Total	293
Fuel, tons/year	
Trucks	4,100[e]
Cooking	56,000[f]
Pump water	5,130[g]
Total	65,230

[a]From Table II (1005 x (2 x 10^6)) /687,000 = 2925 field kitchens.

[b]Based on five men per field kitchen.

[c]Based on 200 gal/day/kitchen for washing and rinsing mess trays plus all else not consumed directly by men.

[d]Based on 15 gal/man/day. Truck water requirements are small in comparison.

[e]Based on 2000 miles/year/truck, average fuel consumption of 5 miles/gal, and 7.0 lb/gal.

[f]Based on 15 gal/day/kitchen.

[g]Based on 1 gal of fuel required to pump 200 gal of water at the water source.

The requirements for mobile bakery companies are shown in Table IV. These companies bake bread for troops in the field, but may also be

assigned to supplement the production of garrison bakery units as the situation demands. During overseas use these bakeries require over 8,000 tons per year of fuel per 2 million men.

TABLE IV

REQUIREMENTS FOR MOBILE BAKERY COMPANIES
(for 2 million men overseas)

Item	Quantity
No. of bakery companies	15[a]
Men	2130[b]
Trucks	315[c]
Water, millions of gal/year	
Equipment	39[d]
Men	12[e]
Total	51
Fuel, tons/year	
Trucks	2200[f]
Ovens	1150[g]
Electric generators	3950[h]
Pump water	890[i]
Total	8190

[a]From Table II $(5 \times (2 \times 10^6))/687,000 = 14.6$ bakery companies.

[b]Based on 142 men/company.

[c]Based on 21 trucks/company.

[d]Based on 100 gal/hr/platoon and three platoons per company.

[e]Based on 15 gal/man/day. Truck water requirements are small in comparison.

[f]Based on 10,000 miles/year/truck, average fuel consumption of 5 miles/gal, and 7.0 lb/gal.

[g]Based on 10 gal/day/oven and six ovens per company.

[h]Based on three generators (25 Kw each) per company running continuously; specific fuel consumption = 0.6 lb/hp-hr.

[i]Based on 1 gal of fuel required to pump 200 gal of water at the water source.

Mobile refrigeration companies deliver perishable foods from depots to supply points, but many also use their semitrailer vans as fixed refrigerators as the situation demands. During use in the theater of operations, these companies require fuel for gasoline-driven refrigerant compressors and trucks. These requirements are illustrated in Table V.

TABLE V

REQUIREMENTS FOR MOBILE REFRIGERATION COMPANIES
(For 2 million men overseas)

Item	Quantity
No. of refrigeration companies	3[a]
Men	564[b]
Trucks	165[c]
Water for men, millions of gal/year	3[d]
Fuel, tons/year	
Trucks	1,150[e]
Refrigeration equipment	10,600[f]
Pump water	50[g]
Total	11,800

[a]From Table II (1 x (2 x 10^6))/687,000 = 2.9 refrigeration companies.

[b]Based on 188 men/company.

[c]Based on 55 trucks/company, of which 48 are 7 1/2-ton semi-trailers and 7 are 2 1/2-ton trucks.

[d]Based on 15 gal/man/day. Truck water requirements are small in comparison.

[e]Based on 10,000 mile/year/truck, average fuel consumption of 5 miles/gal, and 7.0 lb./gal for both truck types.

[f]Based on 5-hp motors, specific fuel consumption of 0.6 lb/hp-hr, and 7000 hr/year of operation.

[g]Based on 1 gal of fuel required to pump 200 gal of water at the water source.

If the use of irradiated foods and freeze-dehydrated foods would permit the elimination of refrigerated warehouses, a substantial saving in the overseas logistical effort shown in Table VI could be attained.

TABLE VI

REQUIREMENTS FOR REFRIGERATED WAREHOUSES
(For 2 million men overseas,
warehouse size, 20 by 100 ft)

Item	Quantity
No. of refrigerated warehouses	173
Men	519[a]
Water, millions of gal/year	
Equipment	6
Men	3[b]
Total	9
Fuel, tons/year	
Electricity generation	12,700
Pump water	100[c]
Total	12,810

[a]Based on three men per warehouse (ORO estimate).

[b]Based on 15 gal/man/day.

[c]Based on 1 gal of fuel required to pump 200 gal of water at the water source.

Table VII summarizes those requirements for kitchens, bakeries, and refrigeration facilities which could be reduced from the total logistical

effort by employment of irradiated foods and dehydrated foods in military rations. Net savings that could be attained in fuel alone are shown in Table VIII. The delivery of bulk fuel by truck is one of the largest problems in theater logistics.

TABLE VII

REQUIREMENTS FOR KITCHENS, BAKERIES, AND REFRIGERATION FACILITIES
(For 2 million men overseas)

Facility	Men	Trucks	Water, millions of gal/year	Fuel tons/year
Field kitchens	14,625	2925	293	65,230
Bakery companies	2,130	315	51	8,190
Refrigeration companies	564	165	3	11,800
Refrigerated warehouses	519	0	9	12,800

TABLE VIII

NET FUEL SAVINGS
(For 2 million men overseas)

Item	Fuels saved, tons/year
Field kitchens	65,230
Bakery companies	8,190
Refrigeration companies	10,600[a]
Refrigerated warehouses	12,700[b]
Total for irradiated foods	96,720[c]
Preparation of freeze-dehydrated food	-5,080[d]
Total for freeze-dehydrated foods	91,640[e]

[a]Refrigerating equipment only.

[b]Refrigerating equipment only. All electricity generation is assumed to be for refrigeration purposes.

[c]Fuel for electricity for truck shops is less than 0.5 percent of this total.

[d]Fuel consumed in heating of water for freeze-dehydrated foods.

[e]Corresponds to 250 tons/day and is not sensitive to the assumed percentage of men who may eat freeze-dehydrated food regularly. For example, 50 percent freeze-dehydrated food in a theater yields 96,720 - 10,160 = 86,560 tons/year, which corresponds to 237 tons/day.

Our cost analysis for processing and transporting irradiated foods showed that these costs were similar and competitive to those of freeze-dehydrated foods and foods preserved by other commercial means.

Our analysis of costs of operation of electron accelerators included the result that the cost of these machines per kilowatt of output power decreases with the increase in the power rating of the accelerator. In addition, at any given radiation dose the output in terms of food processed is proportional. When we applied our operational costs analyses to the problem of establishing a mobilization base, we found that seven 100-Kw accelerators or twelve 30-Kw accelerators would be required to process the total annual meat ration requirements per 1 million men (Fig. 2). The costs of processing this amount of meat is shown in Table IX.

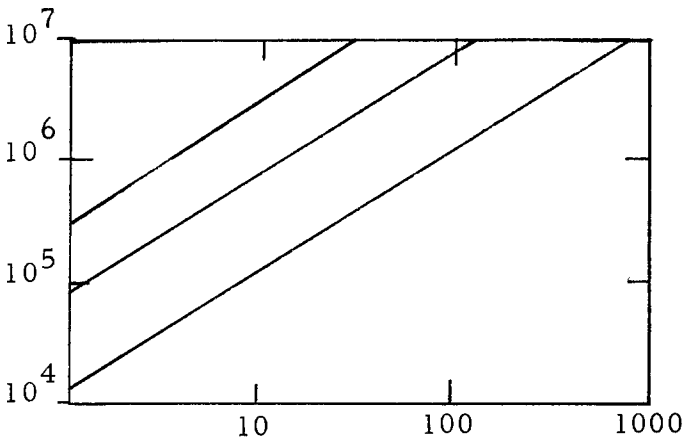


Fig. 2 -- Number of Accelerators Needed to Treat Meats for Armed-Forces Personnel

TABLE IX

ANNUAL CAPITAL AND OPERATING COSTS FOR MEAT PROCESSING
(In millions of dollars)

Manpower level	First Year		Subsequent Years[a]	
	30 Kw	100 Kw	30 Kw	100 Kw
10^4 men	0.039 - 0.057	0.016 - 0.029	0.020	0.010
10^6 men	3.9 - 5.7	1.6 - 2.9	2.0	1.0
10^7 men	39.0 - 57.0	16.0 - 29.0	20.0	10.0

[a]No amortization, operating costs only.

The QMC generally stocks meat reserves in a 15-month supply (composed of a 12-month operational reserve plus a 3-month safety reserve). We determined that a mobilization base reserve of 15 months' rations containing irradiated meat components could be obtained by establishing a production base and operating for one year nine 100-Kw electron accelerators per million men supplied.

About 2 years of total lead time would be required to establish the production base and produce the 15-month reserve supply.

It is likely that only 50% of the meat components of future rations in this system would contain irradiated meats; the rest being processed by other means. Under this consideration the electron accelerators required for the reserve supply would be reduced to 5.

The results of our studies permitted us to make certain conclusions, some of which were as follows:

CONCLUSIONS.

1. The use of rations containing irradiated foods instead of B and C rations and the elimination of field kitchens in general war could result in logistical savings equivalent to 97,000 tons of fuel/year/2 million men in the theater of operations.

2. The logistical savings gained by employing only dehydrated foods instead of B and C rations would be equivalent to 91,000 tons of fuel/year/2 million men.

3. In 1965-1975 irradiated foods could have a distinctive advantage over all other types of foods in providing an operationally suitable individual combat meal that would be well received by fighting men.

4. The estimated cost of radiation processing of foods would be competitive with the costs of the thermal-canning, freezing, and freeze-dehydration processes.

5. About 2 years would be required to obtain a mobilization base composed of a 15-month reserve supply of rations with irradiated meats comprising 50 percent of the meat components. This time includes the estimated 9 to 12 months required to establish radiation facilities and the 12 months required to process the rations.

6. To process the 15-month supply of rations, five 100-Kw accelerators/million men would be required at a cost of $1 million to $1.8 million.

A CRITIQUE OF THE EVIDENCE RELATING DIET AND CORONARY HEART DISEASE

George V. Mann
Division of Nutrition, Vanderbilt University
School of Medicine
Nashville, Tennessee

It may be useful for me to review the problem of coronary heart disease (CHD) from the special viewpoint of a nutritionist. While this view may have some prejudice it seems relevant because of the frequent association of diet with CHD and the widespread lay interest in the problem.

Coronary heart disease may seem to have risen like an epidemic among us. It is a complicated task to determine whether this rise to prominence is real or only made apparent by changing techniques. It would be an interesting task for someone to relate the time course of the prevalence of CHD to the marketing of electrocardiographs. To my knowledge this has not been done. One might have expected a rise of CHD when the ECG became available for diagnosis. Dr. Lew of the Metropolitan Life Insurance Company has shown a remarkable explanation for the distribution by states of coronary heart disease in the United States (1)(2), (Fig. I and II). It must be clear that we see what we look for.

A more subtle influence is that of "competing causes" (3). Even when age specific rates are considered we may be baffled in understanding the entire effect of the removal of diseases which typically kill at an earlier age than does coronary heart disease.

The errors of reporting cause of death are well known (4). It is less than even money that an autopsy will confirm the clinical impression and only a small proportion of all deaths are followed by necropsy. Since reporting causes of death, like ladies hats, tends to change with fashion it is easily possible for the mortality rates to be strongly influenced by the current fashion and this is conveniently done since the selection of the first cause in the presence of multiple causes of death will determine the final tabulations.

Finally we must concede that it is possible that an apparent rise of prevalence of coronary heart disease is real and that this is a reflection

of the introduction of a new and potent causal factor that we must identify and adjust in order to control CHD.

The interest of nutritionists in this problem like their interest in most diseases stems from the ancient judgement that a man may be sick because of "something he et". This explanation has proved so attractive that we have a second epidemic, a scourge of nutritionists. These newcomers, coupled with the food industries, have made food and feeding a highly complicated and even dangerous business.

The essential series of hypotheses upon which most research is presently based may be shown as follows:

Diet ^{1} Hypercholesteremia ^{2} Atherosclerosis ^{3} Clinical Events

The evidence to support the first relationship is at best indecisive. The question was brought to prominence by A. Keys (5) who based this contention on a curious selection of food-mortality data of the World Health Organization (Fig. III). Aside from the fact that the hypothesis is based on tenuous population data that might as easily be explained in other ways (6) it has proven impossible to show in retrospective studies that persons with CHD eat differently than those without (Fig. IV).

The dietary behavior of 983 persons in the Framingham Study has been measured by Georgiana Pearson in the past four years (7)(8)(9). The reproducibility of the method whether by one person (Fig. V) or by a second observer (Fig. VI) is good. We are confident that these people were well classified but we can find no relationship between either cholesterol level (Table 1) or experience with CHD and the way these people eat. Morris, Marr and Heady (10) have found no diet-cholesterol disease relationship in their population of bank clerks (11). Their method of measuring diet does not reproduce quite as well as ours. (Table 2).

The entire problem is complicated by the prevailing imprecision of the measurement of cholesterol. Consider, for example, the data of Rivin (12) (Table 3) who compares hospital and commercial laboratories. We have compared several methods applied in a research setting (13) (Table 4). If one adds to this technical variation the considerable biological variation of serum cholesterol with time (14) it is clear that the

central element of the hypothesis may be so badly estimated that this disqualifies our most convenient index (Table 5).

We are at least as bad off in measuring atherosclerosis, the anatomical lesion we believe to be the basis for the clinical disorders. We cannot visualize these lesions in life and even after death to do more than make qualitative descriptions is difficult. You can appreciate that an element of probability determines whether the plaque is critically placed in the cardiovascular system.

The clinical manifestations of CHD are varied (Table 6). A disturbing number, disturbing at least for the biometrician, are completely occult events called "silent coronaries" because they do not cause important clinical signs. The cerebral events, strokes, are even more obscure because we have less precise ways to determine and localize these, having no equivalent for the ECG.

There are several prospective dietary studies under way which propose to change the experience with CHD by altering the diet. The dietary regimens of some of these are summarized in Table 7. The most ambitious of these called the National Diet Heart Disease Study is directed by Dr. Irvine Page and sponsored by the National Heart Institute (15). It is now in the feasibility phase, that is, the determination of whether families can be recruited, supplied with suitable food and kept under surveillance while consuming the diet for the measurement of cholesteremia and the evaluation of cardiovascular disease status. If proven feasible, this experiment will be extended to larger numbers in order to answer the critical question--will dietary changes modify the course of CHD?

The smaller trials of diet, for example, that of Dayton at Los Angeles (16) and Rinzler with the Anti-Coronary Club (17) in New York have usually obtained about a 15% reduction of serum cholesterol in the best circumstances, that is, when the starting level is high. However, many subjects who do follow the diet do not respond and some who respond initially drift back up with time. We must conclude that dietary treatment, if effective, is a relatively impotent agent. We must conclude also that diet has been overemphasized as a cause of CHD and that dietary modifications are proving relatively ineffectual control measures.

Table 1

DIETARY INTAKES - AMERICANS 1957-58

FRAMINGHAM HEART STUDY

ARRANGED BY SERUM CHOLESTEROL LEVEL

Men	N	Calories	Fat g.	Protein g.	Chol. mg.
High Cholesterol	17	3127	149	113	703
Low Cholesterol	39	3487	163	126	721
Random Sample	133	3333	157	122	735

Table 2

COMPARISON OF REPEATABILITY FOR 2 METHODS OF MEASURING DIET

Nutrient	Heady - Bank Clerks 1 week's weighed intake	Framingham research diet history
	r - consecutive weeks	r - 2 year interval
Calories	0.80	0.92
Protein (gm)	0.67	0.72
COH (gm)	0.84	0.90
Fat (gm)	0.79	0.88

Table 3

Cholesterol Measurement

Rivin, et al., J. A. M. A., 166:2108, 1959

Values in Mgm%

Serum	Author	V. A.	Univ.	Commercial 1.	Commercial 2.	Commercial 3.
A	529	479	480	598	513	411
	487	418	541	500	451	318
B	260	240	255	291	273	183
	273	233	296	263	272	191
C	218	213	275	312	255	180
	249	220	288	252	246	172
Method	K-S	K-S	B1	B1	PSG	Sheftel

Table 4

Evaluation of Methods for Serum Cholesterol

		Methods							
		Abell		FeSac		Pearson		Sackett	
Level (mgm%)	N	$\bar{X}$	T. E.	$\bar{X}$	T. E.	$\bar{X}$	T. E.	$\bar{X}$	T. E.
<210	15	188	6.4	180	5.6	179	7.7	237	3.2
211-274	15	232	4.1	228	5.7	225	8.8	288	3.2
275-499	14	368	5.2	384	7.6	340	15.1	452	12.9
>499	15	672	7.7	667	10.4	643	26.9	837	14.2
all levels	59	365	6.0	365	7.6	347	17.1	454	14.2

Table 5

Serum Cholesterol Variation

68 men - measured twice weekly - 10 weeks

S_T = total variation S_E = laboratory variation S_B = biological variable

where $NS_B^2 = NS_T^2 - 1/2\, NS_E^2$

$\overline{X}S_T = 20$

$\overline{X}S_E = 7$

$\overline{X}S_B = 13$

Then: For 95% assurance of effect 2 x 20 = 40 mgm % minimum change.

Table 6

THE MANIFESTATIONS OF CORONARY HEART DISEASE

Of 100 Men with "Events"

30 drop dead

20 are "silent"

10 die a little later

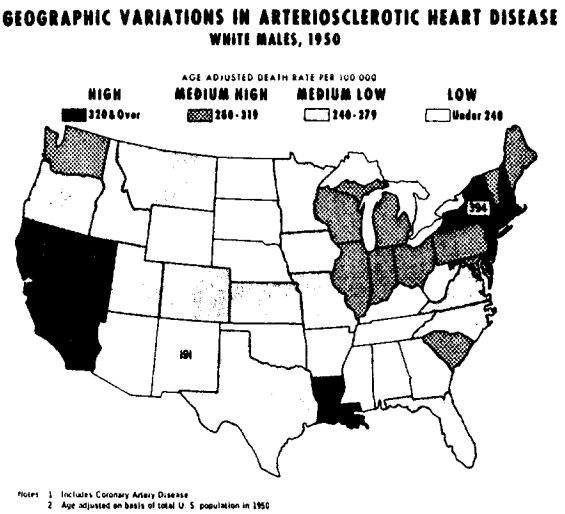
40 recover

Table 7

DIETARY REGIMENS -- OBSERVED AND FAT RESTRICTED

	Framingham (Men)	Page	Rinzler	Dayton	American Heart
Calories	3075	2000	2400	2430	2800
Protein (gm)	112	70	140	94	85
Fat (gm)	154	90	81	106	75
% Cal.	45	41	32	40	36
Cholesterol (mgm)	705	<200	200	380	200
PUS/S	0.3	1.5	1.0	1.7	1.1

Figure I



Data of Enterline and Stewart, Reference 1.

Figure II

CORRELATION BETWEEN MORTALITY FROM ARTERIOSCLEROTIC HEART DISEASE AND INTERNISTS PER 100,000 WHITE PERSONS

United States 1950

Region	Age-adjusted Death Rate per 100,000	Number of Internists** per 100,000
Middle Atlantic	273	12.6
New England	250	10.5
Pacific Coast	233	9.7
East North Central	214	7.3
South Atlantic*	193	7.6
West North Central	183	6.7
Mountain	179	6.5
West South Central	176	5.8
East South Central	160	4.5

* Excludes District of Columbia
** Includes cardiologists

Includes Coronary Heart Disease. Death rates age adjusted on basis of total U.S. population in 1950.

This material was published by E. A. Lew - Reference 2.

Figure III

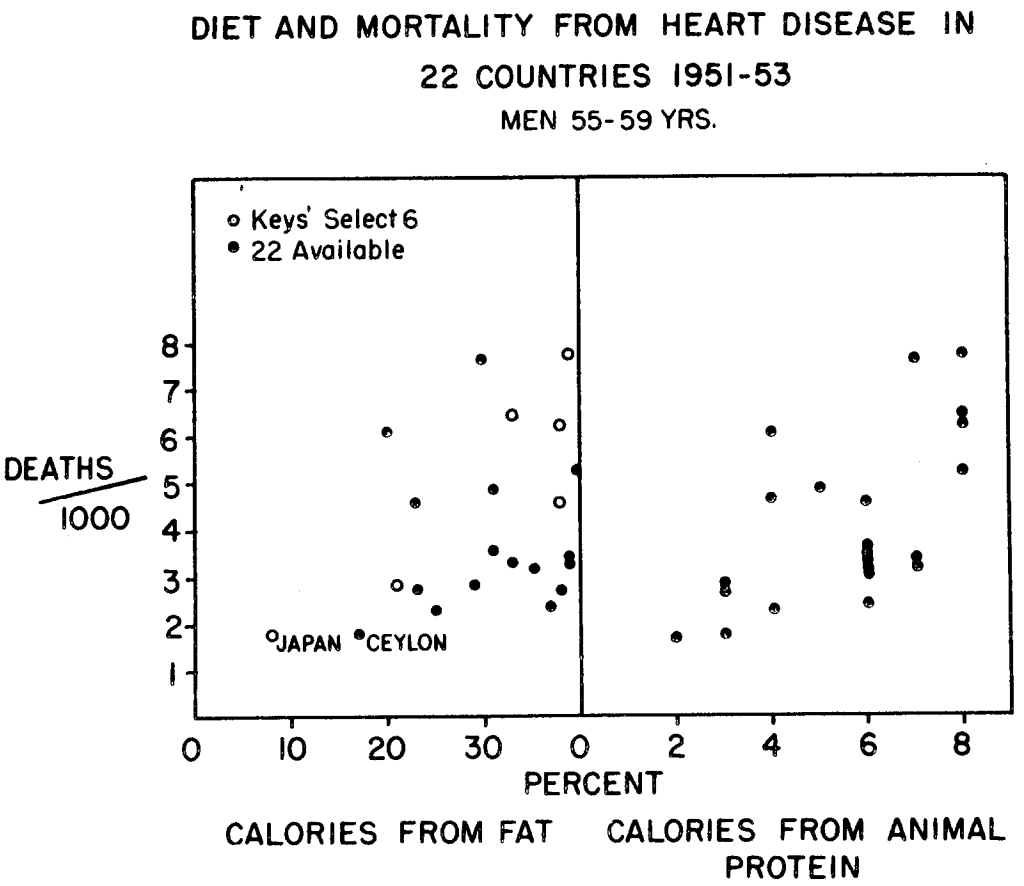

DIET AND MORTALITY FROM HEART DISEASE IN
22 COUNTRIES 1951-53
MEN 55-59 YRS.
o Keys' Select 6
• 22 Available
DEATHS / 1000
8
7
6
5
4
3
2
1
oJAPAN •CEYLON
0
10
20
30
0
2
4
6
8
PERCENT
CALORIES FROM FAT
CALORIES FROM ANIMAL PROTEIN

Figure IV

OBSERVATION OF DIET PATTERN AND EXPERIENCE WITH CORONARY HEART DISEASE

Observer		Association Observed
Wilkinson	Michigan	no
Rosenman	California	no
Zukel	North Dakota	no
Mann	Massachusetts	no
Morris	London	no
Keys	Minnesota	?

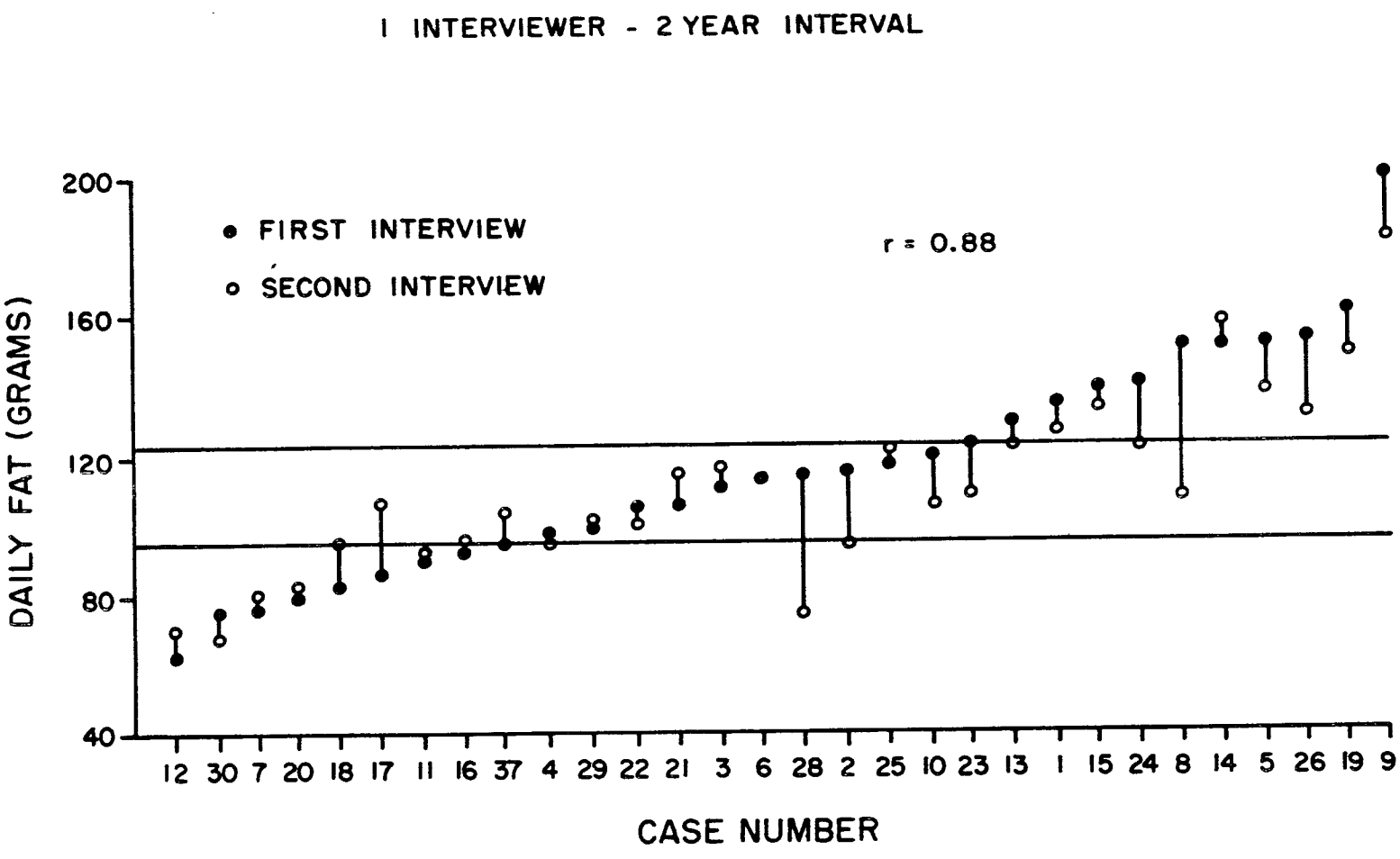
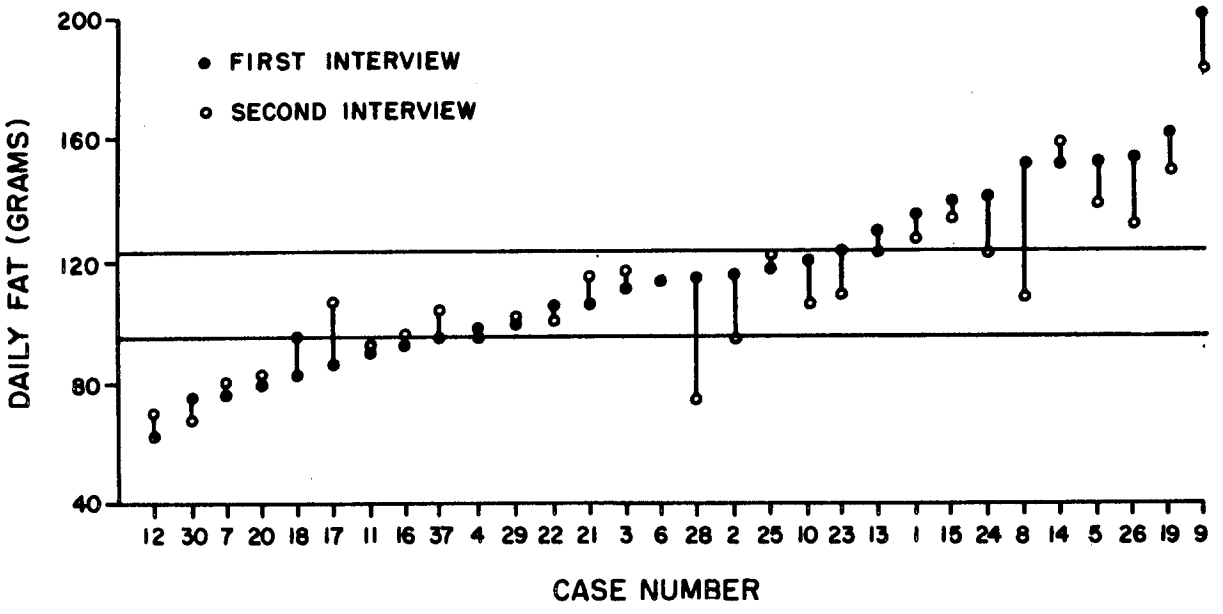

1 INTERVIEWER - 2 YEAR INTERVAL
FIRST INTERVIEW
SECOND INTERVIEW
r = 0.88
DAILY FAT (GRAMS)
200
160
120
80
40
12 30 7 20 18 17 11 16 37 4 29 22 21 3 6 28 2 25 10 23 13 1 15 24 8 14 5 26 19 9
CASE NUMBER

Figure V

Figure VI


200
160
120
80
40
DAILY FAT (GRAMS)
• FIRST INTERVIEW
o SECOND INTERVIEW
12 30 7 20 18 17 11 16 37 4 29 22 21 3 6 28 2 25 10 23 13 1 15 24 8 14 5 26 19 9
CASE NUMBER

BIBLIOGRAPHY

1. Enterline, D. E. and W. H. Stewart, Geographic Patterns in Deaths from Coronary Heart Disease, Publ. Health Rep. 71 849 (1958).

2. Lew, E. A., Some Implications of Mortality Statistics Relating to Coronary Artery Disease, Jour. Chron. Dis. 6 192 (1954).

3. Cornfield, J., Estimation of the Probability of Developing a Disease in the Presence of Competing Causes, Am. J. Publ. Health 57 601 (1957).

4. James, G., R. E. Dalton and A. S. Heslin, Accuracy of Cause-of-Death Statements on Death Certificates, Publ. Health Rep. 70 39 (1955).

5. Keys, A., The Diet and the Development of Coronary Heart Disease, Jour. Chron. Dis. 4 364 (1956).

6. Yerushalmy, J. and H. E. Hilleboe, Fat in the Diet and Mortality From Heart Disease. A Methodologic Note., New York J. Med. 57 2343 (1957).

7. Mann, G. V., G. Pearson, T. Gordon and T. R. Dawber, Diet and Cardiovascular Disease in the Framingham Study I. Measurement of Dietary Intake, Amer. Jour. Clin. Nutr. 11 200 (1962).

8. Dawber, T. R., G. Pearson, P. Anderson, G. V. Mann, W. B. Kannel, D. Shurtleff and P. McNamara, Dietary Assessment in the Epidemiologic Study of Coronary Heart Disease: The Framingham Study II. Reliability of Measurement, Amer. Jour. Clin. Nutr. 11 226 (1962).

9. Pearson, G., G. V. Mann and T. R. Dawber, Diet and Cardiovascular Disease in the Framingham Study III. Food Intakes, Amer. Jour. Clin. Nutr. (in press).

10. Marr, J. W., A Technique of Carrying Out Individual Weighed Diet Surveys, Proc. Nutr. Soc. 20 39 (1961).

11. Morris, J. N., J. W. Marr, J. A. Heady, G. L. Mills and T. R. E. Pilkington, Diet and Plasma Cholesterol in 99 Bank Men, Brit. Med Jour. i 571 (1963).

12. Rivin, A. U., J. Yoshino, M. Schickman and O. A. Schjeide, Serum Cholesterol Measurement: Hazards in Clinical Interpretations J. A. M. A. 166 2108 (1958).

13. Mann, G. V., A Method for Measurement of Cholesterol in Blood Serum, Clin. Chem. 7 275 (1961).

14. Watkin, D. M., E. Y. Lawry, G. V. Mann and M. Halperin, A Study of Serum Beta Lipoprotein and Total Cholesterol Variability and its Relation to Age and Serum Level in Adult Human Subjects, Jour. Clin. Invest. 33 875 (1954).

15. Page, I. H., The National Diet-Heart Disease Study, J. A. M.A. (in press).

16. Dayton, S., M. L. Pearce, S. Hashimoto, L. J. Fakler, E. Hiscock and W. J. Dixon, A Controlled Clinical Trial of a Diet High in Unsaturated Fat, New Engl. Jour. Med. 266 1017 (1962).

17. Rinzler, S., Nutrition in Relation to Heart Disease. The Anti-Coronary Club of New York., Arch. Environ. Health 5 60 (1962).

SOME CONSEQUENCES OF SOME ASSUMPTIONS WITH RESPECT TO THE PHYSICAL DECAY OF A CHAMBER AEROSOL CLOUD*

Theodore W. Horner*
Project Statistician, Booz-Allen Applied Research, Inc.,
4815 Rugby Avenue, Bethesda, Maryland

The usual piece of equipment for studying the survival characteristics of organisms suspended in an atmosphere is a gas-tight chamber controlled with respect to relative humidity and temperature. The mathematical formulation of the behavior of aerosol clouds injected into these chambers and the viability of organisms contained in the particles of these clouds are of great interest to aerobiologists. This paper is concerned with some of the consequences of a particular set of assumptions with respect to the physical decay of chamber aerosol clouds. In presenting the material, I will first touch on those aspects of chambers and aerosol clouds that must be taken into consideration in mathematical formulations. Biological recovery curves will be touched on next. A discussion of relationships among parameters associated with the physical recovery of the cloud will follow -- hitting first the mathematical characterization of the assumptions, then the mathematical relationships among the parameters and finally by means of slides, the relationships will be pointed up visually. The paper will conclude with a short discussion of possible applications of the work and an indication of further work that remains to be done.

Chambers vary enormously in size. Usually they are cylindrical in shape, being oriented either horizontally or vertically. Occasionally they may have a spherical or some other type of shape. The chamber may or may not be revolving. In using a chamber the procedure is to disseminate an aerosol cloud from a liquid slurry containing viable organisms into the chamber.

The aerosol cloud is composed of liquid droplets and the disseminating device produces a spray from the liquid slurry which is similar to the ordinary nose spray used to fight the common cold. Immediately after dissemination, the suspended particles start disappearing from the chamber due to gravitational fallout and impingement on the sides of the chamber. Since

---

* Report on work under Task 2 (Biological Aerosol Decay) or Contract No. DA-18-064-CML-2810 with the Program Coordination Office at Fort Detrick, Frederick, Maryland.

the larger particles fall more rapidly than the smaller particles, the distribution of particle sizes in the aerosol cloud changes with time. The usual assumption is that the particle number distribution is log normal immediately after dissemination. The fraction of those particles with radii between r and $r + dr$ is $f(r)\,dr$ where $f(r)$ is the frequency density function of the particle number distribution. Because of the differential fallout of the various sized particles, the particle number distribution does not remain log normal. The usual expression for differential fallout is

$$h(r, t) = \exp(-Kr^2 t)$$

where $h(r, t)$ is the fraction of those particles with radii between r and $r + dr$ that remain suspended at time t of those suspended initially. K is a constant that depends on chamber dimensions, gravitational acceleration and other factors. This formula which traces to Stokes law was first derived for stirred stationary chambers by Boyd*. Later Calder** showed that a similar formula held for revolving chambers.

Immediately after dissemination the aerosol particles undergo an equilibration process with respect to their moisture content and the chamber atmosphere. This process ordinarily is accomplished in about a second and so it is convenient to refer to time zero as that instant at which the equilibration process is completed. Both the equilibration and the dissemination process are quite drastic events in the life of an organism and so it is not surprising that many organisms which were viable in the slurry are dead at time zero. The organisms continue to die after time zero. The percentage of those organisms which were viable in the slurry, which remain viable at time zero is known as the initial recovery percentage.

The biological recovery percentage is the ratio, expressed in percentage form, of the number of viable suspended organisms at time t to the number of suspended organisms at time t. In this definition only the

---

* Boyd, Charles A., "The Theory of Sedimentation and Decay of Aerosols", Interim Report BLIR-7, Fort Detrick, July 1952.

** Calder, Kenneth L., "Some Theoretical Aspects of the Rotating Drum, Aerosol Chamber", BWL Technical Note 13, 1958.

organisms which were viable in the slurry are considered. We will represent the biological recovery percentage as B(t) and thus

$$B(t) = 100 \left\{ \frac{\text{Number of viable and suspended organisms at time } t}{\text{Number of suspended organisms at time } t} \right\}$$

Characteristics of the biological recovery curve B(t) as it varies with chamber size and shape, relative humidity, temperature, organism and slurry additives are of great interest to investigators studying the viability of organisms suspended in an atmosphere. The typical biological recovery curve when plotted versus time on semilog paper is concave upward. An estimate of the biological recovery percentage at time t depends on data from a sample of the aerosol cloud withdrawn from the chamber at time t.

There appear to be a number of empirical mathematical expressions that do an excellent job of fitting biological recovery data. These expressions will often explain 99.5 per cent of data variability. The expressions generally have no theoretical basis and give rise to differing consequences. Thus inferences based on these empirical curves are always suspect.

As a step toward deriving biological recovery curves from a more fundamental foundation, BAARINC suggested some time back the heterogeneous initial recovery model. The model postulates that the concave upward curvature of semilog plots is due to nothing more complicated than:

(1) Distribution of particle sizes,

(2) Differential fallout of the various sized particles, and

(3) Higher initial recovery percentages for organisms contained in the larger particles.

These three assumptions are sufficient to generate the type of curvature normally observed. There appears to be no question about the first two postulates. The validity of the third remains to be proven, although it does appear to be quite reasonable to the aerobiologists with whom I have been in contact.

Essentially the heterogeneous initial recovery model states that the biological recovery percentage at time t is a weighted average of the biological recoveries associated with the various sized particles. The weights are the fractions of the suspended organisms that are contained in the various sized particles. These weights continuously change with time. The mathematical formula for the biological recovery percentage at time t is

$$(1) \qquad B(t) = \bar{B}(r, t) = \frac{\int_0^\infty Nf(r)\, h(r, t)\, ar^s \; B(r, t)\, dr}{\int_0^\infty Nf(r)\, h(r, t)\, ar^s\, dr}$$

where B(r, t) is the biological recovery percentage for organisms contained in particles with radii between r and r + dr. The number of organisms contained in a particle of radius r is assumed to be proportional to the radius raised to the s^{th} power. Some work by Dr. William C. Day* at Fort Detrick tends to indicate that s may be different from 3.

The weights mentioned a few moments ago are indicated by the trapezoid drawn in the equation above. To point up the logic of these weights, we let N be the number of suspended particles at time zero. Nf(r) dr is then the number of suspended particles with radii between r and r + dr at time zero. Multiplication of Nf(r) dr by h(r, t) yields the number of suspended particles at time t with radii between r and r + dr. Further multiplication by ar^s yields the number of suspended organisms. Hence the denominator of (1) is the number of suspended organisms at time t and the trapezoid ratio is the fraction of suspended organisms contained in particles with radii between r and r + dr.

From equation (1), it is evident that characteristics of the biological recovery curve are intimately tied to the physical aspects of the cloud. In any case an understanding of these physical aspects must precede attempts to ascertain the validity of the heterogeneous initial recovery model.

*This work is described by Horner in a Biomathematics Analysis Note. Horner, Theodore W., Fort Detrick, Maryland, Biomathematics Analysis 5082, "A Relationship Between Spore Number and Particle Size", September 14, 1961.

What are reasonable assumptions and what are their consequences. It appears reasonable to assume that f(r) is log normal and h(r, t) is of the form exp $(-Kr^2t)$. Further the mass of a particle, say m(r) is probably proportional to the cube of the particle radius.

To check the validity of these assumptions and to estimate the relevant parameters is not easy for three reasons:

(1) The particle number distribution does not remain log normal. Part of the present investigation was designed to gain an understanding of the extent of this non-log normality.

(2) Chambers cannot be sampled at time zero and hence estimates of the parameters of the log normal distribution at time zero must be obtained by indirect means based on data collected after time zero.

(3) The physical recovery fraction of the cloud involves still another factor; namely, the mass of the particle.

The physical recovery fraction, normalized to 100 per cent recovery at time zero, is given by the formula

$$R(t) = \frac{\int_0^\infty Nf(r)\, h(r,\, t)\, br^3\, dr}{\int_0^\infty Nf(r)\, br^3\, dr} \tag{2}$$

The mass of a particle is $\bar{m}(r) = br^3$. The total mass suspended at time t and time zero respectively is given by the numerator and denominator of (2).

Following the assumptions made earlier, the recovery fraction is a function of the mean (u), the variance σ^2 of the initial particle number distribution and the chamber constant K. Thus

$$R(t) = R(t;\, u,\, \sigma,\, K).$$

In gaining information about the physical recovery curve one can, on the basis of aerosol samples, do several things. Thus you can:

(a) Estimate the physical recovery fraction at time t.

(b) Estimate characteristics of the particle number distribution such as the mean and the variance of $Y = \ell n\ r$. At time zero, y would be a normally distributed variable.

Estimates of $R(t \mid u, \sigma, K)$, $E(y \mid u, \sigma, K, t)$ and $Var(y \mid u, \sigma, K, t)$ can be obtained from the data of aerosol samples, where u and σ are parameters of the log normal distribution at time zero. Knowing these three quantities, it would be desirable to know u, σ, and K. This would provide a basis for checking the validity of the theory concerning K and the log normal, $\exp(Kr^2t)$ system in general.

Our work has led to several equations that should be useful in this connection. As mentioned earlier, y was defined as $y = \ell n\ r$. The variables v and q will now be defined as

$$v = (1/\sigma)(y - u)$$

and

$$q = -(1/2)\ \ell n\ Kt - u. \tag{3}$$

Using these definitions, the physical recovery fraction at time t can be written as

$$R(t \mid u, \sigma, K) = R(q, \sigma)$$

where

$$R(q, \sigma) = (1/\sqrt{2\pi}) \int_{-\infty}^{\infty} \exp\left[-(1/2)\ v^2 - e^{-2(q - 3\sigma^2 - v\sigma)}\right] dv \tag{4}$$

Thus initially, the tabulation of R would have required the four quantities t, u, σ, and K to be taken into account. The q formula relates t, K and u and thus tabulation becomes simpler in that R needs to be computed only as a function of q and σ. One of the slides a little later will show the relationships among R, q, and σ^2. An approximation formula

has also been developed for computing q as a function of the physical recovery percentage and σ^2. This formula is useful in generating a starting value of q for iteration procedures for the solution of q given R and σ. The approximation formula is

$$(5) \qquad q = -(1/2)\ \ell n\,(-\ell nR) + \left\{ A + B\ \ell n\left[-\ell n\,(1 - R)\right]\right\} \sigma^2$$

where A and B are appropriate constants.

Using this formula suppose q is calculated given R and σ^2. If the approximation q is now used to calculate R using equation (5), the calculated R will not differ from the original R by more than O.O2 for the range of σ values pertinent to the present investigation.

Some additional equations are listed below.

$$(6) \qquad E(y \mid u,\ \sigma,\ K,\ t) = u + \sigma E(v \mid q,\ \sigma)$$

$$(7) \qquad Var(y \mid u,\ \sigma,\ K,\ t) = \sigma^2\ Var\,(v \mid q,\ \sigma)$$

$$(8) \qquad R(t,\ u,\ t,\ K) = R(q, \sigma)$$

where the frequency density of v is

$$m(v) = \frac{(1/\sqrt{2\pi})\ \exp\left[-(1/2)\ v^2 - e^{-2(q - v\sigma)}\right] dv}{(1/\sqrt{2\pi})\int_{-\infty}^{\infty} \exp\left[-(1/2)\ v^2 - e^{-2(q - v\sigma)}\right] dv}$$

Let us look at equations (7) and (8). These are two simultaneous equations. Having estimates of Var y and R(t) available at time t, one can solve for estimates of q and σ. Estimates of q and σ make possible an estimate of $\sigma E(v \mid q,\ \sigma)$. This latter estimate, when coupled with an estimate of

$E(y \mid u, \sigma, K, t)$, makes possible an estimate of u. Equation (3) can be solved for the chamber constant K. This constant is a function of q and u. Since estimates of q and u are available, it is now possible to estimate K. Thus from estimates of q and u are available, it is now possible to estimate K. Thus, from estimates of R, E_y and Var y at time t, one can determine u, σ, and K.

The relationships among these quantities can best be pointed up by means of graphs. For the graphs to be meaningful, it is necessary to make appropriate choices for the parameters u and σ o the log normal distribution of particle sizes at time zero. For u, the range from O. 5 to 5 μ seems appropriate in the subject matter area to which this investigation is related. Similarly, an upper bound for σ appears to be in the neighborhood of 1. 5. To arrive at this latter number we define two radii r_1 and r_2 such that 5O and 84.13 per cent of the particles at time zero have radii less than r_1 and r_2, respectively. Under these assumptions,

$$\ell n\, r_2 = u + \sigma$$

and

$$\ell n\, r_1 = u$$

Hence

$$\sigma = \ell n\, (r_2/r_1)$$

and

$$r_2/r_1 = e^{\sigma}$$

When σ is 1. 5, the ratio of the two radii is 4.4817. Thus the radius associated with the 84.13 per cent point is 4.48 times the radius for the 5O per cent point when σ = 1. 5. Thus σ = 1. 5 appears to be a reasonable upper bound for the subject matter under investigation. In developing tables, the values of σ shown below were used.

σ	r_2/r_1
O.4O547	1.O
O.69315	2.O
1.O9861	3.O
1.38629	4.O

Figures 1, 2, 3, and 4 [figures are at the end of this article] illustrate various parameter relationships. Figure 1 shows the relationships among the physical recovery fraction, q and σ^2. Each line is associated with a different physical recovery fraction. These lines are almost but not quite straight; the curvature is most pronounced in the lines associated with the lower physical recoveries.

Figure 2 shows the relationships between q, R, and Var y, each line being associated with a different physical recovery fraction. Again these lines are almost straight; the greatest curvature being associated with the lowest physical recovery fractions.

The third figure shows the relationships among (1/Var v), Var y and R. This graph can be used to estimate σ^2 from estimates of R and Var y. The estimate of σ^2 is the product of the estimates of Var y and (1/Var v). Thus suppose R at time t is estimated as 5O per cent and Var y is estimated as O.43. The value of (1/Var v) is then estimated as 1.12. The estimate of σ^2 is then (O.43) x (1.12) = O.48.

The final figure shows relationships among ($-\sigma$ Ev), R and the standard deviation of y. The estimate of u is found by adding ($-\sigma$ Ev) to the estimate of Ey. Again suppose R = O.5O and σ_y^2 = O.43. In this case σ_y = O.66 and the estimate of ($-\sigma$ Ev) is O.O65μ. Thus the estimate of u is found by adding O.O65 to the estimate of u_y.

Hopefully, the relationships which have been developed will lead to

(1) Quick and efficient estimation procedures for the parameters which characterize physical decay in aerosol chamber trials,

(2) Aids useful in designing and interpreting chamber experiments

(3) Procedures for testing the validity of the common assumptions with respect to physical decay, and

(4) Ways of evaluating the bias in methods of estimating biological recovery percentages which employ mass tracer data.

Finally and most important of all, these relationships constitute a start toward developing methods for testing the validity of the heterogeneous initial recovery model for biological recovery.

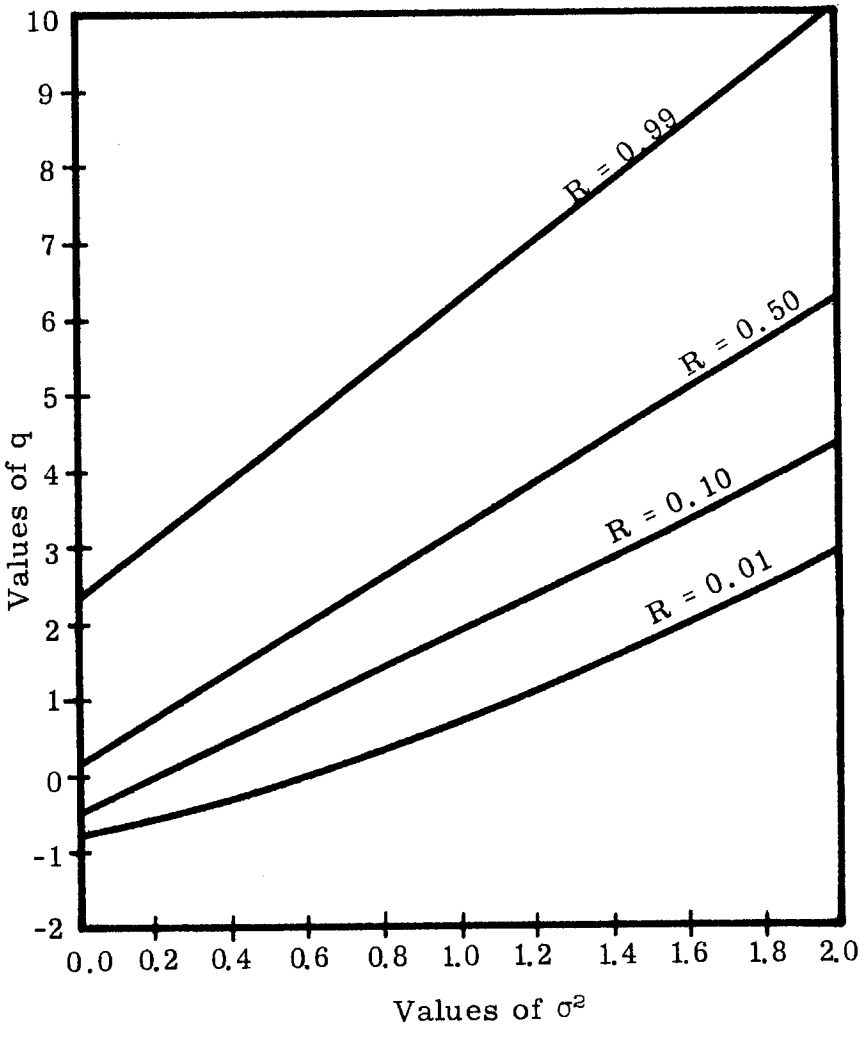


Figure 1. Graph of q versus σ^2 for various values of R where

$$R = (1/\sqrt{2\Pi}) \int_{-\infty}^{\infty} \exp\left[-1/2)\, v^2 - e^{-2(q-3\sigma^2-v\sigma)}\right] dv.$$

This page intentionally left blank.

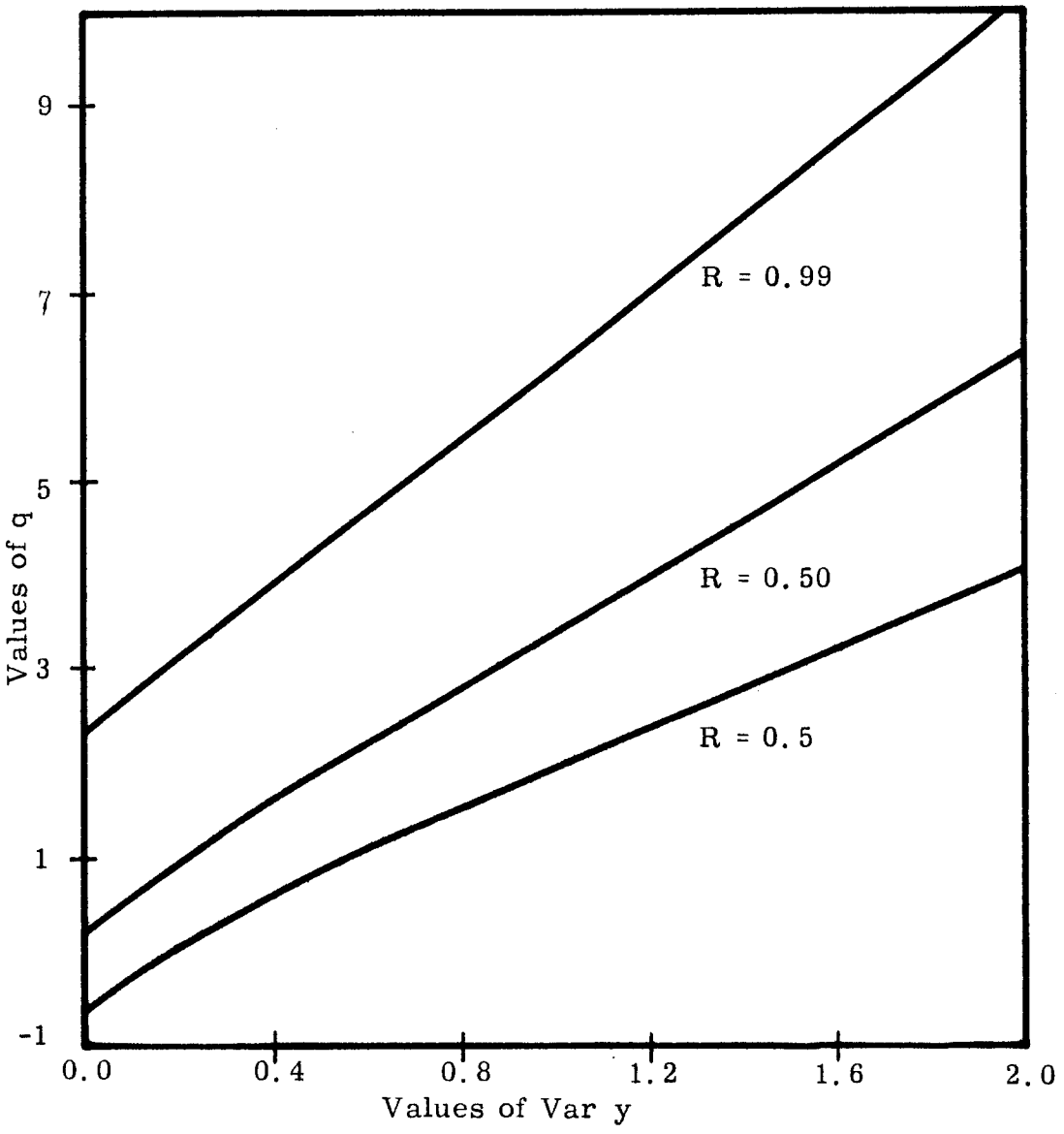


Figure 2. Values of q versus Var y at various physical recovery fractions.

This page intentionally left blank.

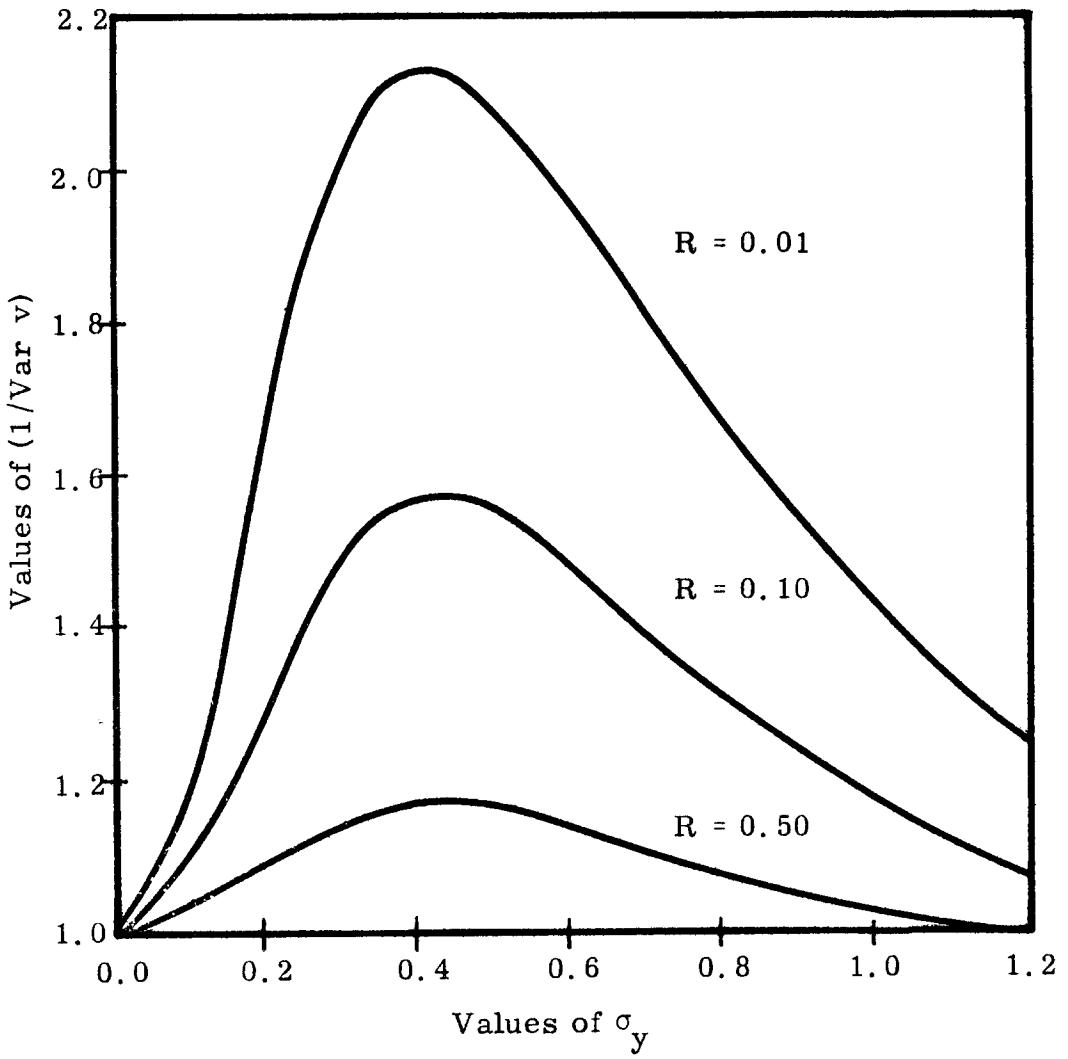


Figure 3. Values of (1/Var v) versus σ_y at various physical recovery fractions.

This page intentionally left blank.

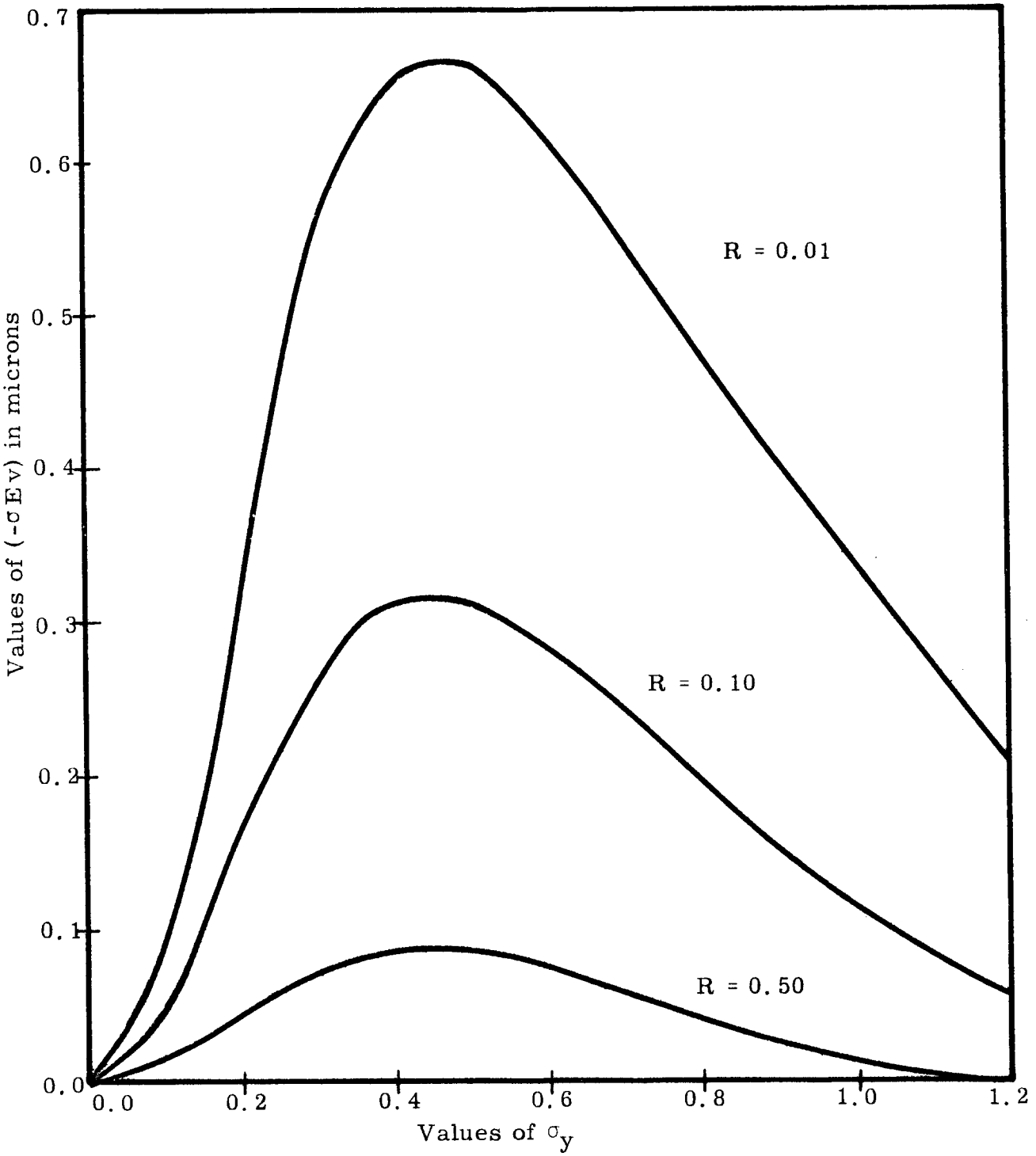


Figure 4. Plots of $(-\sigma E v)$ versus σ_y at various physical recovery fractions.

This page intentionally left blank.

THE ROLE OF INTUITION IN THE SCIENTIFIC METHOD

Nicholas M. Smith
Research Analysis Corporation

ATSTRACT. By virtue of Huygens' careful appraisal of the Newtonian method of inquiry, a logical fallacy has long been detectable in Newton's contention that he had deduced truth from observations of nature. The logical fallacy is concerned with two aspects over which empiricism has no control: (1) the observation that what constitutes a fact in the test of a theory is not determined by empirical principles alone, and (2) the observation that the constructs of science cannot be demonstrated to be other than sufficient with respect to the so-called facts. The necessity--that is, the uniqueness--of these constructs can never be established. The realization has gradually emerged that whatever is "true," "valid," or "warrantable," is not to be determined by absolutely singular discoveries produced in flashes of insight, but by selection among alternative creative insights on the basis of systematic tests constituting the reasoning process. The possibility of alternative insights of equal predictive applicability makes necessary the imposition of some principles other than empirical ones to decide among them. These principles are intuitive and categorical. That is, there exists a set of cognitive controls (of which empirical tests are members) which are established for the sole purpose of preventing ambiguity. Some of these principles have appeared in modern and contemporary science, notably the principle of relativistic invariance, which can be traced directly to the need of preventing procedural ambiguity introduced by transformations. There are other important introspective controls which delimit forms acceptable for application in the cognitive act. Many of these introspective principles are to be found separately in contemporary theories of value. We have formulated a theory of these cognitive controls based on our attempts during the past two years to establish the foundations of a rational methodology for systematic and organized prescriptive activities, that is, the decision process in all its generality. Science, as a decision system which has as its purpose the production of predictive theories, is shown to be a reduction of the more general axiologies. Consequently, many of the important so-called laws of science are not singular discoveries of properties of the external world entirely, but are in addition necessary properties of admissible forms which may serve as objectifications for applicable symbolic scientific models. Among these principles may be some of the most cherished scientific discoveries: the relativistic properties of space-time, the conservation of momentum, the conservation of energy, the second law of thermodynamics, and the Heisingberg uncertainty principle.

INTRODUCTION. The purpose of this discussion is to present the findings of a study* conducted at the Research Analysis Corporation for the past several years concerning the nature of the rational or cognitive process. This study has revealed that intuition (or introspection, as we shall call it) plays a much greater role in the process of rational thinking than we had heretofore suspected. The nature and complexity of the subject is such that a detailed presentation in a systematic and convincing step-by-step procedure would require many hours. We shall therefore resort to a presentation of our material in the time available in the form of an elaborate abstract. It cannot be hoped that such a shortened version of our presentation can be wholly convincing. It is hoped, however, that the attention of the reader will have been directed to some shortcomings in prevalent notions of the scientific method as applied to the design of experiments. We also hope that our method of resolving these problems will seem plausible and that your own interest in this exciting field will be aroused.

The study I am describing has been motivated by a search for the foundations of management science. In the term "management science" I mean to include such other terms as operations research, operations analysis, industrial engineering, economics, and the like. Those who practice these professions are not in complete agreement as to a statement of their mission; but in general these ingredients will be found in any definition: management science is somehow to provide a client with aids--quantitative or otherwise--to one of his decision processes. Or, the analyst may even go so far as to recommend specific decisions to the client. These aids, or these recommendations, are formulated with respect to the client's value system; it is further claimed, either implicitly or explicitly, that the management scientist employs a method which will somehow lead to better decisions. These definitions are charged with highly significant but poorly defined words. These words are "decision," "values," "method" and "better." The search for an understanding of the ideas behind these words has triggered an escalade of theoretical projects.

In the first place we have committed our interest to the field of practical decisions and therefore have become interested in the theory of decision algorithms. To many persons who practice our profession this subject may appear as the sole content of management science. This general field covers such divisions as mathematical programming, queueing theory, logistics theory, game theory, etc. The central commodity in terms of which decision

* This paper describes work done under RAC Study 5.4, "Valuation and the Cognitive Process," by N. M. Smith and M. C. Marney.

makers operate in reaching their decisions, however, is value. Since values are the determinants of decisions, a whole new theoretical field in the theory of value is developing. Value theory, on the other hand has drawn attention to the decision process in context of system. One cannot understand the act of evaluation without understanding the nature of the system in which the evaluation process is undertaken. This situation thus leads to a third theoretical project--the theory of selective systems.

Finally one must turn to the question of the validation or warranting of values, policies, and ethical systems. The question of such warrant, together with questions concerning the adequacy of the methodology of professional management scientists, have drawn out interest into the general theory of cognitive processes.

RECONSTRUCTION OF PHILOSOPHICAL FOUNDATIONS. The first impediment one encounters in the search for a method of warranting a value-decisions process (or, as we shall call it, a prescriptive process) is the conculsion that contemporary scientific method is inadequate. This inadequacy arises because one of the chief controls in the scientific method is a predictive process. One attempts to test or "warrant" a scientific theory by means of predicting future observations. A comparison of actual observations, with a suitable definition and range of measurement, will then define a warrant for a scientific theory. In the prescriptive process, on the other hand, one cannot confirm the adequacy of a value or policy by predicting one's own decisions, since these values are the indices which determine these decisions. Such a test would merely demonstrate a degree of consistency with respect to policy. We have gradually become aware that the prescriptive process is somehow different from the predictive process. In subsequent developments of the theory we have found that the differences and similarities between prescription and prediction are fairly complex, as I shall attempt to demonstrate.

Failing to find a rational prototype for the validation of prescriptive processes, we turned to a survey of historical and contemporary ethical and value theories. Although we found literally dozens of philosophical schools which purported to provide a means of selection and control of ethical systems, all of them exhibited inadequacies of various kinds. Failures of these systems and the historical failures of older ethical systems and scientific methods have occurred in a characteristic pattern: ultimately they have been confronted with situations which could not be resolved by the principles espoused.

It may also be observed that, to a large degree, it has been supposed that three great rational methodologies are treated as if they were separate processes. I am referring to (1) axiomatics, a selective system that produces valid formal systems, (2) scientific method, a selective system that produces predictive theories, and (3) axiological method, a method that produces prescriptives, policies, ethics. It has been assumed that axiomatics may be adequately controlled entirely by the rules of logic. On the other hand, the history of the scientific method has been characterized by the accretion of both logical and empirical controls and, in modern science to some degree, by the injection of intuitional controls. The axiologies have been presumed to have been controlled entirely by intuition. There have been, of course, attempts to approach ethics and values from a naturalistic viewpoint as predictive entities, but these studies can be shown to be concerned with value systems as objects, wherein our chief concern has been with value systems as subjects. (That is, what should my policy and my values be in order to determine my decisions?)

Having failed, then, to fine a rational prototype for the warranting of the prescriptive process, we have been forced to attempt a reconstruction in the philosophical foundations of the rational method in order to incorporate axiologies into the group of systematic rational pursuits. This reconstruction has taken the nature of a synthesis among modern scientific methods and contemporary and historic value and ethical theories. It promises, besides its direct application to axiology, to yield additional enlightenment on axiomatics (that is, the control of mathematical method) and the scientific method. This intimation is pertinent to the objectives of this conference and represents the specific subject of my discussion.

THE DEVELOPMENT OF A METATHEORY. It is desirable to distinguish between metatheory and object-theory. An object-theory is a theory which objectifies, or externalizes, objects. Such theories create the following types of objects: the objects of mathematics and logic are sets of self-consistent formal statements together with their consequent theorems; the objects of science are particular predictive theories and the elements thereof; and the objects of axiologies are particular policies that determine or prescribe practical decision. A metatheory, on the other hand, is a theory about object-theories. In particular it is a theory about the methods of admission or control and warrant of object theories. Our theory is a metatheory in which we are attempting to synthesize the metatheories of mathematics, science and axiology under one conformal perspective.

Any such theory presupposes, explicitly or implicitly, certain ontological and epistemological commitments, i.e., commitments as to what constitutes existence and knowledge respectively. Central among our commitments is the notion of relativism in three facets: the first is onotological relativism. This refers to the doctrine that existence of an object-construct is determined by its testability in principle or its connectability by inference to other object-constructs which are testable in principle. In other words, one rejects the notion of things-in-themselves or concepts which, by their very nature, are not subject to test. The term "test," of course, refers not only to empirical tests but to intuitive and formal tests as well.

The second facet is relativism in epistemology. This refers to the doctrine that certainty of knowledge of object-constructs, i.e., the establishment of apodictic truth (truth by necessity) is not obtainable. One must observe that the proofs of validity or warrantability of any scientific theory merely determine the efficiency of that theory in coordinating and clarifying the information obtained under a consistent predictive format. There is never any certainty that some other theory may not be developed which would describe the observations equally well or better; nor is there any certainty that the presently accepted theory will be adequate with respect to any future information that may be obtained. The consequence of these observations is that absoluteness at the object level is not meaningful.

A third facet is perspective relativism. This refers to the doctrine that an absolute reference for the judgment of object-theories is not obtainable. As we shall see in a moment, the consequence of this commitment is Einstein's principle of invariant transformations.

A second commitment presupposed by our meta-theory is that the sole function of the metacognitive process is the assurance of decidability of object-statements--that is, decidability with respect to their admissibility. The concept here is that relativism in object-space leads to degrees of freedom. Decision as to which object-constructs in this range of freedom are to be admitted must be accomplished in terms of some metaprinciple or control. This control then becomes a new absolute replacing the absolutes relinquished at the object level. That is, the controls are categorical, and they are metacontrols. The consequence is the conclusion that ambiguity is the sole motivation for decision.

THEORY OF COGNITIVE CONTROLS. There are, however, many kinds of ambiguities and each type of ambiguity necessitates a corresponding control. As we have said before, these controls

are categorical and their sole function is to resolve ambiguities of the class to which they apply. We have classified the controls in terms of three factors which we call formal, extrospective, and introspective. Besides these reflexive or internal controls there are also sets of external controls which we refer to as evolutionary and aesthetic. One of the great difficulties in developing an acceptable metatheory is collection of provision for selection among alternative object-theories which purport to apply to the same problematic situations. This selection is accomplished by means of evolutionary control--a generalization of the Darwinian principle--and aesthetic control (elegance). The evolutionary controls (fruitfulness, adaptability, and survival) represent ultimate commitments. Since the general thesis of this presentation can be developed without an elaboration of these important concepts, and since time does not permit such an elaboration, we shall forego any further discussion on these topics.

Central to our theory is the concept "objectification." Objectification represents the emergent result of a creative act which externalizes, at the level of a cognitive agent or self, a set of new conceptual entities or object-constructs on a trial basis as an act of policy and subject to a warrant to be established for predictive or prescriptive purposes by a set of cognitive controls. In this viewpoint all rational process is undertaken in terms of object-constructs, a special class of object-constructs being theories or models.

The formal controls of an object-construct apply to its format or formal properties. They insure admissibility under tests of consistence, completeness and independence.

Extrospective Controls. There are two acts in the extrospective control. One is the determination of the criteria of fact--that is, the selection of the specification of what constitutes a relevant fact based on a formal objectification selected among an indefinite set of objectifications as an act of policy. The criteria of fact becomes a filter through which extrospection is admitted as relevant to the problematic situation at hand. Thus, in the act of its admission, any "fact" has formal, introspective and extrospective components. There is no such thing as a purely extrospective fact. This supports the views of contemporary philosophers of science. So let me repeat: this conference, concerned as it is with the design of experiment, or as I have called it, the criteria of fact, is concerned with much more than extrospective information or data. In particular, it is concerned with formal and introspective (that is, intuitive) properties. Now "extrospection" means a looking outwards, or receptivity to

information processed through transducers and subsystems whose outputs are presented to mediation at conscious level. On the basis of the objectification or model one undertakes a prediction, that is, a symbolic projection forward in time beginning with an extrospectively determined initial state and in terms of a specific objectification. This leads to an expectation. A significant discrepancy at a later time between expectation and extrospection engenders extrospective ambiguity. In order to define extrospective ambiguity, one must first select (a) a range of initial admissible expectations, (b) a range of admissible divergencies between expectation and extrospection at a later time, and (c) a frequency measure. We can now define extrospective ambiguity as follows: a set of final expectations and extrospections are empirically nonambiguous if, and only if, a set of histories all beginning with initial states in the admissible range are examined and are found to contain a subset of final states lying in the admissible range around expectation, such that the ratio of the number of final admissible histories to the number of initial admissible histories is equal to or greater than the frequency measure.

The decisions as to the admissible initial and final ranges and the frequency measures are determined by the problematic situations which are desired to be resolved by the objectification. This range of application represents an aesthetic decision. One could, for example (see Table 1), set the frequency measure equal to zero, in which case he is saying he is indifferent to the correspondence between expectation and extrospection. He then becomes, by this aesthetic orientation, primarily concerned with the formal properties of his objectification. That is, he becomes a mathematician. He maintains an interest in the residual substantive properties of his constructs as exhibited by his attention to the nature and efficiency of his notation.

If the range of problematic situations desired to be faced includes prediction of situations, then the frequency measure is set at a non-zero value. We shall call such a person a scientist provided that he has also set his norms with respect to action implied by his objectification at null values, such that he is indifferent to such action. If he becomes aesthetically oriented completely toward action with respect to all immediate and mediate problematic situations, he will, in general, find that he has greater difficulty in satisfying all of the cognitive controls and hence, facing more restrictive constraints, must reduce the scope of comprehensiveness of his models. A primary control is that of practicability. What is practicable with respect to an action problem may be oversimplified with respect to a predictive problem. What is practicable to a scientist may be impractical

Table 1

A UNIFIED META-CONTROL SYSTEM

Metasystem	Operation	Range of problematic situations	Aesthetic Decisions	Scope of objectifications practicable
Axiology	Retrodiction	All practical problems	All norms effective	Most severely restrained, most reduced
Science	Prediction	Specific predictive situations	Action norms at indifference	Restrictions moderate, richer range of objectifications
Axiomatics	Formal extension	Consistent axiomatic systems	Action norms at indifference, extrospective ambiguity measure at null	Least restricted, richest in formal content

to a man of action, etc., the objectifications becoming correspondingly richer as one goes from axiology, to science, to mathematics, as the cognitive controls become, in some sense, degenerate.

There is also a very important difference between the viewpoint of prescription and the viewpoint of prediction--that is, the prescriptive operation, although it may warrant its objectification or model predictively, when it is used in prescription it is turned around and used retrodictively. Now, retrodiction is not the exact reverse of the predictive process. It is this difference between retrodiction and prediction which makes science and axiology acquire complementary characteristics. One is said to be adjoint to the other.

This property has very important philosophical as well as methodological implications. In particular, the primal or predictive viewpoint represents the view of a construct as an object whereas the complementary or dual can be interpreted as a representation of the construct as a subject. Thus, in terms of value theory, predictive value theory is a system by means of which one observes the decisions of another person as data and makes a theory the value system of that person as an object. On the prescriptive side of value theory, one is concerned with one's own values as determinants of one's own decisions. It is this process that is retrodictive.

Introspective Controls. Time will not permit a detailed discussion of introspective controls. We shall endeavor, however, to say enough about these so that their function and importance can be realized. Let us look at perspective control. This is the direct application of our epistomological commitment to perspective relativism. One may refer a statement in an objectification to a particular context of coordinate systems. Ultimately, they may be transformed into another and a description made in terms of another coordinate system. If this transformation depends upon the procedure or path taken from one system to another, one would naturally get a different result from the transformation depending upon the path taken. This would result in what we may call perspective ambiguity. If there existed an absolute point of reference, then a natural algorithm for transformations would be indicated. One would simply transform from the first coordinate system to the absolute origin and from there to the new coordinate system. In the absence of any such absolute perspective, one must limit the transformations to those having a particular property.

We seek a class of transformations which do not lead to ambiguity, regardless of the procedure or path taken. These are

called invariant transformations and they result in a formal description in the new coordinate system which is identical to the formal description in the old coordinate system. This is the principle of invariance. While it has a rather abstract title, and while the discovery of invariant transformations may sometimes be difficult, the intent and meaning of the principle is very simple. It says merely that one must avoid procedurally induced ambiguities.

In a space-time transformation of a physical theory, this leads directly to the Lorentz-Einstein conditions for a space-time transformation. Now it is also true and also of interest that if one looks at an object-space determined by a Markov stochastic system and asks for a nonambiguous or invariant transformation of a velocity in a Markov space (i.e., the velocity of movement of a probability configuration), half of the conditions for an invariant transformation emerge as a result. The adoption of the second half, as necessary for an invariant transformation, is equivalent to the introduction of the set of imaginary probabilities which, together with the real Markov probabilities, are to be associated with each transition. The results*, which may not surprise you by now, are none other than, again, the Lorentz-Einstein transformation equations in the space defined by the Markovian system.

Before the time of Einstein, science and axiology were presumed to be entirely separate, science being the province of empiricism and formal logic, whereas axiology, separate and disconnected, was the province of intuition. Then Einstein shook the very foundations of physical theory by a brillant and successful modification of the cherished concepts of space and time--a modification which depended not on empirical discovery but upon application of an intuitional requirement.

Even today the commitment to empiricism is sufficiently strong, and naive realism is so firmly established, that the full significance of Einstein's principle is not realized. This principle does not refer to a singular discovery of a property of the external world, but instead to a necessary property of admissible forms which may serve as objectifications for applicable symbolizations. We are constrained to think in terms of perspectively invariant forms, or we are inevitably led to ambiguity. Einstein, having achieved a nonambiguous formulation of mechanics, was then able to proceed to show a relation between energy and mass. The relation between energy and mass is not a substantive consequence of relativistic invariance; it is merely a formal result educed by an enlightened procedure which was made possible by a form of nonambiguous thinking.

*Smith, Nicholas M., "A Calculus for Ethics: A Theory of the Structure of Value," Behavioral Science, Vol. 1, Nos. 2, 3. 1956.

Other Implications of Invariance. Einstein's invariance has other far-reaching implications, particularly when we generalize the principle to state that all formal objectifications must be invariance with respect to significant transformations. "Significant" transformations are those in which the ambiguity arising from noninvariancy will be distinguishable from the range of admissible extrospective error. Generalized invariance has particular importance and implication in value-decision theory. One demands by application of this principle that the transformation of decision from a present to a future state by means of the Chapman-Kolmogorov transformation shall lead to a form of the value-decision equation identical with the initial one. If this were not true, then the decision indicated by the value-decision equation would depend upon the procedure in which a decision was staged into parts for analysis. The requirement of invariance with respect to time-translation transformation is insured first by the nature of the Chapman-Kolmogorov equation, and second by the placing of an important restriction on the decision operator. This restriction is one of commutation. A decision operator which commutes through the stages of decision process will permit an invariant transformation of the equation as applied from one point of reference in time to another. This property is also known by another name. It is the principle of optimality of dynamic programming. The latter is connectable to Euler's Weirstrasse and Legendre conditions of steepest descent algorithms.

It may also be shown that the Chapman-Kolmogorov equation, as it enters into value theory, introduces a concept analogous to momentum by virtue of the fact that the value equation is analogous to the conservation of momentum. Again the selection of a model in which the Chapman-Kolmogorov equation applies has been based upon the need for an invariant model as a starting point for the building of a theory. It also may lead, one adds, to a suggested generalization or modification of the law of conservation of momentum.

Other Introspective Controls. There are other introspective controls, each in its way fully as important as the principle of invariance; and each, when stripped of technical verbiage, merely assures nonambiguity and therefore decidability in the object-model.

One of these controls refers to the context of an object-construct. It requires that the context be specified in order to complete the meaning of the construct and it further specifies that an object-construct may have only one context, since if it had more than one context, it would be ambiguous. This particular

control, a modification of the Russell-Whitehead theory of types, can be expected to have important significance in the removal of certain kinds of paradoxes from modern logic.

Another introspective control, which is a direct statement of onotological relativity, constrains all object-constructs to those which are testable in principle. A third control requires furthermore that the test of the construct must not only be attainable in principle, it must be attainable and interpretable in terms of finite processes. This control will rule out infinite processes and continuous time-space as directly applicable to substantive constructs. Such concepts must assume a secondary status--that of operating constructs which guide the interpretation of finite extrospection in the context of a particular objectification. Examples of such secondary or operating constructs are: the wave functions of wave mechanics (which operate away in the act of evaluating a measurable entity), the concept "true" probability, which is never attainable; the optimum in a decision process, which is never achievable; also included is the class of decision variables as contrasted with the class of object variables.

The effect of introspective controls is to restrain the selection of object-models which are admissible for serving as the formal content of object-constructs. It therefore should come as no surprise that the form of all successful theories (that is, theories which prove to be admissible under extrospective, introspective and formal tests) will shown strong analogies.

Nor is it surprising that scientists have discovered introspective principles in the course of empirical investigations and have believed them to be part of the extrospective content of their observations.

This is not to say that these explicit principles, when they appear, are wholly intuitive, but rather that they are the consequents of intuitive requirements. A successful theory--no matter in what terminology it is formulated--will contain these principles in order to be nonambiguous.

Modern mathematicians have rediscovered Einstein's principle of invariance recently and have given it a name implying an extrospective connotation--they call it the principle of causality, and further, go so far as to say that it is the basic principle of classical physics.

These sets of controls alone are not sufficient to determine decidability. They are reflexive controls only. Ultimate decidability depends also on evolutionary controls, aesthetic controls, and on an intuitively established and evolutionary validated set of norms. Time does not permit discussion of them here. Their introduction and application merely serve to support the statement that extrospection is only a part of a concept, indeed, that the criterion of fact, although a necessary and desirable part of the rational process, must be imbedded for its understanding in the context of the metasystem. The nature of the evolutionary control is to insure fruitfulness, adaptability and survival of a concept as a workable construct. Formalizations which have inherent ambiguities must necessarily sooner or later reach a condition where decidability cannot be established; and they must surely fail. This does not imply that once a method of rational inquiry is devised which accomplishes decidability it can be expected to retain this property indefinitely.

Novelty is a characteristic of emerging concepts. Novelty will inevitably occur in the method of inquiry itself. The appearance of higher orders of abstraction will make necessary a re-establishment of cognitive control evolving through a repetitive cycle of ambiguity, undecidability, and finally the discovery of new rational principles.

CONCLUSION. The conclusion I wish to draw from these remarks is that knowledge depends as much upon intuition as it does upon extrospection and logic; and that these aspects are interdependent. I have hoped to make you aware of the implication that the nature of the rational act is much more complicated than heretofore supposed and that the simplistic views of cognition must irrevocably be discarded.

This page intentionally left blank.

HOW TO DESIGN WAR GAMES TO ANSWER RESEARCH QUESTIONS

William L. Pierce
Research Analysis Corporation

INTRODUCTION. In war gaming, to produce data for analysis, the game itself and the forms of data extraction must be designed to give outputs conveniently usable in answering the specific research questions that the game seeks to solve. This paper presents methods developed with this purpose in mind and employed in a rigidly-assessed, manually-played war game. With a brief historical sketch of the uses of war gaming from the 19th century to the present as background, present war games are classified into three groups, manual, computer-assisted, and computer-programmed, and defined. The design of manually-played war games is then considered separately in the context of a laboratory research tool. Finally, the application of the design requirements is illustrated by a description of TACSPIEL, RAC's division level war game.

Historically, war games have been developed and played to train officers and to test war plans. The former purpose was evident in the 9th century when Rigid Kriegspiel and Free Kriegspiel were developed. The testing of war plans by war games was used extensively by the Germans in the first half of the 20th century. After World War II, the technique of employing war games as an analytic tool was developed in an attempt to answer military questions pertaining to the battlefields of the future. With the advent of high speed computers, war gamers acquired a tool that permitted more comprehensive and complex games to be played. Also, the computer brought about a classification of war games by the war gaming community. A straightforward classification is to consider war games as either manually-operated, computer assisted, or computer programmed.

A manually-operated game is one in which all game orders are written, and assessments are made by people who are governed by strict or informal game rules, in essence, a rule book or an umpire.

A computer-assisted war game is a manually-operated game with the additional attribute that some of the bookkeeping and assessments are accomplished by a computer.

The third classification, the computer war game, is now in a prominent position in the war gaming field. In this type of game, after the start button is pushed, the computer plays the game without human intervention. Each and every situation thought to be important must be anticipated and simulated in the program with a suitable response.

THE DESIGN OF A WAR GAME AS A RESEARCH TOOL. Let us now consider a war game as a laboratory tool for military research. While much of what follows applies to all classifications of war games, I will be directing my words toward manual war games as a prelude to the later description of TACSPIEL.

One use of a laboratory tool is to enable the experimenter to investigate an area which would be inaccessible without the tool. For the military, the battle field of the future is the area at which attention is focused. To open this area for investigation, the war game becomes the tool. But for the experiments, or research plays, if you will, to be meaningful, the war game must be analytic. That is, it must be engineered to present a realistic environment for controlled experimental simulations with a view toward securing data for analysis.

It is not a difficult task for the military customer and the designer of war games to agree that war games can aid in the solution of military problems. However, when the research questions are directed at echelons from platoon to army, the designer must step back and take a sharp look at the design problems both obvious and subtle.

What does he see in the way of problems? First, there is the resolution problem. What is the gamet of resolution that should be considered in the game? Can the military units be played at company and battery level? Where and when can platoon, patrols, and radars be introduced? Do the research questions permit the game to be designed with divisions as the lowest echelon? To what resolution shall the unit deployments be recorded and played? How often should the game interactions be assessed?

The second problem which goes hand-in-glove with the resolution problem is the aggregation of the game models. If the basic unit is the company, then the models must reflect the capabilities of the company to move, fight, and receive casualties. There is a paramount requirement here when the designer builds the game models. Once he has chosen the resolution for the military units, he must be extremely careful to avoid constructing a game model for which no predictive data exists at the designed echelon. Should the input data to the model be lacking for the

echelon designed, the designer must re-examine the unit resolution. In short, resolution and aggregation as reflected in the game models are the two sides of the same coin.

Another problem which relates to the desired analytic nature of a manual war game is assessment of the play. A game can be played under a set of general rules with an umpire to assess battles, contacts, and other interactions. Or the game can be played under a set of rigid rules which are as detailed as the designer can make them. In this case, umpiring occurs infrequently and only when situations and capabilities arise that are not provided for in the rules. This latter method will produce the most objective and well-defined experimental conditions for a manually conducted game that can be attained.

Once the war gamer has decided on the resolution and aggregation level, he must now consider the basic tactical structure of the game. There are three characteristics which identify combat. They are the movement of units, the meeting of units known as contact, and the engagement by fire and maneuver of opposing units called battle. These characteristics are the basic tactical structure of a war game whether the game depicts ground, sea or air warfare. Any war game design must start by constructing models to represent these three characteristics.

Once the basic models of movement, contact, and battle exist, the war game is ready to consider the specific research questions of the military customer. When the research question is asked, the war gamer must ask himself three questions.

What models must be built such that the events to which the question is addressed will necessarily occur in the course of the play? What additional models must be designed to reflect the player's usual military capabilities, for example, artillery and tactical aircraft? How should all these models be constructed so that the output of each is presented in both usable form for analysis and with tactical realism for the players?

As an example relating to the first question, if the research question was to investigate the surveillance capability of a division in order to determine the detection rate of ground and airborne sensors, the war gamer would have to build, in detail, one model depicting the capability of each type of ground and airborne sensor including its associated delivery vehicle. These models would presumably allow the division commander as much flexibility as would be expected in actual combat and

would produce sufficient data for analysis. If the particular play was not directed at the surveillance question, an aggregate model depicting the intelligence acquisition capabilities of the airborne sensors could be built.

The third question concerning the presentation of the output of the models contains several requirements. The desired data for analysis must be presented in a format that can be easily manipulated manually or by a computer program. With the same format, the game assessments which contain the data for analysis should be presented unambiguously to the players and contain as much but no more information as could be expected to arise in actual combat under the same conditions that the model attempts to simulate. Without relaxing the above requirements, the recording of the assessments in a data format must not be time consuming. Otherwise the time saved during the analysis by preplanning the organization of the game data will be lost by the data recording process during the play of the game.

DESCRIPTION OF TACSPIEL. So far I have discussed the design of manually played war games pointing out the requirements for a basic tactical structure, associated sub-models, and data format.

I will now describe TACSPIEL, RAC's division-level war game, as an illustration of the application of the foregoing design principles and requirements.

The objective of TACSPIEL is ". . . to study operational problems of ground combat at division and lower echelon by analysis of play of a detailed tactical war game". The objective sets the framework within which the game was designed.

TACSPIEL is two-sided, free-play, analytic, rigidly-assessed, and manually operated. It is a free play game since after each side has been given their forces, scenarios, assigned missions, and approximate location of their respective reconnaissance elements, they are not constrained in their concepts of operation and organization other than the requirement to stay within the 45 x 200 km area of play.

It is analytic in that it is engineered to support research as a tactical environment for controlled experimental simulation of operational capabilities with a view toward securing data for analysis of their performance.

Design of Experiments

By rigid assessment it is meant that the game is conducted using a manual of rules that are as detailed as the years of operation and present ingenuity can make them, and that they limit Control as well as the players. The assessment of movement, contact, and battle are done by two ground assessors, one representing Blue, and the other Red. They must agree on the assessments that arise from interaction between the opposing forces that the game rules recognize. When unreconcilable situations, or situations and capabilities unprovided for by the rules arise, the senior controller intervenes,decides or umpires the offending assessment, and a new rule or a clarification of an old one is generated. Thus, the game manual is a living document, and some umpiring does occur. But this organization of game assessment is believed to provide the most objective and well-defined experimental conditions for a manually-conducted game that can be attained.

The game is played on 1:25,000 scale relief models of an area 44 x 90 km. The terrain selected is considered to influence but not dominate the operations at division level. The relief models depict three classes of roads, the railroads, all military streams, three classes of vegetation, all built-up areas and other limitations to movement. The relief is shown as 20 meter contour steps, with 2:1 vertical exaggeration. The interval of play is 30 minutes, and unit deployments are known to the 1 x 1 km square. The organization of units is resolved to the company, battery and armored platoon levels. This resolution at company level was dictated by the lack of predictive data at lower echelons for use as inputs to the battle model. Battery and armored platoon resolution was required to permit flexibility and realism in the employment of artillery and reconnaissance elements. The foregoing details apply to the basic tactical structure that exists in order to generate the operational environment. For research purposes, locations of point objects such as radars, OPs, minefield lanes, and boundaries can be specified to the nearest 100 x 100 m; timing of events can be reckoned to the nearest 5 minutes; and individual patrols can be assessed. The presence or absence of this latter kind of expensive detail largely depends on research objectives requiring them in any given play.

As discussed before, the basic tactical structure of a war game must contain the models for movement, contact, and battle or resolved units. These and other models developed to reflect the division's capabilities and tactical realism are shown in Figure 1. (Figures can be found at the end of this article.)

MOVEMENT MODEL. The movement model is designed to reflect a unit's maneuver capability. At the level of resolution played, this model represents in effect the movement rate of the center of mass of the unit. The movement model permits cross-country and road movement, deployed or in column for tracked and wheeled vehicles and dismounted troops. Reductions and restrictions on the rate of movement occur whenever the slope of the terrain is excessive, dense vegetation is encountered, escarpments occur, or when moving under artillery fire.

CONTACT MODEL. The contact model was developed to reflect a unit's visual and oral detection capability. In order to permit terrain to play its natural role on the battlefield, ridge lines are present on the terrain board in the Control Room. If any part of a unit's deployment area is in a terrain cell, as defined by the Control terrain model ridge lines, it is assumed that the unit exercises the full surveillance capability in that cell. Further, enough of the unit is in that cell to provide a complete sample of the unit for enemy visual surveillance and possible complete reporting.

Since in actual combat, initial visual contact does not always result in perfect identification as to the size and type of the contacted force, a distribution of possible identifications is used to determine the content of the contact report.

A three-digit code is used to report the size and type of the contacted unit. The size of the contacted unit can be reported accurately, or simply as "unknown". The type of unit may be reported accurately, i. e., medium tanks, 8-in How, etc., or given a general identification, i. e., armor, artillery, etc. or just as "enemy". The output of the assessment of contact is a contact report. The information contained in this report includes the size and type of the enemy unit as obtained from the identification distribution, its location when contacted, or its location when contact was broken, its attitude, i. e, moving halted, deployed, in column, and if moving, its direction of movement.

BATTLE MODEL. The third part of the basic structure, the battle model, is the most difficult model to design. If the battle model attempted to depict squad combat, how would one predict the behavior under a set of battlefield conditions of a squad without including the make-up of the individuals? At TACSPIEL's level of resolution, the company, group prediction is possible and feasible because of available basic knowledge from experience. This model must motivate the players to win. At the same time, the model must be as simple and quick to use as possible.

While the basic time resolution is one-half hour, battles are assessed on an hourly basis. That is, engagements are assessed once each hour of engagement. No winner or loser is declared but rather each force may accumulate casualties and the battle location can change if the attacker is successful. At the end of each hour of battle, the commanders receive reports of their own casualties, movement of the forces, and an estimate of casualties inflicted on the enemy. At that time they may attempt to reinforce or withdraw engaged units.

In the assessment of a battle, the engaged units basic combat effectiveness, their casualties at the start of the battle hour and the amount of casualties caused by enemy supporting artillery are used to calculate an attacker to defender force ratio. This force ratio is then used to determine by random number selection that hour's battle casualties on the attacker and defender, and the penetration of a successful attacker.

ARTILLERY MODEL. Since artillery has a capability which, in the real world, the division commander is able to employ with a high flexibility and effectiveness, the artillery model must be built to realistically reflect this capability. The emplacement time, rate of fire of the weapons, and the availability of the ammunition are the limitations on the employment of artillery. Since the effectiveness of rounds of different caliber to produce casualties vary, a standard unit of effectiveness called the fire unit or FU is used. One FU is equivalent in casualty production to 24 105-mm Howitzer HE rounds. The effectiveness of rounds of other calibers is equated to this measure. Thus, all fire missions are described and assessed, and a battery's basic load and resupply, computed in fire units.

The assessment of casualties is based on the number of fire units delivered, the type of target (armored, personnel) and posture of the target (exposed, attacking, defending, in woods, etc.) and the extent of observation on the target.

The results of a fire mission are reported to the side firing and the side receiving the shelling. The number of fire units expended, the type of target fired upon, and the target's location, are reported from the firing battery. If the fire is observed, an observer's report will contain the type and location of the target, an estimate of the damage inflicted, and the firing battery's designation to indicate what fire mission was being observed. For the unit receiving the fire, a shell report is generated containing an estimate of the amount of fire received and the amount of casualties suffered.

AIR OPERATIONS MODEL. The air operations model is designed to include the employment of air defense artillery, tactical aircraft against ground targets in support of engaged troops and against ground targets behind the enemy lines, airlift capability for divisional troops, and organic helicopters used in a reconnaissance role or to deliver fire from the air.

The effectiveness of the tactical aircraft's ordnance is equated to the artillery fire unit. Three stages of tactical air alert are played: No Alert, Standby, and On-station CAP. Three types of air missions are played, specific target, armed reconnaissance, and battle support. Appropriate planning times are assessed against aircraft when ordered from one of the air alert stages into one of the air missions.

Air defense artillery kill probabilities are based on their rate of fire, engagement ranges, altitude and speed of the aircraft.

The output of the air operations model is a report indicating the number of aircraft in the mission, the number surviving, and the result of the mission.

GROUND AND AIR SURVEILLANCE MODELS. To reflect the division's surveillance capability, OPs, patrols, and surveillance radars for the detection of moving personnel and vehicles are played in the ground surveillance model. The characteristics of the radars are obtained from the results of field tests.

The reconnaissance and surveillance capabilities of several air-borne devices and agencies are amenable to war game simulation. For the purpose of TACSPIEL, however, only those systems designed to concentrate in the 10-km area immediately beyond the line of contact are played.

Based on the previous plays in which each surveillance mission beyond 10-km from the LC was individually played, an aggregate effectiveness has been developed for the surveillance capability in the zone in excess of 10 km beyond the LC. This aggregated deep penetration surveillance includes information gathered by air photos, infrared devices and side-looking air-borne radars.

The output of both the ground and air surveillance models reflects the normal capabilities of each sensor.

LOGISTICS AND VEHICLE BREAKDOWN MODELS. The next model illustrates how TACSPIEL was applied to generate data to support analysis of a research question. The research question concerned an analysis of consumption and resupply in the ROAD Division of Class III and V Supplies. To generate the data, a tactical logistics model was developed.

The model assumed an infinite stock of ammunition at the Army Supply Point. Using the organic transportation available in the division, the player was required to order up the ammunition he needed under a side condition that the fuel for the divisional units must be hauled simultaneously out of the same transportation capability.

All basic loads and ordering of ammunition were reduced to one unit, the "fire unit" of effect. The POL consumption rates and basic loads of the various units were also reduced to a single "consumption unit" or CU, equal to 18 gallons of gasoline. Finally, the transport available to haul the basic loads and resupply Class III and V was reduced to a "transportation serial" equivalent to 7 1/2 tons of lift. In this manner, players could requisition ammunition, POL, and transport in a system of units that was independent of the detailed tables of equipment of the organization concerned.

TACSPIEL has undertaken on a trial basis a model to simulate vehicle maintenance and mechanical failure of combat vehicles. Vehicle maintenance in the division is simulated by assuming that the level of availability of wheeled vehicles is in a steady state during the play of the game. The level of availability is assumed to be 80 percent for units having five or more trucks over one ton carrying capacity. This 20 percent loss of hauling capacity is reflected through reduction in basic loads of Class III and V supplies available to the player, and in resupply capabilities.

The breakdown model has been developed from data from field trials on the M60 tank and M113 APC. This model simulates mechanical failure of APC, armored vehicles and SP artillery. During each interval in which these types of units move, the unit is assessed for breakdown. If breakdown occurs, the unit effectiveness is degraded 5 percent. At appropriate times during the game, vehicles repaired at the divisional support group are returned to the game.

By adding new simulation models and proving them out TACSPIEL can expect to increase its potential to produce useful data for research.

MECHANIZATION OF TACSPIEL. The method by which game data are made available for analysis is by employing IBM punch cards as a medium of exchange for the vast bulk of orders and reports and for the recording of data. A vocabulary of codes has been developed to transmit the game messages between the players and Control. The orders recognized by the game rules and the assessments are punched on IBM cards using IBM port-a-punch holders and 40-column partially preperforated cards. These 40-column cards are divided into several fields. One format is used for player orders and another for assessments (Figure 2). These formats are designed to include a three-digit order code or report code, the unit's order of battle, its coordinates, and all pertinent information in the order or assessment.

The flow of orders from the players to Control, the assessment of the interactions pursuant to the orders, and the reporting of the results to the players is called the Order-Assess-Report Cycle (Figure 3).

The player-commanders write their mission orders and organize their units for combat using overlays and mission order forms, (Figure 4). The written orders are translated into TACSPIEL order codes in the ORDERS column of this form (Figure 5). After coding the orders, the orders are punched on the 40-column IBM cards, using the ORDER format. These cards then go to Control with their ground mission order.

Upon receipt by Control of the IBM cards, the data on these cards are transferred to standard 80-column IBM cards by an IBM Summary Punch. The 80-column cards are then used to prepare a worksheet for the assessors called the Unit History Form. The format (Figure 6) groups the combat units with their initial coded orders as they are organized for combat on the mission order to provide continuity in time and to help organize the assessorls work. Additional headings (ORDer, ACTion, POSition) are printed to permit the assessors to enter interval by interval any changes in orders from the players and assessment notes for each unit. Enough space on each page is available for listing units that become attached to an organization during the game. When the first page is filled, additional pages are produced which may reflect changes in the make-up of the combat organization.

After the Unit History Forms have been prepared, the action of the opposing forces is assessed. Reports generated by interactions are punched on the IBM cards in the Assessment Format, transferred by the IBM Summary Punch to standard IBM cards containing prepunched

military English, listed, and distributed to the players. For example, the output of an artillery assessment would use Code 801 for the firing unit's report and Code 750 for the observer's report, and the assessment by Control would be listed from the IBM cards (Figure 7). The percent casualties to the enemy unit and its order of battle would be deleted from the player's copy.

The player's response to the reports results in new orders by which a new Order-Assess-Report cycle is generated. Unless the general mission of a combat team is changed, the mission order form is not required for transmitting additional orders affecting that combat team.

By the use of the IBM cards and the mission order forms, a complete rapidly accessible record of the game is available for analysis. In addition, by the expedient of reproducing the cards, the data become accessible to any qualified study outside the gaming group itself.

SUMMARY. To summarize, in the design of a war game, a basic operational structure of movement, contact, and battle is required. Within this structure detailed simulations of the real world events to be studied are introduced in order to generate data to answer the specific research questions of the military customer.

In order to extract the data to answer the research questions rapidly and efficiently, the TACSPIEL war game has developed a method which combines a vocabulary of order and assessment codes with IBM cards. The result is a compact and complete game record for analysis and a data source on which analytic research can be based.

This page intentionally left blank.

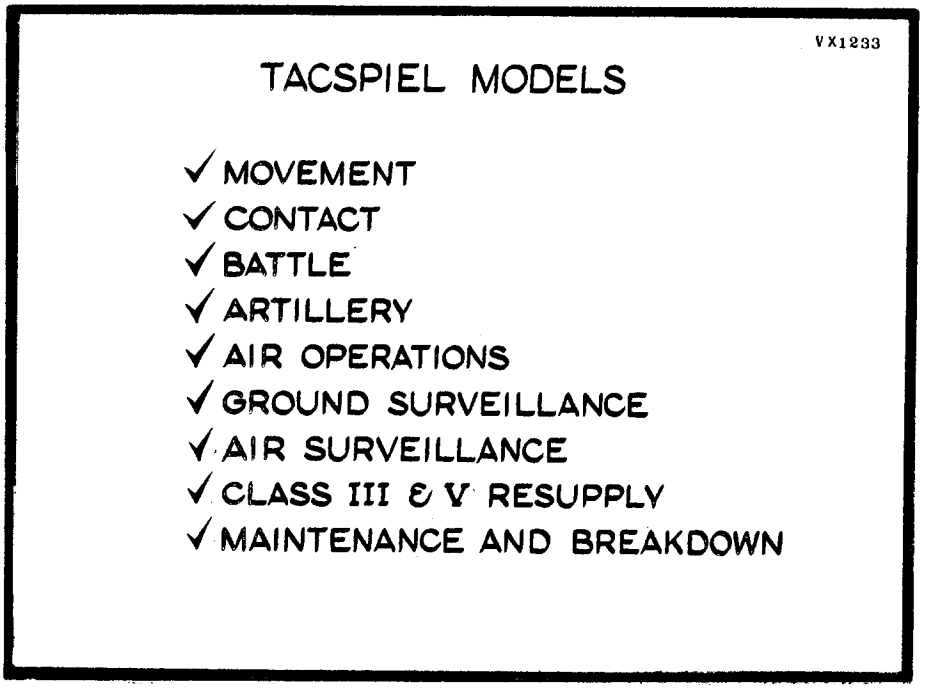


FIGURE 1- Tacspiel Models

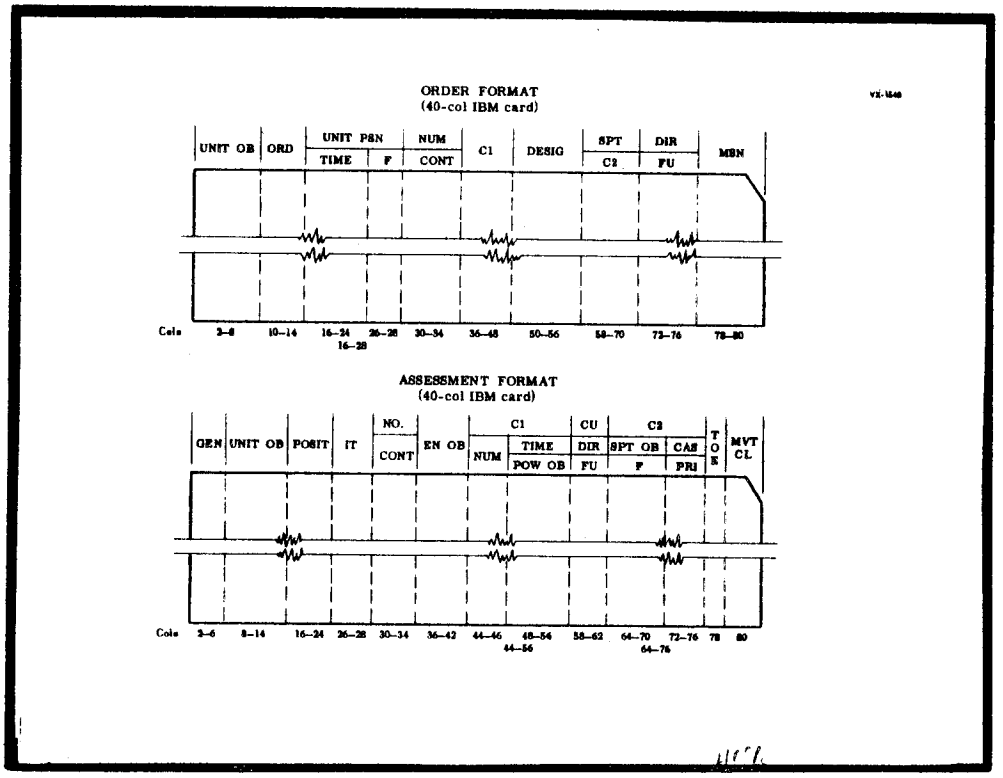


FIGURE 2- Order and Assessment Formats

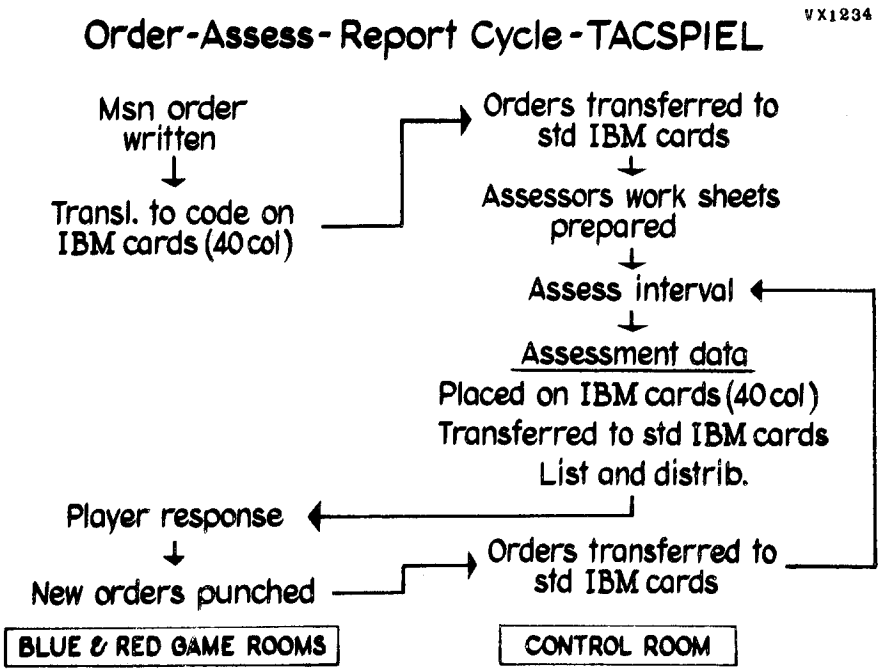


FIGURE 3- Order-Assess-Report Cycle

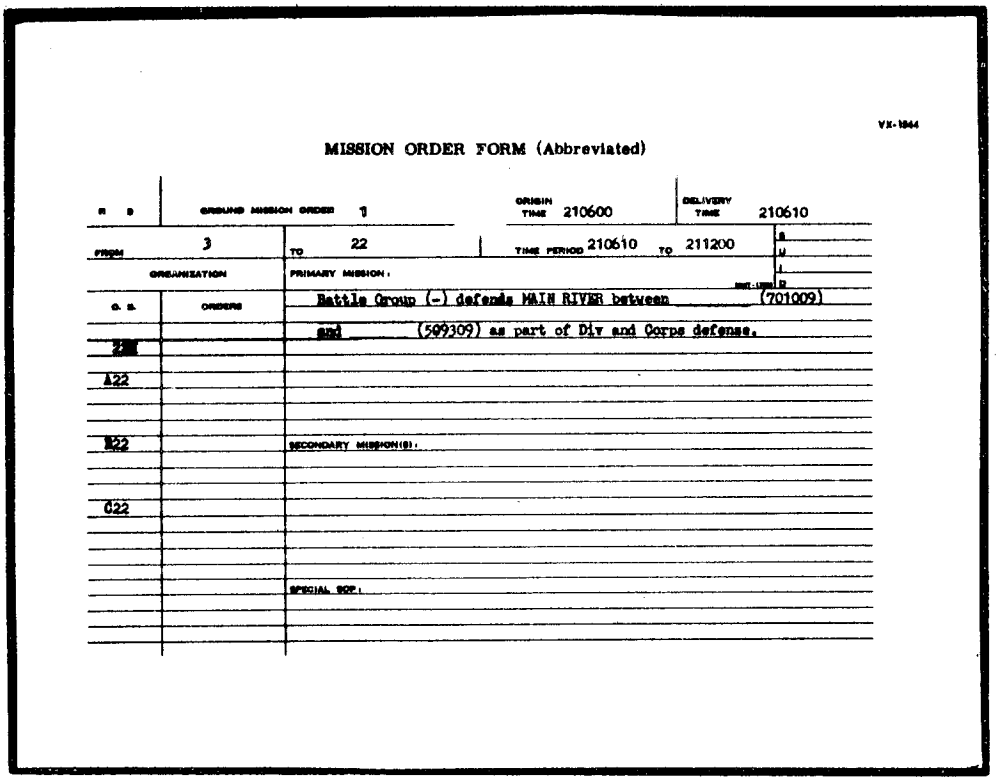


FIGURE 4- Mission Order Form(Abbreviated)

This page intentionally left blank.

MISSION ORDER FORM (Abbreviated)

with Coded Orders Added

VX-1545

R B | GROUND MISSION ORDER 1 | ORIGIN TIME 210600 | DELIVERY TIME 210610

FROM 3 | TO 22 | TIME PERIOD 210610 TO 211200

ORGANIZATION | PRIMARY MISSION:

Battle Group (-) defends MAIN RIVER between _____ (701009)

and _____ (509309) as part of Div and Corps defense.

O. B.	ORDERS
22H	004 22
A22	033 1 020 701009 507305
B22	033 1 020 700005 509309
C22	032 1 020 690003 509307
	131

SECONDARY MISSION(S):

SPECIAL SOP:

004 Assigned command of (designated organization) DESIG

020 Assigned zone/sector of responsibility C1 to C2

032 Assigned reserve role in MSN

033 Assigned forward defense role in MSN

131 Defend assigned sector

FIGURE 5- Mission Order Form (Abbreviated) with Coded Orders Added

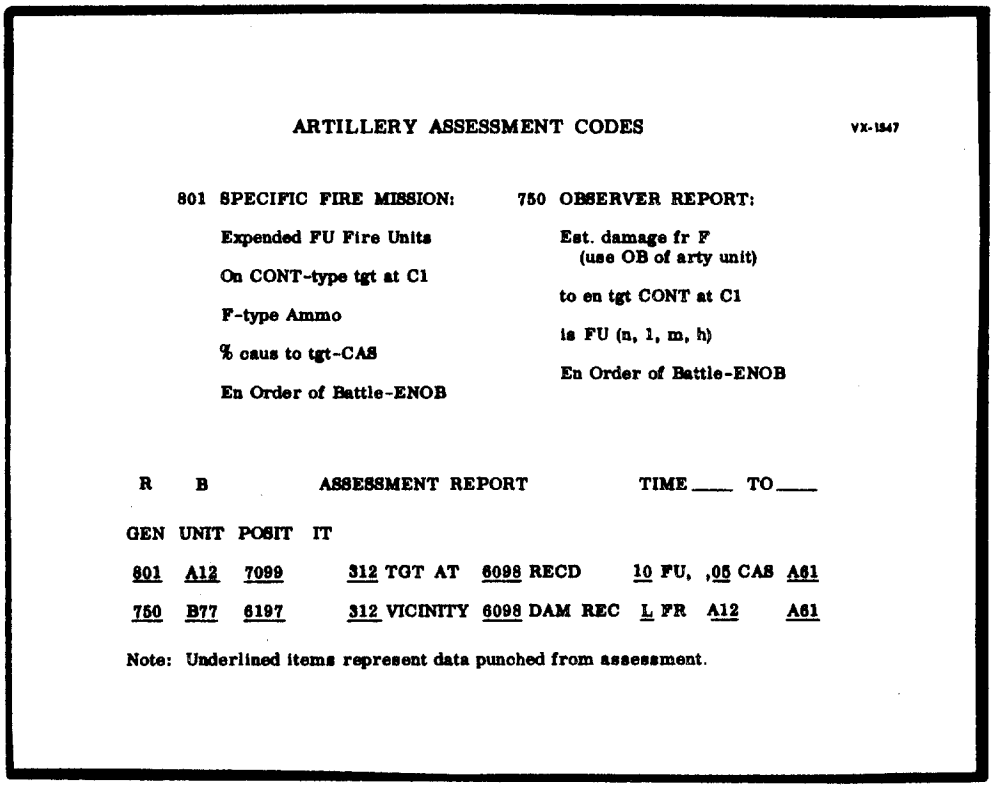
VX-1546

R B GAME _____ UNIT HISTORY FORM PAGE ___ OF ___ TIME ___ TO ___

CMD ECH	UNIT	ORD	UNIT TIME	PSN F	NUM CONT	C1	DESIG	SPT C2	DIR FU	MSN	TIME: ACT POS	TO ORD ACT POS
22BG	22H	004					22					
	A22	033								1		
		020				701009		507305				
	B22	033								1		
		020				700005		509309				
	C22	032								1		
		020				690003		509307				
		131										

FIGURE 6- Unit History Form

This page intentionally left blank.



ARTILLERY ASSESSMENT CODES VX-1547

801 SPECIFIC FIRE MISSION:
Expended FU Fire Units
On CONT-type tgt at C1
F-type Ammo
% caus to tgt-CAS
En Order of Battle-ENOB

750 OBSERVER REPORT:
Est. damage fr F
(use OB of arty unit)
to en tgt CONT at C1
is FU (n, l, m, h)
En Order of Battle-ENOB

R B ASSESSMENT REPORT TIME____ TO____

GEN	UNIT	POSIT	IT								
801	A12	7099		312	TGT AT	6098	RECD	10	FU,	,05	CAS A61
750	B77	6197		312	VICINITY	6098	DAM REC	L	FR	A12	A61

Note: Underlined items represent data punched from assessment.

FIGURE 7- Artillery Assessment Codes and Report

This page intentionally left blank.

EVALUATION OF PERFORMANCE RELIABILITY USING REGRESSION MODELS

Seymour K. Einbinder, Picatinny Arsenal
and
Ingram Olkin, Stanford University

0. ABSTRACT. Performance reliability is the probability that a weapon will perform its prescribed function under given conditions of environment at some particular time. Performance reliability models are defined for continuous performance response variables. Procedures are then described for the evaluation of reliability with emphasis on the application of univariate and multivariate regression analysis to single and multiple continuous response variables, respectively. Point and confidence interval estimation methods for performance reliability are discussed, and a sample problem is presented illustrating some of the basic concepts and results.

1. INTRODUCTION. A major problem during the research and development of a weapon or warhead is the assurance of high functioning reliability of the final prototype design. The reliability concepts and evaluation methods to be described are general and are applicable to a wide variety of systems and components.

A weapon during its lifetime may be subjected to many environmental factors or stresses such as temperature, vibration, acceleration, rough handling, etc. In addition, the stresses may be encountered singly, simultaneously or in sequence. The problem of testing and estimating reliability is of importance to the weapon developer in order to assure the user of a reliable weapon for use in any potential combat situation.

The establishment of high reliability with a high level of confidence generally requires the testing of larger numbers of items than are usually available during a development program for a complex and expensive item. Thus, it is generally necessary to obtain the most information with a minimum number of samples and tests. Improved and more efficient statistical methods are required in many cases to solve the reliability estimation problem.

Before solutions to a problem can be obtained, it is important to delineate the problem so that a representative mathematical model can be developed. Obviously, any solutions obtained can be no better than the underlying mathematical model which is assumed as a prototype of the problem.

In a previous paper [3], the emphasis was on test-to-failure and stress vs. strength analyses for single and multiple environments (stresses). The present paper is concerned mainly with reliability in the case of a continuous regression response surface, and thereby is an extension of [3]. Methods of point and interval estimation for the univariate and multivariate regression problems are discussed.

2. RELIABILITY CONCEPTS. Reliability of a weapon may be defined as the probability of a successful functioning under required conditions of the environment at some particular time. Successful functioning might require the successful operation of several or all components of a system. The outputs or responses of each component may be attribute or continuous. In the case of the continuous response, success may require that the response lie within certain limits (possibly specification limits).

To illustrate these concepts, a hypothetical shaped charge warhead section for a missile will be used as an example. Successful functioning of the warhead section may require that the S and A (Safety and Arming Device) must arm and detonate the warhead on target impact, and the warhead must then penetrate at least a specified distance into an armor plate target. This example could be further complicated by specifying arming limits for the S and A. Failure of the warhead section may occur in two fundamental ways:

(1) a complete dud or catastrophic failure may result such that no warhead detonation takes place, or,

(2) the warhead explosive train may be initiated but the armor penetration requirement may not be met.

The reliability of the warhead section is given by

$$R = (1 - P_D) R_{WHD},$$

where R is the overall warhead section reliability, P_D is the probability of a dud or catastrophic failure, and R_{WHD} is the conditional probability that the warhead exceeds the specified performance requirement. The latter will be referred to as the performance reliability and is of prime concern in this paper. The dud probability can be broken down

further according to various components. In general it is necessary to evaluate the dud or catastrophic failures separately from the performance failures since they do not have the same distribution and are mutually exclusive. Dud failures, being attribute, normally require larger sample sizes for evaluation with the same precision and confidence levels as would performance reliability based on continuous variables.

The remainder of this paper will be concerned with the evaluation of the performance reliability, R_{WHD}.

3. MATHEMATICAL MODELS. In this section we define the mathematical models upon which subsequent analysis is based. Univariate and multivariate responses and single and multiple stresses are considered. Estimation procedures are described in the subsequent sections.

3.1 Univariate Response. We have defined performance reliability as the probability that a continuous performance variable lies within certain specified limits. Thus, it may be required that the arming time for an S and A Device be greater than some minimum time required for safety. The performance variable or response may be thought of as a dependent variable which is a function of one or more environments or stresses which are the independent variables. In general, this functional relationship is unknown; however, we can approximate the response function over small regions of the function space by linear regression methods. Generally, we are concerned with the reliability under some critical stress conditions. These conditions will be referred to as a critical point or critical reliability boundary. The regression experiment is designed to provide the best information in the vicinity of this point. Figure 1 illustrates the experimental design in general terms for the univariate case. $(x_1, \ldots, x_m)$ represent the applied stresses or environments such as temperature, vibration, etc., and the elements of the design matrix represent the levels of each of these stresses. For example, x_{12} represents the second level of the stress x_1, etc. The column titled response vector represents the observed response obtained with each treatment combination. The response, y, is a continuous variable such as arming time in the case of the fuze, or possibly depth of penetration in the case of a shaped charge warhead; the response, y_i for the i^{th} treatment combination may be expressed as a linear combination of the treatment levels plus some random error. The regression model and underlying assumptions are:

$$y = X'\beta + u ,$$

y: n x 1 is the observation vector,
u: n x 1 is the vector of random errors, and is normally distributed with mean vector O and covariance matrix $\sigma^2 I$, i.e., $E(u) = O$, $E(u\,u') = \sigma^2 I$,

X: m x n is the design matrix of rank r, $r \leq m \leq n$,
β: m x 1 is the vector of regression coefficients.

A geometrical interpretation of this regression model and its relation to performance reliability is shown in Figure 2 [Tables and figures can be found at the end of the article.] for the univariate response case. $(x_1, \ldots, x_m)$ are the stress variables, $(c_1, \ldots, c_m)$ are the components of the critical reliability boundary vector c. The points shown in the $(x_1, \ldots, x_m)$ plane represent the treatment combinations for the regression experiment, and the average response, y, is represented by the regression surface shown above the points. The regression equation provides estimates of the response for any point in the $(x_1, \ldots, x_m)$ space. The response y at some point such as the critical reliability boundary c : m x 1 is denoted by $y^{(c)}$ and is distributed according to $N(c'\beta, \sigma^2)$. The lower performance limit which the response $y^{(c)}$ is required to exceed is denoted by $y^{(O)}$, and consequently, the performance reliability R is given by:

$$(1)\quad R(c) = P\left\{Y^{(c)} \geq y^{(O)}\right\} = \int_{y^{(O)}}^{\infty} n\left(y^{(c)} \middle| c'\beta,\ \sigma^2\right)\, dy^{(c)} \equiv g_1(c'\beta, \sigma^2; c) ,$$

where $n(x \mid \mu, \sigma^2)$ is the density function for a normal population with mean μ and variance σ^2. This expression represents the shaded area under the normal curve shown in Figure 2.

Thus, our problem is to estimate g which is a function of the unknown parameters β, σ^2, and the fixed point c, based upon a sample of size n treated as a single or multiple regression experiment.

3.2 Multivariate Response. The univariate model will now be extended to include cases where more than one continuous response may be observed on a single experimental unit such as S and A arming time, functioning time, and self-destruct time; also, the responses may be correlated. Multivariate analysis techniques permit the correlation between responses to be investigated. As before, the problem is best illustrated by examining the table shown in Figure 3. The design matrix X is exactly the same as for the univariate case; $(x_1, \ldots, x_m)$ is still the vector of applied stresses. However, instead of a single response vector of y's, we now have p responses $(y_1, \ldots, y_p)$. Thus, for each treatment combination we observe p responses so that our response vector for the univariate case has now become a response matrix where the column vectors may be correlated, and the rows which represent independent response vectors are uncorrelated. The multivariate model and assumptions are:

$$Y = X' B + U,$$

Y: n x p is the response matrix
X: m x n is the design matrix of rank $r \leq m < n$, $p \leq n-r$,
B: m x p is the matrix of regression coefficients,
U: n x p is the error matrix,
u_j, j=1, ..., n are the rows of U and are independently and identically distributed, each having a p-variate normal distribution with mean vector O and positive definite covariance matrix Σ.

From the multivariate model, a p x 1 response vector $y^{(c)}$ is obtained for the response at the critical reliability boundary vector c: m x 1. The response vector $y^{(c)}$ is distributed according to $N(c'B, \Sigma)$, in which the p x p covariance matrix Σ takes into account any correlations between responses.

The performance reliability R for the multiple response case is given by:

$$(2) \quad R(c) = P\left\{Y_1^{(c)} \geq y_1^{(O)}, \ldots, Y_p^{(c)} \geq y_p^{(O)}\right\}$$

$$= \int_{y_1^{(O)}}^{\infty} \cdots \int_{y_p^{(O)}}^{\infty} n(y^{(c)} \mid c' B, \Sigma)\, dy_1^{(c)} \cdots dy_p^{(c)}$$

$$= g_p (c' B, \Sigma; c) \ .$$

A graphical representation of the performance reliability in two dimensions is shown in Figure 4 for the multivariate case. The above integral represents the volume of the multivariate normal density function over the shaded quadrant whose vertex $y^{(O)}$ is the vector of specification limits.

Thus, the general problem may be summarized as follows: Based upon the results of a suitable experimental design with a sample of size n, it is required to estimate the g function for the univariate and multivariate cases, both by point estimation and confidence limits.

4. POINT ESTIMATION. The general problem is to estimate the performance reliability functions defined for the univariate and multivariate responses both by a point estimate and confidence limits based upon responses observed on a sample of size n subjected to various stress treatments in accordance with a suitable experimental design. The experimental designs used for exploring response surfaces [1, 2] are generally suitable for exploring the region around the critical reliability boundary.

4.1 Univariate Point Estimates. The g or R functions to be estimated may be written as follows for the univariate case:

$$R(\beta, \sigma^2) \equiv g_1(c'\beta, \sigma^2; c) = \int_{(y^{(O)} - c'\beta)/\sigma}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2)\, dt,$$

where c and β are $m \times 1$ vectors.

We consider three types of point estimates. Suppose we write

$$K(\beta, \sigma) = (y^{(O)} - c'\beta)/\sigma \quad ,$$

then

$$R(\beta, \sigma^2) = \int_{K(\beta, \sigma)}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2)\, dt. \tag{3}$$

The first estimate of R is based on using $K(\hat{\beta}, \hat{\sigma})$, where $\hat{\beta}$ and $\hat{\sigma}$ are appropriate estimates of β and σ.

The least squares estimate of β is given by

$$\hat{\beta} = (XX')^{-1} X\, y \quad , \tag{4}$$

where X: $m \times n$, of rank $m \leq n$, is the design matrix, and y : $n \times 1$ is the response vector. An unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{(y - X'\hat{\beta})' (y - X'\hat{\beta})}{n - m} \quad . \tag{5}$$

Thus, we may use the estimate

$$K(\hat{\beta}, \hat{\sigma}) = (y^{(O)} - c'\beta)/\hat{\sigma} \quad , \tag{6}$$

from which we obtain

$$R(\hat{\beta}, \hat{\sigma}^2) = \int_{K(\hat{\beta}, \hat{\sigma})}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2)\, dt. \tag{7}$$

A second estimate is based on the UMVU estimate of K, namely

$$(8) \qquad \tilde{K}(\beta, \sigma) = \sqrt{\frac{2}{f}} \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})} \frac{y^{(0)} - c'\hat{\beta}}{\hat{\sigma}} = \sqrt{\frac{2}{f}} \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})} K(\hat{\beta}, \hat{\sigma}) \ ,$$

where $f = n-m$, from which we may use

$$(9) \qquad \tilde{R}(\beta, \sigma^2) = \int_{\tilde{K}(\beta, \sigma)}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2) \, dt$$

as an estimate at $R(\beta, \sigma^2)$.

Although $\tilde{K}$ is a UMVU estimate of K, it is not the case that $\tilde{R}$ is a UMVU estimate of R. Consequently, a third procedure is based on the UMVU estimate of R, [4], and is given by

$$(10) \qquad \bar{R}(\beta, \sigma^2) = \int_{0}^{\max[0, \eta]} \frac{t^{\left[\frac{f-1}{2} - 1\right]} (1-t)^{\left[\frac{f-1}{2} - 1\right]}}{B(\frac{f-1}{2}, \frac{f-1}{2})} \, dt \ ,$$

where $\eta = \frac{1}{2} - \frac{(y^{(10)} - c'\beta)}{2\hat{\sigma}\sqrt{f(1-c'(XX')^{-1}c)}} = \frac{1}{2} - \frac{K(\hat{\beta}, \hat{\sigma})}{2\sqrt{f(1-c'(XX')^{-1}c)}}$.

Note that $\bar{R}(\beta, \sigma^2) = 1$ if $\eta > 1$, and that the estimate is valid for critical vectors c such that $c'(XX')^{-1}c < 1$.

Unfortunately, comparisons of the risk of these estimators are unavailable, since the determination of the variance is quite complicated, and was not attempted in this paper.

4.2 Multivariate Case. In the multivariate case, we have

$$R(B,\Sigma) = \int_{y_1^{(O)}}^{\infty} \int_{y_p^{(O)}}^{\infty} \frac{e^{-1/2 \ \mathrm{tr}\ \Sigma^{-1}(Y-X'B)'(Y-X'B)}}{|\Sigma|^{pn/2}(2\pi)^{pn/2}} \ dY,$$

where $Y : n \times p$, $X : m \times n$, $B : m \times p$. As in the univariate case, we can consider $R(\hat{B}, \hat{\Sigma})$ as an estimate of $R(B, \Sigma)$, where $\hat{B} = (XX')^{-1}XY$, and $\hat{\Sigma} = (Y - X'\hat{B})'(Y - X'\hat{B}) / [p(n-m)]$. The problem, however, is still to evaluate the multivariate normal distribution over an orthant. In fact, whether we use this estimation procedure or another, the difficulty of carrying out such an integration still remains. However, for any particular problem, one can employ numerical procedures to yield an answer. Another possibility which has not been considered in the literature is to obtain a lower bound for $R(B, \Sigma)$ in terms of known functions. Further work in this area is required.

5. CONFIDENCE INTERVALS. The problem of obtaining confidence intervals for the g or R functions is considered next. The general method is discussed in [3], and is now extended to the regression model. In Section 4, three estimates were presented. For only the second procedure is the distribution theory known, so that exact confidence intervals can be obtained. However, the first procedure does lead to approximate or asymptotic intervals based on the normal distribution.

5.1 Exact Confidence Intervals. Since $R(\beta, \sigma^2)$ is a monotone function of $K(\beta, \sigma)$, if we can find a confidence interval (K_1, K_2) for K, we will then have a confidence interval (R_1, R_2) for R, where

$$R_i = \int_{K_i}^{\infty} (2\pi)^{-1/2} \exp(-y^2/2) \ dy.$$

It is shown in Appendix A that $K(\hat{\beta}, \hat{\sigma})/\|a\| \equiv t(f, \delta)$, where $\|a\|^2 = c'(XX')^{-1}c$, has a non-central t-distribution with $f = n - m$ degrees of freedom and non-centrality parameter.

$$\delta = (y^{(O)} - c'\beta)/(\sigma \|a\|) = K(\beta, \sigma)/\|a\| \quad .$$

Thus, a lower and upper confidence limit with confidence coefficient $1 - \alpha$ may be obtained by finding the values of δ_i for which

(11) $$P\left\{t > \frac{K(\hat{\beta}, \hat{\sigma})}{\|a\|} \,\middle|\, f, \delta_i\right\} = \alpha_i, \quad i = 1, 2, \text{ where } \alpha_1 = 1 - \alpha/2, \text{ and}$$

$\alpha_2 = \alpha/2$. Table IV in [6] may be used to obtain δ_i for seventeen values of ϵ.

The tabulation by Resnikoff and Lieberman [6] of the percentage points of the non-central t-statistic may be conveniently used to obtain the limits δ_1 and δ_2 that satisfy (11). The entries in the table give the values of x such that

$$P\left\{\frac{t}{\sqrt{f}} > x\right\} = \epsilon \quad .$$

The table should be entered for the degrees of freedom $f = n-m$, the probability ϵ_i corresponding to α_i, and $x = K(\hat{\beta}, \hat{\sigma})/(\|a\|\sqrt{f})$. The required non-centrality value $\delta_i = \sqrt{f+1}\ K_p$, where K_p is the standardized normal random variable exceeded with probability p. The present concern was with one sided tails (one sided specification limits) for both the univariate and multivariate cases. A review of available point and confidence methods for two sided tails is given in [3].

5.2 Approximate Confidence Intervals. If we expand $R(\hat{\beta}, \hat{\sigma}^2)$ about $R(\beta, \sigma^2)$, we obtain the result that

$$R(\hat{\beta}, \hat{\sigma}^2) - R(\beta, \sigma^2) \sim N(O, V_\infty(\beta, \sigma^2)) \quad ,$$

where

$$V_\infty(\beta, \sigma) \equiv \sigma^2 \left[n(y^{(O)} | c'\beta, \sigma^2)\right]^2 \left\{c'(XX')^{-1} c + \frac{(y^{(O)} - c'\beta)^2}{2\sigma^2 f}\right\} .$$

Consequently (see Appendix B),

$$\frac{R(\hat{\beta}, \hat{\sigma}^2) - R(\beta, \sigma^2)}{\sqrt{V_\infty(\hat{\beta}, \hat{\sigma})}} \sim N(0, 1) \quad ,$$

from which we obtain the confidence interval

$$\left[R(\beta, \sigma^2) + z_{\alpha/2}\sqrt{V_\infty(\hat{\beta}, \hat{\sigma})} \ , \ R(\hat{\beta}, \hat{\sigma}^2) + z_{1-\alpha/2}\sqrt{V_\infty(\beta, \sigma)}\right],$$

where z_α is the $100\,\alpha$ % point of the $N(0; 1)$ distribution.

6. SAMPLE PROBLEM. In order to illustrate the results of the previous sections, a sample problem will be solved. A model representing the performance of a hypothetical shaped charge warhead section for a missile will be described, and the performance reliability will be evaluated based upon a Monte Carlo simulation of test results. The point estimates and confidence intervals obtained using the methods previously described will be compared with the true value of the reliability. Only the univariate or independent response cases are considered.

6.1 Performance Model. The warhead section to be evaluated consists of a shaped charge warhead and a Safety and Arming Device. It is assumed that the warhead is required to penetrate at least 10 inches into an armor plate target and that the minimum arming time of the S and A is 0.5 seconds. The warhead section is expected to meet these performance requirements under all possible combinations of vibration and temperature shock that may be encountered. To facilitate the illustration, only two stresses are considered in this problem, but the procedure is easily extended to more than two stress variables.

The two stresses, vibration in g's and temperature shock in standard cycles, are denoted by X_1 and X_2, respectively. Coded levels of the stresses are used throughout this problem to facilitate the analysis and simulation of test results. The relationship between the coded and actual stress units is of no importance with regard to illustrating the reliability evaluation methods and will be disregarded.

The critical reliability boundary is defined by the vector $c' = (c_0, c_1, c_2)$ where c_1, c_2 are the upper stress limits specified for vibration and temperature shock, respectively and c_0 is a dummy variable required to make the vector c consistent with the design matrix X and is equal to 1. The coded variables, in this example, are centered on the critical reliability boundary so that $c' = (1, 0, 0)$.

The warhead performance is measured in terms of depth of penetration t_w into monolithic armor, and S and A performance is measured by arming time t_f. The distribution of warhead penetration t_w and arming time t_f for the S and A is each distributed according to $N[\beta_0 + \beta_{1x1} + \beta_{2x2}, \sigma^2]$. The true values of the parameters are

	β_0	β_1	β_2	σ
Warhead	13''	-0.6	-0.4	1.5''
S and A	0.6 sec.	0.07	0.03	.033 sec.

These models thus assume that the average penetration decreases linearly with increasing stress and that the average arming time increases very slowly with increased stress within the region of interest. Thus, by (1), we see that the performance reliability for the warhead and S and A, respectively, are

$$(12) \qquad R_{WHD} = P\{t_w^{(c)} \geq 10''\} = \int_{10''}^{\infty} n(t_w^{(c)} \mid c'\beta, \sigma^2)\, dt_w^{(c)}$$

$$= g_1(c'\beta, \sigma; c) \equiv g_1(13'', 1.5''; c) = 0.977$$

$$(13) \qquad R_{S \text{ and } A} = P\left\{t_f^{(c)} \geq 0.5\right\} = \int_{0.5}^{\infty} n(t_f^{(c)} \mid c'\beta, \sigma^2)\, dt_f^{(c)}$$

$$= g_1(c'\beta, \sigma; c) \equiv g_1(0.6, .033; c) = 0.999 .$$

The performance reliability of the warhead section is thus

$$R = R_{WHD} \cdot R_{S \text{ and } A} = .976 .$$

Dud probabilities were not considered in this model. The evaluation of dud rates requires attribute test methods which are not as efficient as the variables plans and require much larger sample sizes. In conducting the type of test program described herein an estimate of the dud reliability may be made by noting the number of dud failures. However, useful interval estimation with these results may not be possible with reasonable confidence coefficients. When a dud occurs, it is desirable to repeat the appropriate test under the same conditions in order to avoid or minimize having to work with missing data in the test plan.

6.2 Multiple Regression Analysis. Multiple linear regression experimental designs of the type used in exploring response surfaces were used to evaluate the performance reliability of the warhead and S and A based upon the stated performance model. In particular, central composite rotatable experimental designs [3] were used. The experiments were conducted with sample sizes (n) of 8 and 30 for both the warhead and S and A. The treatment combinations and the responses generated by Monte Carlo simulation of the performance models are shown in Tables 1 to 4.

Least Squares estimates of the regression coefficients and error variance were made for the test results, and goodness-of-fit tests were conducted. In all four cases, a linear regression model was found to represent the data adequately. The least squares estimates of the regression coefficients obtained for each case are as follows:

Item	n	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
Warhead	8	13.99	-.175	.675
	30	12.75	-.975	.215
S and A	8	.59	.055	.020
	30	.60	.0705	.0205

Tests of significance at the .05 level performed for the regression coefficients gave the following results. The estimates of β_1 and β_2 for the warhead based on $n = 8$ were not significantly different from zero. For $n = 30$, $\hat{\beta}_2$ was not significantly different from zero, but $\hat{\beta}_1$ which corresponds to the effect of the vibration stress X_1 was found to be significantly different from zero, which corresponds to the true situation for the model. In the case of the S and A, $\hat{\beta}_2$(temperature shock) was not significantly different from zero, and $\hat{\beta}_1$(vibration) was significantly different from zero for n=8 and 30.

Point estimates of the performance reliability $R(\beta, \sigma^2)$ at $c'=(1, 0, 0)$ were made using the UMVU estimate $\tilde{K}(\beta, \sigma)$ of $K(\beta, \sigma)$. Exact one sided lower confidence limits using the non-central t-distribution were also obtained. A summary of these results is tabulated below, and a sample computation is given in Appendix C for one case. Estimates of $R(\beta, \sigma^2)$ based on the estimates $K(\hat{\beta}, \hat{\sigma})$ and $\tilde{K}(\beta, \sigma)$ are also included in the appendix.

Case	Item	n	$\tilde{R}$	R(.95)	True R
1	Warhead	8	.974	.832	.977
2		30	.967	.905	
3	S and A	8	.983	.891	.999
4		30	.997	.981	

where $\tilde{R}$ is the Estimate of Performance Reliability based on the UMVU estimate $\tilde{K}(\beta, \sigma)$ or $K(b, \sigma)$ and R(.95) is the one-sided lower 95% confidence limit for Performance Reliability.

A point estimate of the warhead section reliability is given by

$$\tilde{R} = \tilde{R}_{WHD} \cdot \tilde{R}_{S \text{ and } A} .$$

Conservative .90 confidence intervals for the warhead section reliability are obtained by multiplying the lower .95 confidence limits for the warhead and S and A. This result is easily proven by applying the Bonferroni inequality to obtain a conservative simultaneous confidence region T for R_{WHD} and $R_{S \text{ and } A}$ and by making use of the fact that the product is monotone in each variable. Thus, we obtain the following results for the warhead section reliability.

n	$\tilde{R}$	$R(\geq .90)$	True R
8	.957	.741	.976
30	.964	.888	

REFERENCES

1. Box, G. E. P. and Draper, N. R., "A Basis for the Selection of a Response Surface Design", J. Amer. Statist. Assoc., 54 (1959) 622-654.

2. Box, G. E. P. and Hunter, J. S., "Multifactor Experimental Designs for Exploring Response Surfaces", Ann. Math. Statist., 28 (1957) 195-241.

3. Einbinder, S. K. and Olkin, I., "Reliability Testing and Estimation for Single and Multiple Environments (Preliminary Report)", Proceedings of the Seventh Conference on the Design of Experiments in Army Research Development and Testing, U. S. Army Research Office, Report No. 62-2, August, 1962, 261-291.

4. Ghurye, S. G. and Olkin, I., "Unbiased Estimation of Probability Densities", Abstract, Annals. Math. Statist., 32 (1961) 1346.

5. Johnson, N. L. and Welch, B. L., "Applications of the Non-central t-Distribution", Biometrika, 31 (1940) 362-389.

6. Resnikoff, George J. and Lieberman, Gerald J., Tables of the Non-central t-Distribution, Stanford University Press, Stanford, California, 1957.

TABLE I

WARHEAD

EXPERIMENTAL DESIGN N=8

Sample	x_1	x_2	t_w
1	-1	-1	13.3
2	+1	-1	14.9
3	-1	+1	16.6
4	+1	+1	14.3
5	0	0	12.6
6	0	0	12.7
7	0	0	12.2
8	0	0	15.3

x_1 = Vibration

x_2 = Temperature Shock

t_w = Penetration (inches)

TABLE 2

WARHEAD EXPERIMENTAL DESIGN N=30

Sample	x_1	x_2	t_w
1	-1	-1	13.8
2	+1	-1	12.1
3	-1	+1	13.8
4	+1	+1	12.7
5	0	0	12.1
6	0	0	10.9
7	0	0	12.5
8	0	0	12.2
9	-1	-1	13.7
10	+1	-1	10.5
11	-1	+1	12.7
12	+1	+1	11.0
13	0	0	14.8
14	-1	-1	13.7
15	+1	-1	9.5
16	-1	+1	11.5
17	+1	+1	13.0
18	0	0	13.3
19	-1	-1	12.0
20	+1	-1	12.5
21	-1	+1	14.4
22	+1	+1	11.4
23	0	0	13.9
24	-1	-1	14.8
25	+1	-1	11.5
26	-1	+1	15.6
27	+1	+1	14.3
28	0	0	12.0
29	0	0	16.0
30	0	0	12.3

x_1 = Vibration

x_2 = Temperature Shock

t_w = Penetration (inches)

TABLE 3

S & A EXPERIMENTAL DESIGN N=8

Sample	x_1	x_2	Arming Time (seconds)
1	-1	-1	.48
2	+1	-1	.63
3	-1	+1	.56
4	+1	+1	.63
5	0	0	.61
6	0	0	.56
7	0	0	.60
8	0	0	.64

x_1 = Vibration

x_2 = Temperature Shock

TABLE 4

S & A EXPERIMENTAL DESIGN N=30

Sample	x_1	x_2	Arming Time (seconds)
1	-1	-1	.55
2	+1	-1	.63
3	-1	+1	.56
4	+1	+1	.71
5	0	0	.57
6	0	0	.68
7	0	0	.57
8	0	0	.60
9	-1	-1	.50
10	+1	-1	.66
11	-1	+1	.53
12	+1	+1	.73
13	0	0	.57
14	-1	-1	.58
15	+1	-1	.66
16	-1	+1	.57
17	+1	+1	.72
18	0	0	.55
19	-1	-1	.46
20	+1	-1	.64
21	-1	+1	.54
22	+1	+1	.65
23	0	0	.59
24	-1	-1	.51
25	+1	-1	.67
26	-1	+1	.56
27	+1	+1	.70
28	0	0	.53
29	0	0	.61
30	0	0	.60

x_1 = Vibration

x_2 = Temperature Shock

UNIVARIATE RESPONSE

No.	Response Vector Y	Design Matrix x_1	x_2	.	.	.	x_m
1	y_1	x_{11}	x_{21}	.	.	.	x_{m1}
2	y_2	x_{12}	x_{22}	.	.	.	x_{m2}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
i	y_i	x_{1i}	x_{2i}	.	.	.	x_{mi}
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
n	y_n	x_{1n}	x_{2n}	.	.	.	x_{mn}

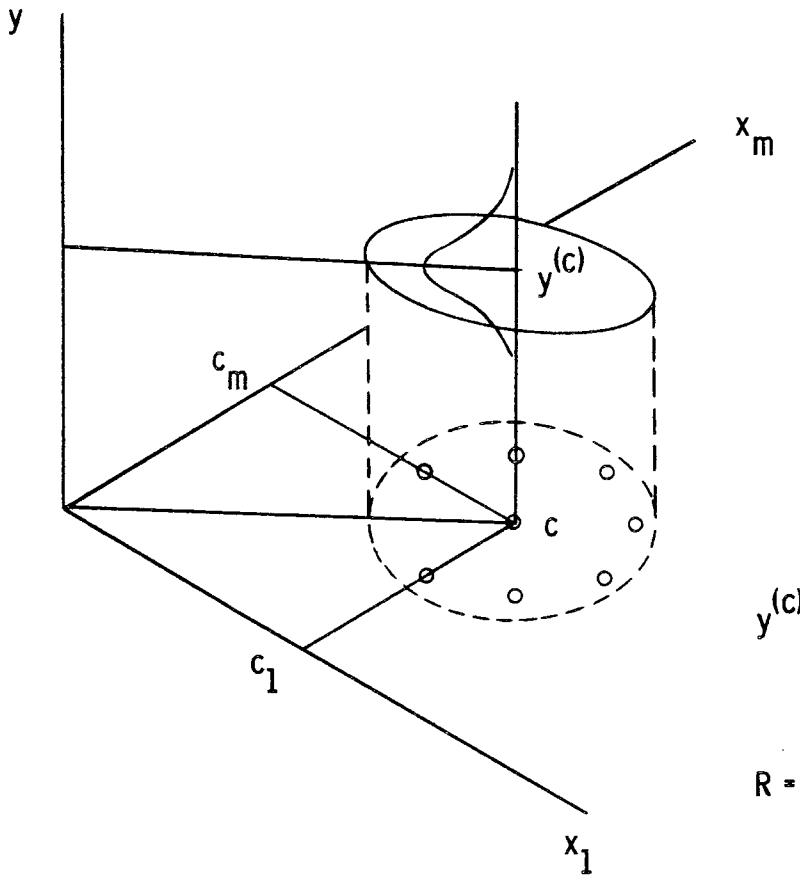
$$\Omega: \; y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{mi} + u_i, \; i = 1, \ldots, n$$

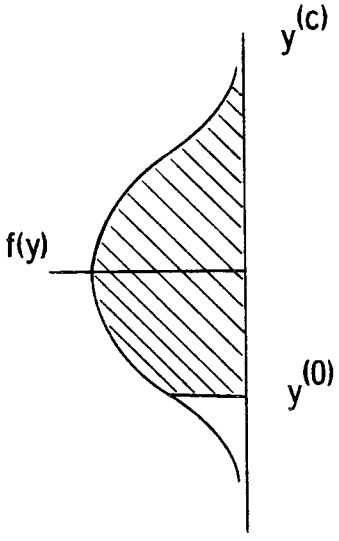
$\{u_1, \ldots, u_n\}$ are independently and identically distributed with mean 0 and variance σ^2.

Figure 1

PERFORMANCE RELIABILITY

UNIVARIATE RESPONSE





$$y^{(c)} \sim N[c_1\beta_1 + c_2\beta_2 + \cdots + c_m\beta_m, \sigma^2]$$

$$\equiv N[c'\beta, \sigma^2]$$

$$R = P\{Y^{(c)} \geq y^{(0)}\} = \int_{y^{(0)}}^{\infty} n(c'\beta, \sigma^2)\, dy^{(c)}$$

Figure 2

MULTIVARIATE RESPONSE

No	Response Matrix						Design Matrix					
	y_1	y_2	.	.	.	y_p	x_1	x_2	.	.	.	x_m
1	y_{11}	y_{21}	.	.	.	y_{p1}	x_{11}	x_{21}	.	.	.	x_{m1}
2	y_{12}	y_{22}	.	.	.	y_{p2}	x_{12}	x_{22}	.	.	.	x_{m2}
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
i	y_{1i}	y_{2i}	.	.	.	y_{pi}	x_{1i}	x_{2i}	.	.	.	x_{mi}
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
n	y_{1n}	y_{2n}	.	.	.	y_{pn}	x_{1n}	x_{2n}	.	.	.	x_{mn}

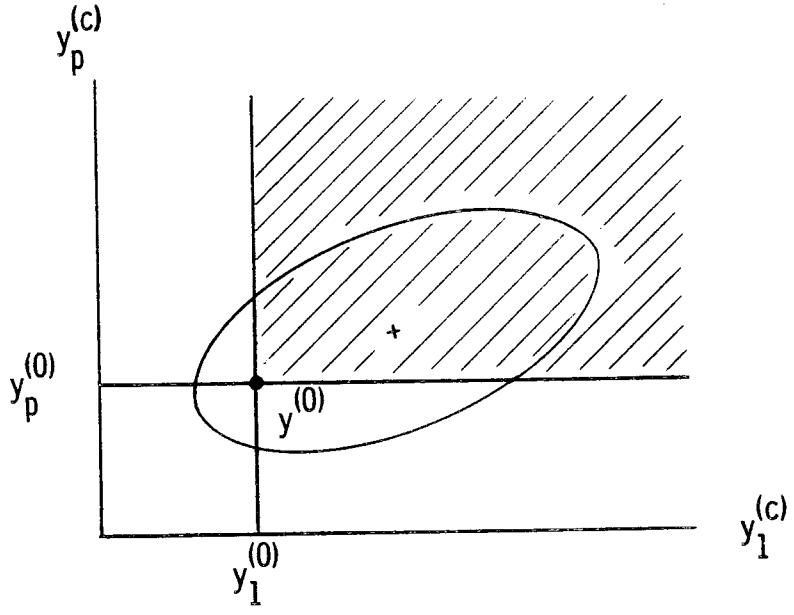
Ω : $Y = X' B + U$,

$\{u'_1, \ldots, u'_n\}$ are independently identically distributed with mean vector 0 and common p x p positive definite covariance matrix Σ .

Figure 3

PERFORMANCE RELIABILITY

MULTIVARIATE RESPONSE (VECTOR)



$$y^{(c)} \sim N[c' B, \Sigma] ,$$

$$R = P\{Y_1^{(c)} \geq y_1^{(0)}, \ldots, Y_p^{(c)} \geq y_p^{(0)}\}$$

$$= \int_{y_1^{(0)}}^{\infty} \cdots \int_{y_p^{(0)}}^{\infty} n(c' B, \Sigma)\, dy^{(c)}$$

Figure 4

APPENDIX A

Define $K = (y^{(0)} - c'\beta)/\sigma$ and $K(\hat{\beta},\hat{\sigma}) = (y^{(0)} - c'\hat{\beta})/\hat{\sigma}$. We first note that $\hat{\beta}$ and $\hat{\sigma}$ are independently distributed. Since $E\hat{\beta} = \beta$, we have that $E(y^{(0)} - c'\hat{\beta}) = (y^{(0)} - c'\beta)$. Also $v \equiv f\hat{\sigma}^2/\sigma^2$ has a χ^2_f distribution, $f = n - m$, so that

$$Ev^{-1/2} = \frac{\Gamma(\frac{f-1}{2})}{\sqrt{2}\ \Gamma(\frac{f}{2})} \quad .$$

Hence,

$$E\ \frac{\sqrt{2}}{\sqrt{f}}\ \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})}\ \frac{1}{\hat{\sigma}} = \frac{1}{\sigma}$$

which proves that $\tilde{K}(\beta,\sigma) = \frac{\sqrt{2}}{\sqrt{f}}\ \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})}\ K(\hat{\beta},\hat{\sigma})$ is an unbiased estimator of $K(\beta,\sigma)$. By completeness, it then follows that $\tilde{K}$ is the unique such estimator, and hence is UMVU.

An alternative approach is also useful, namely, that $K(\hat{\beta},\hat{\sigma})/\|a\| \equiv t(f,\delta)$ has a non-central t-distribution with f degrees of freedom and non-centrality parameter

$$\delta = \frac{y^{(0)} - c'\beta}{\sigma\|a\|} = \frac{K(\beta,\sigma)}{\|a\|} \quad ,$$

where $\|a\|^2 = c'(XX')^{-1}c$. To see this, we write

$$K(\hat{\beta},\hat{\sigma}) = \frac{\dfrac{y^{(0)} - c'\beta}{\sqrt{\mathrm{Var}(c'\hat{\beta})}} - \dfrac{c'\hat{\beta} - c'\beta}{\sqrt{\mathrm{Var}(c'\hat{\beta})}}}{\dfrac{\hat{\sigma}}{\sqrt{\mathrm{Var}(c'\hat{\beta})}}} .$$

But $c'\hat{\beta} = c'(XX')^{-1}Xy$, and hence $\mathrm{var}(c'\hat{\beta}) = \sigma^2 \; c'(XX')^{-1}c \equiv \sigma^2 \|a\|^2$. Thus,

$$\frac{K(\hat{\beta},\hat{\sigma})}{\|a\|} = \frac{\dfrac{y^{(0)} - c'\beta}{\sigma\|a\|} - \dfrac{c'\hat{\beta} - c'\beta}{\sigma\|a\|}}{\dfrac{\hat{\sigma}}{\sigma}} = t(f,\delta) .$$

By noting that $Et(f,\delta) = \delta \sqrt{f/2} \;\; \Gamma(\frac{f-1}{2})/\Gamma(\frac{f}{2})$, we can also obtain $E\,\tilde{K}(\beta,\sigma) = K(\beta,\sigma)$.

APPENDIX B

Since $R(\hat{\beta},\hat{\sigma}^2)$ is a function of the sample moments, it follows that $R(\hat{\beta},\hat{\sigma}^2)$ is asymptotically normal with mean $R(\beta,\sigma^2)$ and variance

$$\sum_{i,j} \left(\frac{\partial R}{\partial \beta_i} \bigg|_{\beta,\sigma^2} \right) \left(\frac{\partial R}{\partial \beta_j} \bigg|_{\beta,\sigma^2} \right) \mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) + \left(\frac{\partial R}{\partial \hat{\sigma}^2} \bigg|_{\beta,\sigma^2} \right) \mathrm{Var}(\hat{\sigma}^2) .$$

The cross-product terms involving $\hat{\beta}_i$ and $\hat{\sigma}^2$ drop out because of the independence of $\hat{\beta}$ and $\hat{\sigma}^2$. From

$$R(b,s^2) = \int_{(y^{(0)} - c'b)/s}^{\infty} (2\pi)^{-1/2} \exp(-1/2\; t^2)\, dt ,$$

we obtain

$$\frac{\partial R}{\partial b_i} = \frac{c_i}{\sqrt{2\pi}\, s} \exp[-1/2 \frac{(y^{(0)} - c'b)^2}{s^2}], \quad \frac{\partial R}{\partial s^2} = \frac{(y^{(0)} - c'b)}{2s^2} \frac{1}{\sqrt{2\pi}\, s} \exp[-1/2 \frac{(y^{(0)} - c'b)^2}{s^2}]$$

Also, $\mathrm{Cov}(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 a_{ij}$, where $A = (XX')^{-1}$, $\mathrm{Var}(\hat{\sigma}^2) = 2\sigma^4/f$, and hence the asymptotic variance is

$$V_\infty(\beta, \sigma^2) = \sigma^2 [n(y^{(0)} \mid c'\beta;\ \sigma^2)]^2 \{c'(XX')^{-1} c + \frac{(y^{(0)} - c'\beta)^2}{2\sigma^4 f}\} .$$

But, $V_\infty(\hat{\beta}, \hat{\sigma}^2)$ is a rational function of the sample moments, so that, by Slutsky's Theorem, $V_\infty(\hat{\beta}, \hat{\sigma}^2)$ converges in probability to $V_\infty(\beta, \sigma^2)$, and hence

$$\frac{R(\hat{\beta}, \hat{\sigma}^2) - R(\beta, \sigma^2)}{\sqrt{V_\infty(\hat{\beta}, \hat{\sigma}^2)}} \rightarrow N(0, 1) .$$

APPENDIX C

The computation of the point and confidence interval estimates for the performance reliability of the warhead for the sample size $n = 8$ is described in this appendix. From the test data in Table I, we obtain the following results:

Point Estimation

$$\hat{\sigma}^2 = 2.958, \quad c' = (1, 0, 0), \quad f = n - m = 8 - 3 = 5$$

$$K(\hat{\beta}, \hat{\sigma}) = \frac{y^{(0)} - c'\hat{\beta}}{\hat{\sigma}} = \frac{10 - 13.99}{1.720} = -2.320$$

$$\tilde{K}(\beta, \sigma) = K(\hat{\beta}, \hat{\sigma}) \sqrt{\frac{f}{2}} \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})} = -1.95$$

Substituting, these two estimates of $K(\beta,\sigma)$ in

$$R(\beta,\sigma^2) = \int_{K(\beta,\sigma)}^{\infty} (2\pi)^{-1/2} e^{-y^2/2}\, dy$$

gives the two estimates of reliability $R(\hat{\beta},\hat{\sigma})$ and $\tilde{R}(\beta,\sigma)$, respectively. Thus,

$$R(\hat{\beta},\hat{\sigma}) = \int_{-2.320}^{\infty} (2\pi)^{-1/2} e^{-y^2/2}\, dy = .9898 \text{ , and}$$

$$\tilde{R}(\beta,\sigma) = \int_{-1.95}^{\infty} (2\pi)^{-1/2} e^{-y^2/2}\, dy = .974 .$$

Confidence Intervals

$$\|a\|^2 = c'(XX')^{-1}c = \frac{1}{n} \text{ for } c' = (1, 0, 0) .$$

Following the notation of Resnikoff and Lieberman, a confidence interval may be obtained using the non-central t-tables in [6]. The percentage points of t are denoted by $x(f,\delta,\epsilon)$ where x is the value such that $P\{\frac{t}{\sqrt{f}} > x \mid f,\delta\} = \epsilon$.

$$x = \frac{K(\hat{\beta},\hat{\sigma})}{\sqrt{f}\,\|a\|} = \sqrt{\frac{n}{f}}\, K(\hat{\beta},\hat{\sigma}) = \sqrt{\frac{8}{5}}\,(-2.320) = -2.935 .$$

The one sided lower .95 confidence limit for R is obtained by finding the corresponding limit for $K(\beta,\sigma)$ because of the monotone relation between R and $K(\beta,\sigma)$. The $1-\alpha$ confidence limit for $K(\beta,\sigma)$ is obtained by solving

$$x(n - m, \delta_{1-\alpha}, 1-\alpha) = \frac{K(\beta,\sigma)}{\sqrt{f}\ \|a\|} \quad ,$$

$$x(5, \delta_{.95}, .95) = -2.935 \quad .$$

Making use of the relation $x(f,\delta,\epsilon) = -x(f,-\delta,1-\epsilon)$, we obtain

$$x(5, -\delta_{.95}, .05) = 2.935 \quad .$$

From the Resnikoff-Lieberman table of percentage points of t, we obtain a non-centrality value $\delta = \sqrt{f} + 1\ K_p = \sqrt{6}\ (1.107) = 2.712$ by interpolating on K_p. Since $\delta = \frac{K(\beta,\sigma)}{\|a\|}$, the .95 lower confidence limit for $K(\beta,\sigma)$ is

$$\underset{.95}{K(\beta,\sigma)} = \|a\|\ \delta_{.95} = \frac{\delta_{.95}}{\sqrt{n}} = \frac{-2.712}{\sqrt{8}} = -.959$$

Finally, the .95 confidence limit for R is

$$\underset{.95}{R(\beta,\sigma^2)} = \int_{\underset{.95}{K(\beta,\sigma)}}^{\infty} (2\pi)^{-1/2} e^{-y^2/2}\, dy = .832 \quad .$$

$\delta_{.95}$ may also be computed using the Johnson-Welch table IV and following the procedure on page 372 of [5].

This page intentionally left blank.

EVALUATION OF VARIOUS LABORATORY METHODS FOR DETERMINING RELIABILITY

A. Bulfinch
Reliability Concepts Section, Reliability Branch
Quality Assurance Division
Picatinny Arsenal

ABSTRACT. In preparation for evaluating several laboratory methods for determining reliability, an operational definition of reliability has been described in terms of stress and strength. The definition was verified by three independent methods of calculation.

Two classes of laboratory methods have been evaluated by means of Monte Carlo Sampling Procedures:

1. Testing-Without-Failure
2. Testing-With-Failure

The former type was found to be worthless or of limited value. The latter type was found to be capable of accurate and precise determination of reliability values at any level with small sample sizes.

1. INTRODUCTION. This investigation is concerned with the laboratory determination of ultimate* functional reliability with respect to single environments of components from systems to be used only once. To this end, Monte Carlo Experiments have been conducted to evaluate the precision, accuracy, and efficiency of two types of methods:

1. Methods which test-without-failure
2. Methods which test-to-failure.

In preparation for conducting these experiments, consideration was first given to establishing a valid procedure for calculating the true, ultimate[a] reliability of known conditions from the operational definition of reliability described below.

II. OPERATIONAL DEFINITION OF RELIABILITY. The theoretical definition of reliability, the probability of success under specified conditions, gives no clue as to how reliability can be experimentally determined. This

*Ultimate in this sense means maximum or unbiased value.

question, of course, must be answered before an experiment to measure reliability can be conducted. That is, the theoretical concept of reliability must be translated into an operational definition in terms of measurable properties and conditions. The following procedure is proposed as such a definition of the true, ultimate reliability of items to be used only once. The objective of the procedure is to measure the probability of failure--the complement of the probability of success.

The concept of reliability can be thought of in terms of stress and strength.* The relationship between these two elements can be represented graphically by two distributions on a single continuous scale: one (on the left) for the specified condition which will be called the applied stress curve and the other (on the right) for the property of an item which will be called the strength curve. The distance between the means of these distributions represents the margin of safety, or the extra strength built into the item during development. If the mean stress equals the mean strength, then the two curves are superimposed and the reliability is equal to 50%.

It is reasonable to assume that an item will not fail unless the applied stress equals or exceeds the item's strength. Referring to the two distributions described above, the probability of a strength value less than any particular stress value can be measured by the area under that portion of the strength curve which represents items with strengths less than that stress--that is, the area of the strength curve to the left of the stress ordinate. The probability of a particular stress value occurring can be measured by a small increment of area between two ordinates on the stress curve. In order to produce a failure in this manner, the higher stress value must occur simultaneously with the lower strength value. To determine the probability of such a combination of independent events occurring, the product of the probabilities of the separate events occurring must be obtained. There are many such mutually exclusive combinations possible. To obtain the total probability of failure, the sum of all the possible product must be calculated.

Defining the complement of reliability in this manner appears to be logically sound. The mathematics involved is based on well-known laws of probability. Therefore, the complement of the sum of the products of the probabilities of the stress equaling or exceeding the strength can be

*Strength here means the ability to withstand environmental stresses.

taken as the operational definition of the true, ultimate reliability of items used only once.

It would appear that combinations of conditions of this kind can represent, in an elementary way, the cause of failure when a component experiences an environment in actual use. A reliability obtained in this manner then has practical value and can be accepted as the solution to the basic problem of obtaining a valid, unbiased numerical value for reliability.

III. VERIFICATION OF THE OPERATIONAL DEFINITION.

A. Numerical Integration:

Having arrived at these conclusions the following combination of conditions was chosen for use:

Stress	U_1 = 10 (True average stress)
	S_1 = 5 (True std. dev. of stress)
Strength	U_2 = 20 (True average strength)
	S_2 = 5 (True std. dev. of strength)

Both distributions were assumed to be normal.

The reliability associated with this condition was calculated by the numerical integration procedure (calculating the sum of products) described above. The following results were obtained:

STRESS INCREMENT	RELIABILITY
1.0 Unit	.9362
0.5 Unit	.9264

B. Modified Normal Deviate:

A second method of determining the reliability value for the above condition which has its foundation for validity in statistical theory, is the following formula suggested by Dr. William S. Connor of the Research Triangle Institute:

$$Z = \frac{(X_1 - X_2) - (U_1 - U_2)}{\sqrt{S_1^2 + S_2^2}}$$

A failure can occur only when $X_1 \geq X_2$, where X_1 is any stress value and X_2 is any strength value. For this situation the formula becomes:

$$Z \geq \frac{U_2 - U_1}{\sqrt{S_1^2 + S_2^2}}$$

Substituting the above values assigned to U_1, U_2, S_1, and S_2:

$$Z \geq \frac{20 - 10}{\sqrt{25 + 25}} = \frac{10}{\sqrt{50}} = 1.414$$

Entering a table of areas under the standard normal curve, the following value can be obtained.

$$R \geq 0.9213$$

C. Monte Carlo Sampling:

A third method of determining the reliability value for the above condition is the Monte Carlo sampling of both distributions. This method is acceptable as valid on reasonable and logical grounds also. The stress-strength combinations that occur in actual use-conditions result from a random selection of such combinations. This is the basis of the Monte Carlo procedure described below which will be called the Flight-Condition Procedure.

Pairs of applied stress values and assumed strength values were chosen at random from their respective distributions by means of tables of random normal numbers from these distributions. The values in each pair were compared to determine whether the stress exceeded the strength. In those cases where this was true, the result of the observation was declared a failure.

The average reliability value obtained from 7,500 observations with the Flight-Condition Procedure was found to be R = 0.9221.

D. Summary of Results:

Method	Reliability
Numerical Integration	0.9264
Z-Formula	0.9213
Monte Carlo Sampling (Flight-Condition Simulation)	0.9221
Grand Average	0.9233

E. Conclusions of the Verification:

The close agreement of the values obtained by these three independent methods, each of which can be separately accepted as valid, is considered ample justification for concluding that:

1. The operational definition of reliability described above can be taken as valid.
2. The true-ultimate reliability for the condition described above is very close to the average of these three values: 0.9233.
3. The results obtained by the Z-Formula are sufficiently accurate to justify using this formula as the primary standard procedure for calculating the reliability of "one-shot" items under the above operational definition of reliability.

IV. EVALUATION OF LABORATORY METHODS.

A. Introduction:

In view of the above conclusions, the Z-Formula has been used to calculate the true-ultimate reliability values for the known stress-strength conditions used in the Monte Carlo Experiments that follow. These reliability values were used to compare with estimates obtained from the several laboratory methods being evaluated. The magnitude of the differences found between the true and estimated values were used to evaluate the accuracy of the results obtained from these methods.

The purpose of the Monte Carlo Experiments is to determine the accuracy, precision, and efficiency of various laboratory methods. This is done by Monte Carlo sampling of known distributions of stress and strength combinations.

B. Testing-Without-Failure Procedures:

1. Single Stress Level Method:

In practice, this method consists of applying the same environmental stress level to each test specimen comprising the sample and counting the number of failures obtained. The level of stress applied is usually one expected to be experienced in use. In the Monte Carlo Experiment, this method could produce a failure only when the value obtained from a table of random numbers was found to be equal to or less than the assumed stress value. The combination of conditions used for this experiment is:

Stress	Strength
$U_1 = 10$	$U_2 = 20$
$S_1 = 5$	$S_2 = 5$

In conducting this Monte Carlo Experiment, the following assumptions were made:

(1) The stress being applied was 10 units.

(2) The tabled values of random numbers used contained a decimal point to the left of the first digit and that, as such, these decimal values represent the probability associated with the strength values used.

(3) A failure was obtained only when a tabled value was found to be less than 0.0228 - the probability of a strength value equal to or less than 10 units.

Sample sizes of 22, 35, 45, and 230, were used. One hundred samples of each size were taken. The purpose here is to determine the effect of increasing the sample size in addition to the purpose stated above.

a. Results:

TABLE I

SINGLE STRESS LEVEL METHOD

True Probability of Failure = 0.0787

Sample Size = 22

Failures No.	Proportion		Frequency, Per cent
0	0.000	Mode	73
1	.045		23
2	.091		4

For : 0/20 : 90%(1 - sided) upper limit for defects = .099.

The true probability of failure is included in the 90% one-sided upper confidence limit.

TABLE II

SINGLE STRESS LEVEL METHOD

True Probability of Failure = 0.0787

Sample Size = 35

Failures No.	Proportion		Frequency, Per cent
0	0.000	Mode	57
1	.029		35
2	.057		7
3	.086		1

For 0/35 : 90% (1 - sided) upper limit for defects = .063

95% (1 - sided) upper limit for defects = .082

The true value is included in only the 95% upper limit.

TABLE III

SINGLE STRESS LEVEL METHOD

True Probability of Failure = 0.0787

Sample Size = 45

Failures No.	Proportion		Frequency, Per cent
0	0.000	Mode	50
1	.022		36
2	.044		14

For 0/45: 90% (1 - sided) upper limit for defects = .049

95% (1 - sided) upper limit for defects = .064

99% (1 - sided) upper limit for defects = .097

The true value is included in only the 99% upper limit.

TABLE IV

SINGLE STRESS LEVEL METHOD

True Probability of Failure = 0.0787

Sample Size = 230

Failures No.	Proportion		Frequency, Per cent
0	0.000		0
1	.004		4
2	.008	80%(2-sided) lower limit	10
3	.013		7
4	.017	Mode	21
5	.022		17
6	.026		14
7	.030		11
8	.035	80%(2-sided) upper limit	7
9	.039		4
10	.043		4
11	.048		0
12	.052		0
13	.056		1

4/230: 99.5% (1 - sided) upper limit for defects = .053
The true value is not included in even the 99.5% upper limit.

TABLE V

SUMMARY

SINGLE STRESS LEVEL METHOD

True Probability of Failure: P = 0.0787

True Reliability: R = 0.9213

EFFECT OF SAMPLE SIZE

Sample Size	Average	
	R	P
22	.987	.013
35	.985	.015
45	.986	.014
230	.977	.023

b. Conclusions:

The following conclusions can be drawn from the above data about the method which applies a single stress level under the use conditions:

(1) This method produces biased results.

(2) Increasing the sample size did not remove the bias or error.

(3) The confidence interval is worthless for locating the true value.

(4) The ultimate value of the failure rate being measured by this method is 0.0228, the probability of a strength value equal to or less than 10 units - not the true probability of failure as previously defined.

(5) The possibility of a stress value exceeding a strength value greater than 10 units is completely ignored by this method,

(6) This method is worthless to determine reliability as defined above.

2. Multiple-Stress Level Method:

The flight-condition procedure used earlier as one of the intuitively acceptable basic procedures for determining true reliability, is one which also readily lends itself to laboratory use with small samples. For this purpose, a different randomly selected stress is used for each test specimen.

a. Results:

Using the Flight-Condition Procedure in the Monte Carlo Experiment with sample size of 50, the following results were obtained:

TABLE VI

MULTIPLE - STRESS LEVEL METHOD

(Low Reliability)

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 20$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9213

Sample Size = 50

DISTRIBUTION OF FAILURES FOR SAMPLES OF 50

Number of Failures	Frequency	
0	2	
1	6	
2	28	
3	27	
4	39	Mode
5	25	
6	11	
7	5	
8	4	
9	3	
Total Number of Sample Results:	150	

Average Reliability = 0.9224

Standard Deviation = 0.0376

TABLE VII

MULTIPLE - STRESS LEVEL METHOD

(High Reliability)

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 30$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9977

Sample Size = 50

DISTRIBUTION OF FAILURES FOR SAMPLES OF 50

Number of Failures	Frequency	
0	90	Mode
1	9	
2	1	
3	0	
Total Number of Sample Results	100	

Average Reliability = 0.9978

Standard Deviation = 0.0069

b. Conclusions:

Again the Flight-Condition Procedure produced an accurate, unbiased average value. However, in the case of high reliability, this procedure produced no failures in 90% of the samples of 50. In addition, no results could be obtained between 1.00 and .98. This is a distinct disadvantage. In practice only one small sample is usually taken. If no failures are obtained, the result is not only biased but it is worthless for mathematical manipulations of any kind. From this it is concluded that testing under the expected stress conditions leaves something to be desired in determining high reliabilities. This type of method is too insensitive to changes in high reliabilities to be of use with small samples.

V. TESTING-TO-FAILURE PROCEDURES.

A. Introduction:

An alternative approach for determining reliability is "Testing-to-Failure". Instead of applying the stress or stresses expected in use, systemstically increase the level of severity of the stress until failure occurs. Methods for this purpose are available which will generate the ultimate strength distribution. From the results of this type of testing, the average and standard deviation of the ultimate strength can be determined. By means of independent experiments or prior knowledge, the average and standard deviation of the expected stress distribution can be obtained.

1. T-Formula:

With these two sets of data available, reliability, as previously defined, can be estimated from the Z-Formula by replacing the population means and variances with their sample estimates:

$$T \geq \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2 + s_2^2}}$$

where:

$\bar{X}_2$ = Observed Average Ultimate Strength

$\bar{X}_1$ = Observed Average Expected Stress

s_1^2 = Observed Variance of the Expected Stress

s_2^2 = Observed Variance of the Ultimate Strength.

Using tables of areas under the normal curve to obtain reliability values from the T-Formula for samples as small as 20, causes an error in the result. However, this error is small -- less than one per cent.

2. Need for Indirect Methods:

Testing-to-Failure Procedures can be easily performed when the measure of a simple property, such as tensile strength, is required. This is the ideal situation in which:

a. The occurrence of a failure can be detected by visual inspection.

b. The magnitude of the applied stress, at the point of failure, is directly observable.

The results obtained in a test of this kind, permit the calculation of the average and variance of the ultimate strength directly from the observed data. These values can then be used in the T-Formula to calculate the reliability with respect to tensile load.

To measure more complicated properties, such as the integrity of a moisture seal after vibration, indirect methods must be used to obtain the average and variance of the ultimate strength. In this case:

a. The occurrence of a failure cannot be detected by visual inspection. A failure must be determined by means of the results obtained from conducting an appropriate test.

b. The magnitude of the stress at the point of failure is not directly observable. The stress at the point of failure must be calculated.

The observed data in this kind of test is simply success or failure. As the stress is applied at higher and higher levels, higher and higher proportions of the test specimen will fail. If the stress is increased step-wise, and the proportion of failures determined at several points between 0 and 100% failure, the familiar sigmoid-curve of a cumulative frequency distribution can be obtained from these results.

The average and variance of the ultimate strength can be determined from this curve. The average value is the stress level causing 50% failures. The variance is the square of the reciprocal of the slope of this curve at the 50% point.

3. Characteristics of Available Methods:

The class of methods that can generate curves of this type are called Tests-of-Increased Severity. Some of these methods are:

a. The Run-Down
b. The Two-Stimuli
c. The Up-and-Down

The Run-Down Method can be used to determine the shape of the sigmoid-curve. But, it requires the largest number of test specimens.

The Two-Stimuli Method requires the assumption of normality. It is:

a. An abreviated Run-Down Method

b. Highly efficient when the above assumption is valid, since it takes a minimum number of test specimens

c. Easy to conduct

d. Easy to calculate the average and variance.

The Up-And-Down Method is highly efficient for determining the 50% point. However, it gives a poor estimate of the variance without large sample sizes. In addition, this method is difficult to conduct and requires lengthy grouped data calculations to obtain the average and variance.

4. Two-Stimuli vs Up-And-Down Method:

In view of these characteristics and the following data, the Two-Stimuli Method was chosen for conducting most of the initial work on this class of methods.

a. Results:

TABLE VIII

Test-to-Failure Procedures

Conditions Used:	STRESS	STRENGTH	
	$U_1 = 10$	$U_2 = 20$	(True averages)
	$S_1 = 5$	$S_2 = 5$	(True standard deviations)

True Reliability = .9213

METHOD	TOTAL NO. OF TRIALS	AVG.	STANDARD DEVIATION	RELIABILITY
Up-and-Down	2700	19.8	4.88	.9197
Two-Stimuli	1600	20.0	5.03	.9207

b. Conclusions:

Any observed differences in the averages or standard deviations are highly significant for these large sample sizes. Therefore, basing a choice on the absolute magnitude of the observed differences, it can be said that the Two-Stimuli Method is slightly more accurate and precise than the Up-and-Down Method.

B. Two-Stimuli Method:

1. Description:

The Two-Stimuli Method was conducted as follows:

Starting at an assumed stress value of 10 units, the first tabled value (converted to a decimal by placing a decimal point to the left of the first digit) was compared to .0228. If the tabled value was found

to be greater than .0228, the result was declared a success. Then a second tabled value (as a decimal) was chosen and compared to .0668, the probability value associated with a stress of 12.5 units. If the tabled value was greater than .0668, the result was again declared a success. Then a third tabled value was compared with .1587, the probability associated with a stress of 15 units. This process was continued using increments of stress equal to $s_1/2$, until the tabled value (as a decimal) was found to be equal to or less than the probability value associated with the assumed stress. When such a comparison was found, the result was declared a failure. Only half of the cases of exact equality were declared failures. At this level of stress, 19 additional observations were made by the comparison process just described. The proportion of failures at this level of stress was calculated and recorded. If this proportion was less than 50%, then the stress level was increased by two or three increments. If the proportion of failures was greater than 50%, then the stress level was decreased by two or three increments. A total of 20 observations was made, as before, at this latter stress level; the proportion of failures calculated and the result recorded. Only proportions greater than 0 and less than 100%, and which differ by 20% or more, are useful for this purpose. The average and standard deviation of the strength curve was calculated from these two failure proportions by means of the equation for the standard normal cumulative frequency curve.

2. Results:

The following results were obtained using the Two-Stimuli Method:

TABLE IX

TWO-STIMULI METHOD
(Low Reliability)

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 20$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9213

Sample Size = 45

DISTRIBUTION OF RESULTS FROM SAMPLES OF 45

Reliability Cell - Width	Frequency	
.990-.970		
.969-.950	17	
.949-.930	42	Mode
.929-.910	25	
.909-.890	32	
.889-.870	14	
.869-.850	12	
.849-.830	3	
.829-.810	1	
.809-.790	3	
.789-.770	1	
Total Number Of Sample Results	150	

Average Reliability = 0.9136

Standard Deviation = 0.0369

TABLE X

TWO-STIMULI METHOD

(High Reliability)

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 30$
	$S_1 = 5$	$S_2 = 5$

True Reliability = .9977 Sample Size = 45

DISTRIBUTION OF RESULTS FROM SAMPLES OF 45

Reliability Cell-Width	Frequency
.9999 - .9998	2
.9997 - .9990	41
.998 - .995	74 (Mode)
.994 - .991	20
.990 - .987	6
.986 - .983	2
.982 - .979	0
.978 - .975	1
.974 - .971	0
.970 - .967	2
.966 - .963	1
.962 - .959	0
.958 - .955	0
.954 - .950	1
TOTAL NUMBER OF SAMPLE RESULTS	150

Average Reliability = 0.9952

Standard Deviation = 0.0064

3. Conclusions:

It is clear from the above results that estimates of reliabilities at any level can be obtained with acceptable accuracy and precision from small samples when using the Two-Stimuli Method. This method is sensitive to changes that actually take place in either low or high reliability conditions. In principle then, this type of method is the solution to the problem of measuring high reliabilities with small sample sizes.

VI. SUMMARY:

A. Comparison of Testing-With-Failure with the best of the Testing-Without-Failure Methods.

TABLE XI

1. LOW RELIABILITY

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 20$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9213

Sample Size: Two-Stimuli Method, N = 45
Flight Condition Method, N = 50

DISTRIBUTION RELIABILITY RESULTS

Reliability Cell Midpoints	Two-Stimuli	Flight Condition
1.00	0	2
.98	0	6
.96	17	28
.94	42 (Mode)	27
.92	25	39 (Mode)
.90	32	25
.88	14	11
.86	12	5
.84	3	4
.82	1	3
.80	3	0
.78	1	0
Average Reliability =	.9136	.9224
Standard Deviation =	.0369	.0366

TABLE XII

2. HIGH RELIABILITY

Condition:	STRESS	STRENGTH
	$U_1 = 10$	$U_2 = 30$
	$S_1 = 5$	$S_2 = 5$

True Reliability = 0.9977
Sample Size: Two-Stimuli Method, H = 45
Flight-Condition Method, N = 50

DISTRIBUTION OF RELIABILITY RESULTS

Reliability Cell-Width	Flight-Condition Method	Two-Stimuli Method
1.0000 - .9998	90 (Mode)	2
.9997 - .9990	0	41
.998 - .995	0	74 (Mode)
.994 - .991	0	20
.990 - .987	0	6
.986 - .983	0	2
.982 - .979	9	0
.978 - .975	0	1
.974 - .951	0	0
.970 - .967	0	2
.966 - .963	0	1
.962 - .959	1	0
.958 - .955	0	0
.954 - .951	0	1
TOTAL NUMBER OF SAMPLE RESULTS:	100	150
Average Reliability	0.9978	0.9952
Standard Deviation	0.0069	0.0064

B. Conclusions:

There is little to choose between these two methods for reliability values equal to or less than .9213. However, for reliability values as high as .9977, the Two-Stimuli Method appears to be superior for practical use when numerical values for reliability are required.

It can be seen from Table X that the Two-Stimuli Method never produces a result of 1.000 for reliability. This is in contrast to the Flight-Condition Procedure which produced a reliability value of 1.000 in 90% of samples of 50 (Table VII). Seventy-eight percent of the Two-Stimuli Method results are within ≠ 0.002 of the true value for the high reliability condition. In the Flight-Condition Procedure, no result can be obtained between 1.000 and .9800 when a sample size of 50 is used.

VII. CONCLUSIONS:

1. For practical engineering purposes, reliability can be defined in terms of stress and strength. From this, three simple conclusions follow:

a. Reliability is created by the strength built into an item.

b. An item cannot fail until the stress equals or exceeds the strength.

c. To increase the reliability above 50%, the strength must exceed the stress.

2. Reliability is a relative property that depends upon the environmental stresses to be experienced (in use). To state a numerical value for this property requires a careful definition of the expected stresses.

3. The magnitude of a reliability value does not depend upon the number of items tested. Increasing the sample size can only improve the precision with which the reliability value is known. This improvement can be obtained only if the method of testing produces an unbiased estimate of reliability.

4. Applying only a single level of the expected stress to all of the test specimens cannot determine reliability at any level with an acceptable degree of accuracy and precision.

5. Applying a different randomly chosen stress level, from the expected stress distribution, to each test specimen used, produces unbiased results and is only useful for determining low reliability values.

This procedure cannot determine high reliabilities with small sample sizes with an acceptable degree of precision.

6. The Two-Stimuli Method can determine reliability at any level with small sample sizes with an acceptable degree of accuracy and precision.

COMPUTER SIMULATIONS IN RELIABILITY

Jasper Dowling
U. S. Army Munitions Command, Picatinny Arsenal

ABSTRACT. The troublesome problems of calculating a realistic lower confidence limit for systems reliability from component results and writing algebraic probability expressions for complex systems have been investigated. Practical Monte Carlo procedures for routine use of high speed computers are described. Iterative procedures are explained which can:

a. Save time in calculating the lower confidence limits of complex systems.
b. Obtain point estimates of complex systems from Boolean expressions when the writing of algebraic equations is too difficult.

INTRODUCTION. This paper deals with two of the problem areas encountered in obtaining estimates of the reliability of a weapon system from component test data. The accuracy of the reliability estimate of a system is a function of the accuracies of the component reliability estimates. The calculation of the lower bound of the system estimate, or lower confidence limit, is one of the problem areas and is the first of the two discussed. The remainder of this paper explains a computer simulation technique for evaluating system Boolean expressions where algebraic probability expressions are not feasible.

CONCLUSION.

1. It is concluded that the Monte Carlo technique is a valid and practical method of obtaining the lower confidence limit of a system. A 90 % lower confidence limit of O. 88 obtained with the Monte Carlo method corresponds with that obtained by Garner and Vail (Reference 2) using a three component system in a series configuration with component reliabilities of O. 96, O. 97, and O. 99. Their 95 % lower confidence limit of O. 88 using a different technique corresponded with the lower limit of O. 88 obtained with the Monte Carlo method.

2. A five component system of known system reliability was used as a standard in the second part of this report. The reliability of this known system was O. 98. By using the formula

$$\bar{P} \pm 2\sqrt{\frac{pq}{n}}$$

it was determined with 95 % assurance that a Monte Carlo sample size of 78,400 would be needed to obtain values of 0.98 ± .0005. The system was run and resultant values of 0.9796 and 0.9799 were obtained. Although this is only an example it is concluded that the application of the simulation is valid, and hence the procedure is a practical, useful one.

MAIN DISCUSSION

1. To begin the discussion of the lower confidence limit let us assume a three component system in a series configuration. See Figure 1[1]. These components are assumed to be independent. This system reliability is 0.96 x 0.97 x 0.99 = 0.92 (to two decimal places). In this example each success ratio and hence, each binomial probability distribution is depicted for a sample size of 100. By using the binomial probability law we can compute a 90 % lower confidence limit for each of the components, as 0.92, 0.94, and 0.96 respectively. Their product 0.92 x 0.94 x 0.96 = 0.83. This is not a 90 % confidence limit of the system.

2. To achieve the desired system distribution a Monte Carlo technique can be used to perform the product of the three component distributions. The probability of choosing a particular value from the component distribution will have to be equal to the distribution's probability (ordinate). To obtain this with a uniform random number generator, the three component binomial distributions are put into cumulative distributions. See Figure 2. A random number x, $0 \leq x \leq 1$ is generated. If this random number is less than the first value on the cumulative distribution (lowest ordinate) the value of the abscissa at this ordinate is assigned to the value of this component. If the random number is greater than the first ordinate, it is compared to the second ordinate. This continues until an ordinate is greater than the random number. The corresponding abscissa is assigned the value of the component. This is done for each of the three components and the three assigned values are substituted into the system equation and the equation solved. This is one point of the system distribution. As an example, assume the three random numbers .3321645, .2168429O, and .9316420O are chosen. The corresponding reliabilities would be 0.95, 0.96, 1.0. The point on the system distribution would be the product of these three or 0.912O. This procedure is repeated many times to form the system distribution.

[1] All figures are contained in the appendix.

3. Figure 3 is the system distribution of Figure 1. This distribution is based on 5,000 points (repetitions) with a mean of 0.92, a standard deviation of 0.026 and a 90 % lower confidence limit of 0.88. This lower confidence limit can be obtained by assuming normality and calculating the limit or by counting the lowest 500 points (10 % of the total) of the system frequency distribution .

4. Figure 4 represents a 22 component system. Each component is represented by its binomial probability distribution based on a sample size of 100. It should be noted that although there are a number of similar components to be assigned the same probability, the algebraic equation must allow for 22 independent components. Assigning similar components the same Monte Carlo or simulated value will cause both higher and lower system values and hence an excessively high variance and a wider frequency distribution. For accuracy in counting the lowest cells for determining a counted confidence limit, the system frequency distribution is collected in small cell intervals and grouped after counting.

5. Figure 5 is the system distribution of Figure 4 based on 5,000 points. The mean value of this 22 component system is 0.98 and the lower 90 % confidence limit is 0.97. It is interesting to point out that this distribution looses its symmetry as the sample size is decreased. The left hand tail becomes quite long and tapered. It should also be noted that although this distribution is essentially binomial, the sample size of the distribution becomes obscured in the formation of the distribution.

6. The second phase of this paper will be devoted to procedures for the evaluation of a system where the algebraic probability expression is not available. While dependency of components is a contributor in making the algebraic probability expression difficult, a system algebraic probability expression can become "not feasible" even when all components maintain their independence. Examples of situations that add to the complexity of a system probability expression are:

a. Mechanical and electrical couplings

b. One part of a system functions only if a second part of the system fails

c. Multi-option channels.

7. Consider Figure 6. Clearly this is not a simple series-parallel diagram. There are 16 paths of success through this diagram, $A_1B_1C_1D_1E_1F_1$ etc. To write this as a system of 16 in parallel would require a considerable amount of algebra to account for repeating components in more than one of the 16 success paths. One approach to a solution to this problem is to redraw the diagram in a simple series-parallel configuration. See Figure 8. The trouble with this diagram is that there are twenty-four components represented where actually there are only nineteen physical components. Thus, the twenty-four components are not independent and in order to represent the system in a simple series-parallel configuration, it is necessary to draw the same component more than once. This procedure perhaps lessens the algebra required to write an algebraic probability expression, however, it would be preferable to consider a technique that doesn't require indepencence or an algebraic equation.

8. The basic notation and concept of Boolean algebra should be mentioned at this time. A plus sign is used to mean "or", and a dot is used to mean "and". The following expressions are thus introduced:

$$1 + 1 = 1 \qquad 1 \cdot 1 = 1$$

$$1 + 0 = 1 \qquad 1 \cdot 0 = 0$$

$$0 + 1 = 1 \qquad 0 \cdot 1 = 0$$

$$0 + 0 = 0 \qquad 0 \cdot 0 = 0$$

9. For purposes of illustration the example used earlier will be used again. Consider Figure 7. The procedure is as follows: Generate a random number (RN) from a uniform distribution:

If RN $\leq$ reliability of component A, assign A = 1

If RN $\geq$ reliability of component A, assign A = 0.

Components B and C are treated similarly. Now, the probability expression in Boolean notation is $P = A \cdot B \cdot C$ which reads P = A and B and C. In order for P to be a "1" all three components must be "1".

The probability of assigning a "1" to a component is the success ratio of the component. Hence, the probability of assigning all three components a "1" is the system's probability of success.

10. From Figure 8 the following expressions for different groups of components (X) can be obtained.

$$X_1 = A_1 \cdot B_1 + A_2 \cdot B_2$$

$$X_2 = C_1 \cdot (D_1 \cdot E_1 + D_3 \cdot E_2)$$

$$X_3 = C_2 \cdot (D_2 \cdot E_1 + D_4 \cdot E_2)$$

$$X_4 = X_1 \cdot (X_2 + X_3) \cdot F_1$$

$$X_5 = A_1 \cdot B_3 + A_2 \cdot B_4$$

$$X_6 = (C_3 + C_4) \cdot (F_2 + E_3 \cdot F_1)$$

$$\text{PROB} = X_4 + (X_5 \cdot X_6)$$

11. To solve this final expression a random number is generated and compared with component one, A_1. A_1 is assigned a "0" or "1". This process continues until all nineteen (not twenty-four) components have been assigned a "0" or "L". IMPORTANT: when A_1, A_2, E_1, E_2, and F_1 are assigned a value they are assigned the same value everywhere they appear. Upon solving, X_1 will be a "1" or "0". The probability of success of the system can be represented by

$$P = \frac{1}{n} \sum_{i=1}^{n} \text{PROB}_i$$ where n is the number of times the system is simulated.

References

1. Orkand, Donald S., "A Monte Carlo Method for Determining Lower Confidence Limits for System Reliability on the Basis of Sample Component Data", Picatinny Arsenal, Dover, New Jersey, June 1960.

2. Garner and Vail, "Confidence Limits for System Reliability", Systems Design, September-October 1961.

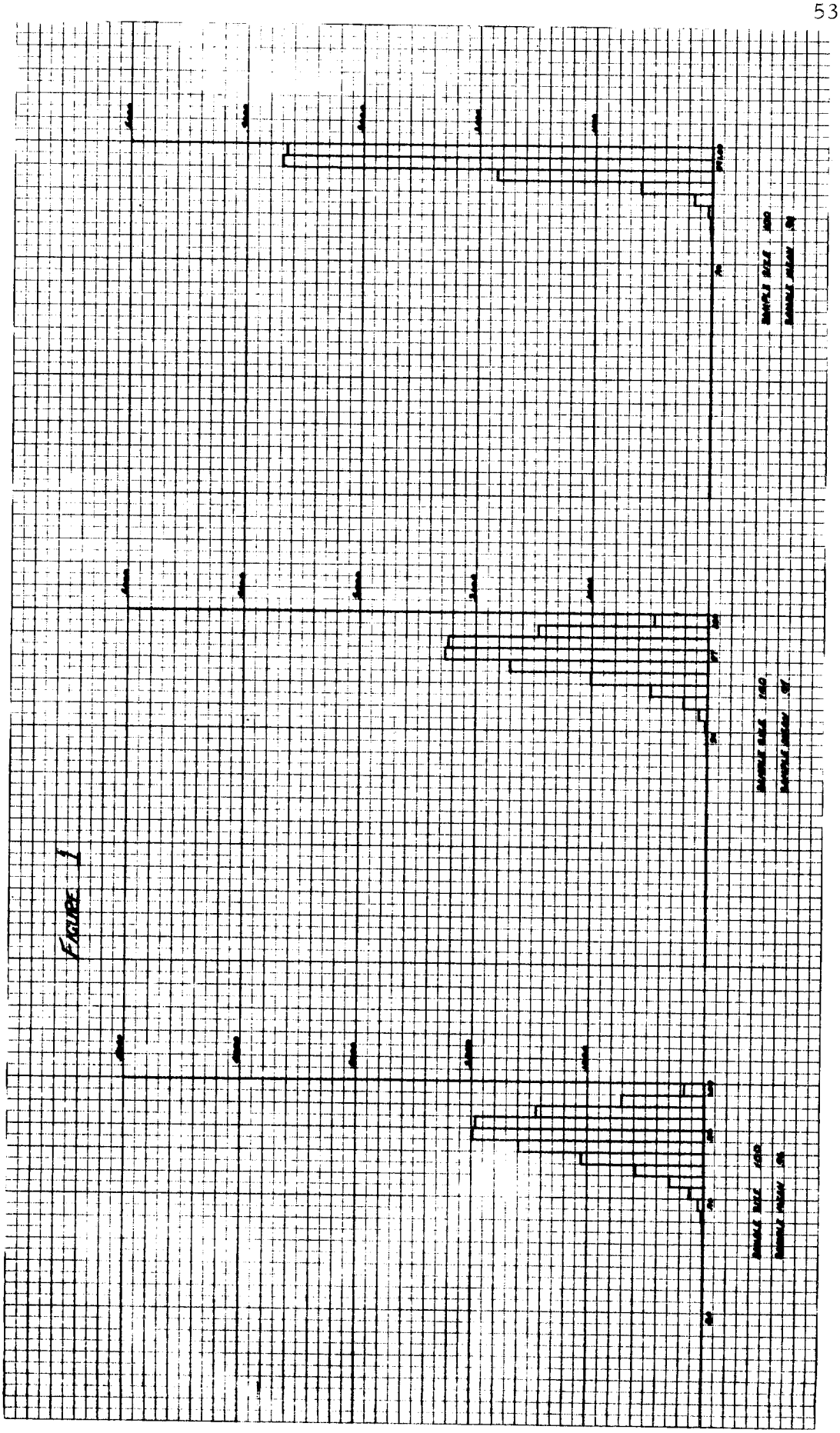
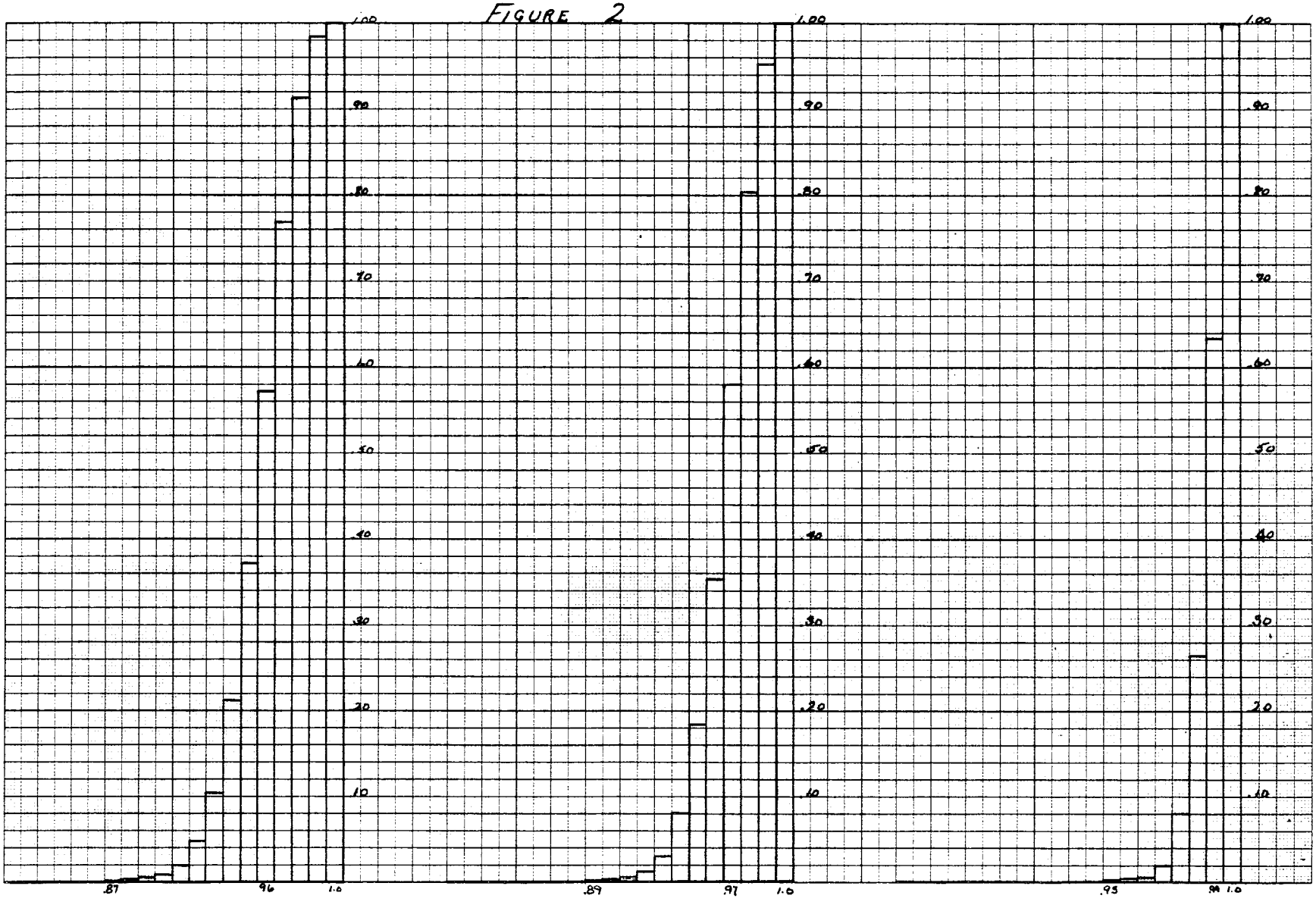


FIGURE 1

This page intentionally left blank.

FIGURE 2



This page intentionally left blank.

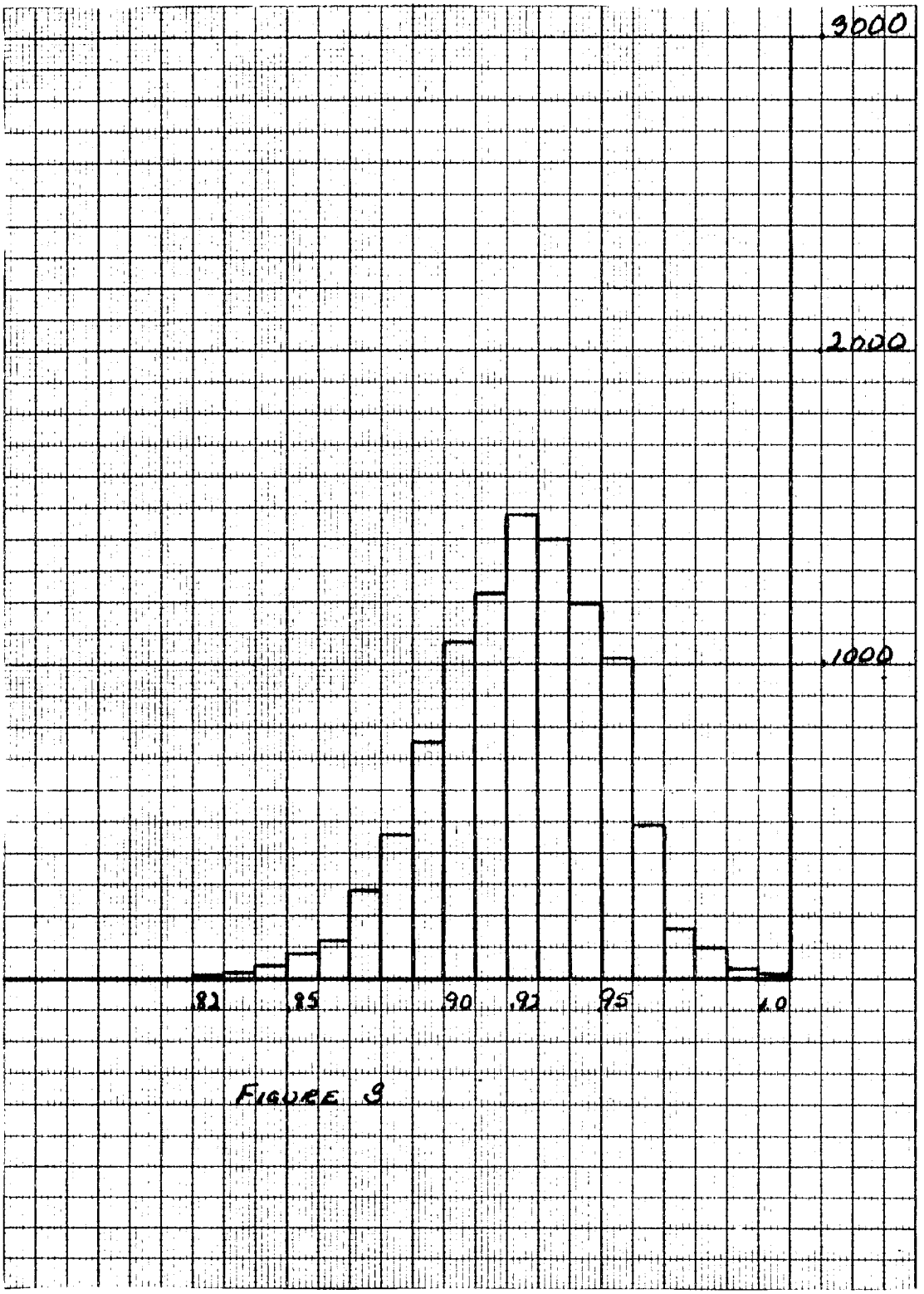


FIGURE 9

This page intentionally left blank.

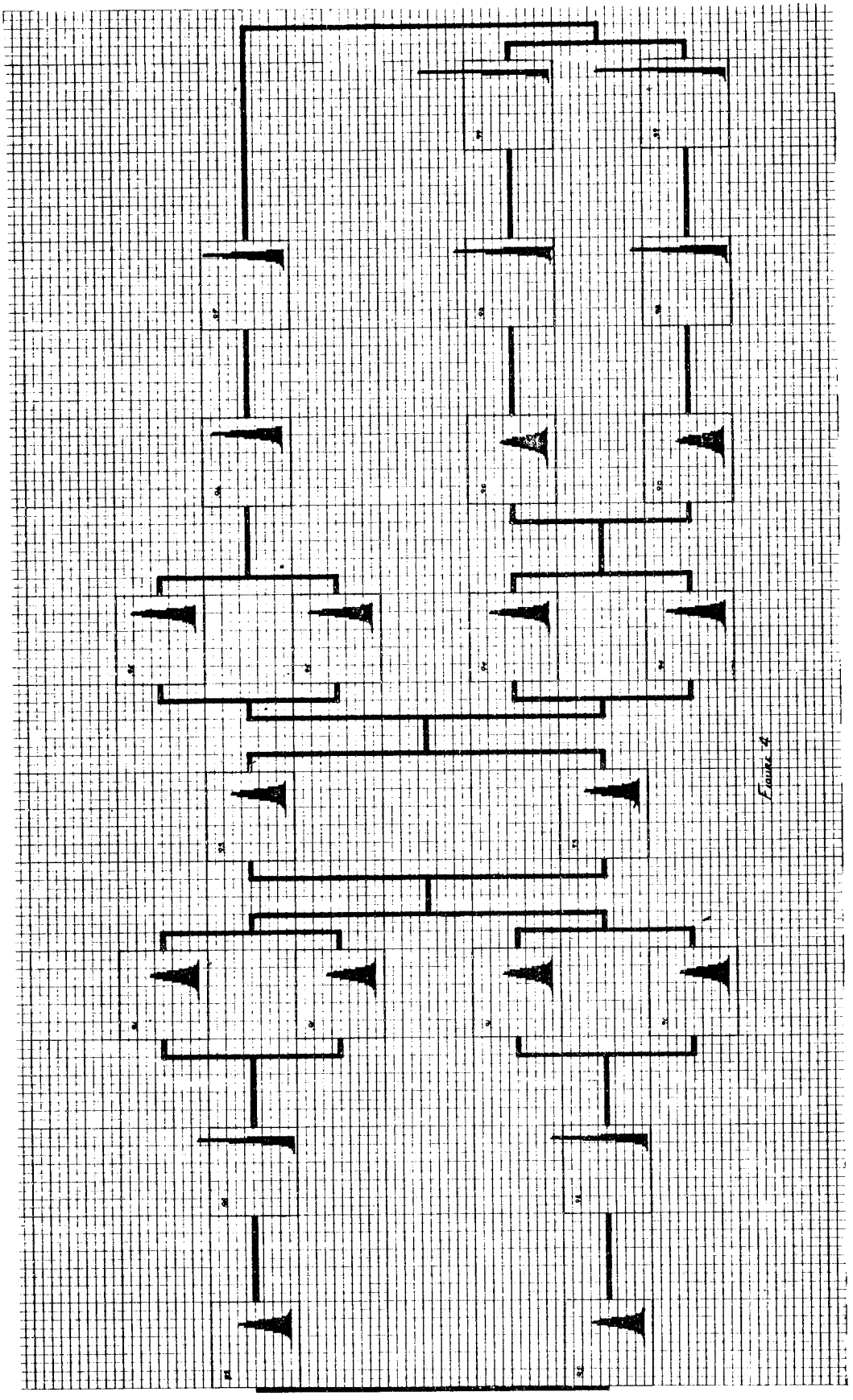
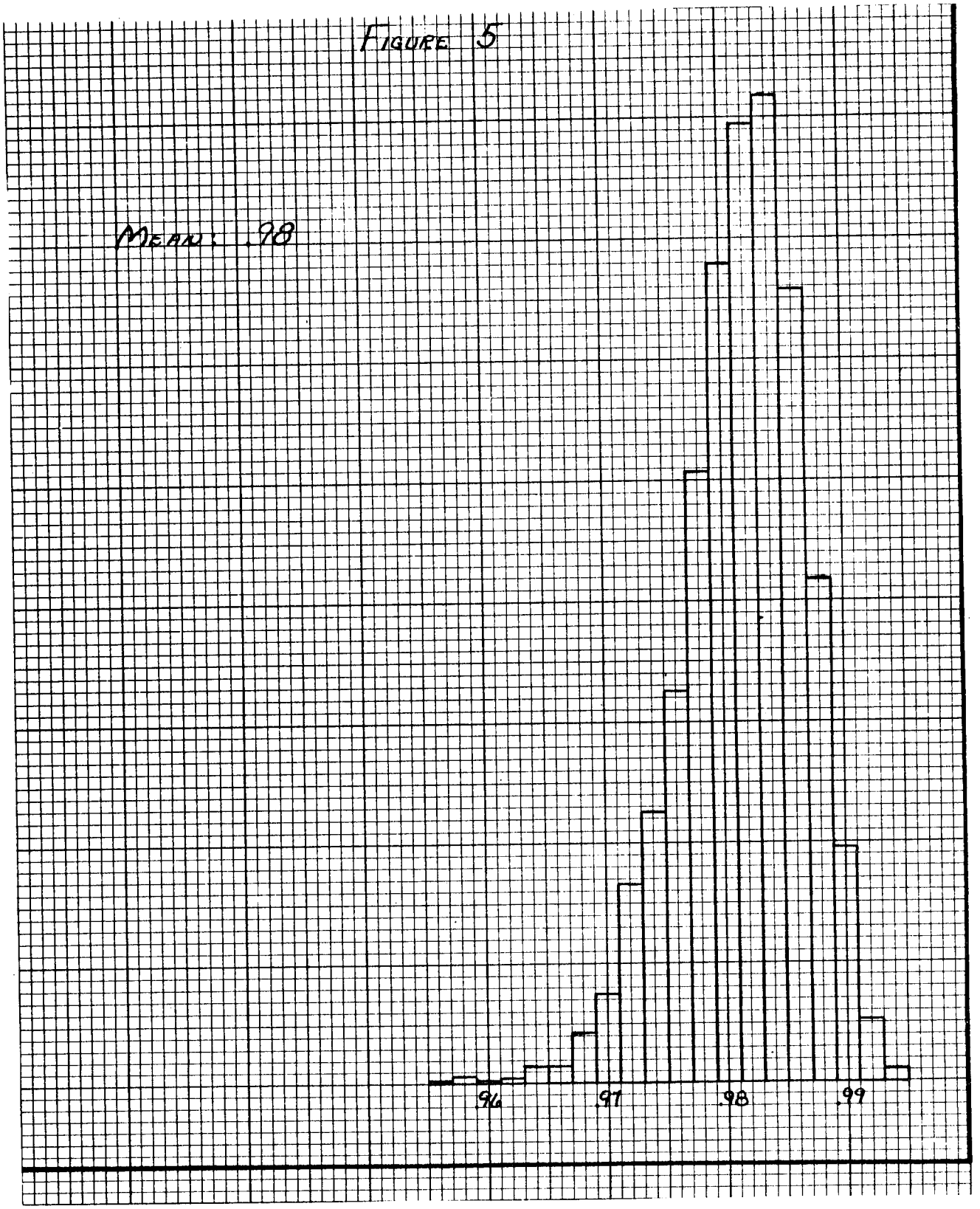


Figure 4

This page intentionally left blank.

FIGURE 5



This page intentionally left blank.

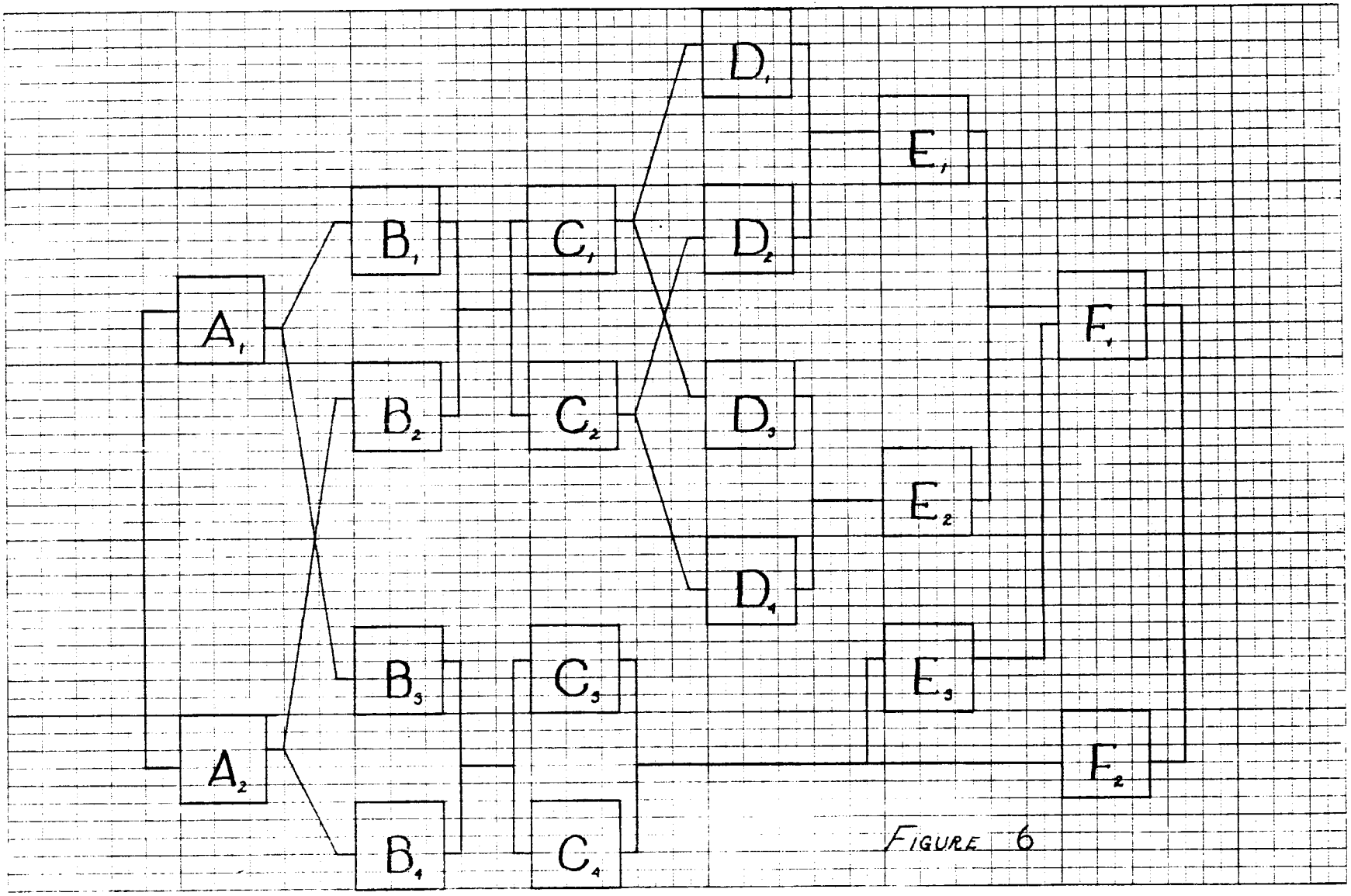


FIGURE 6

This page intentionally left blank.

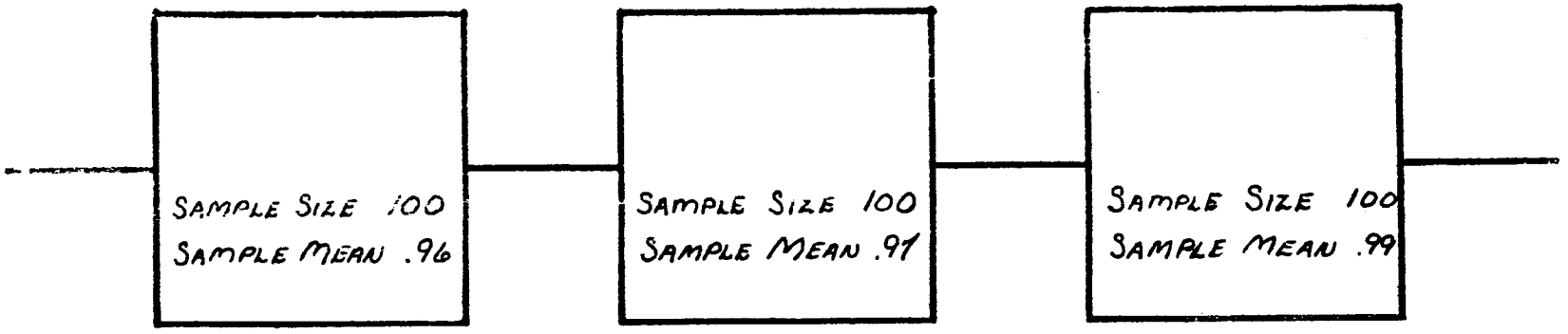
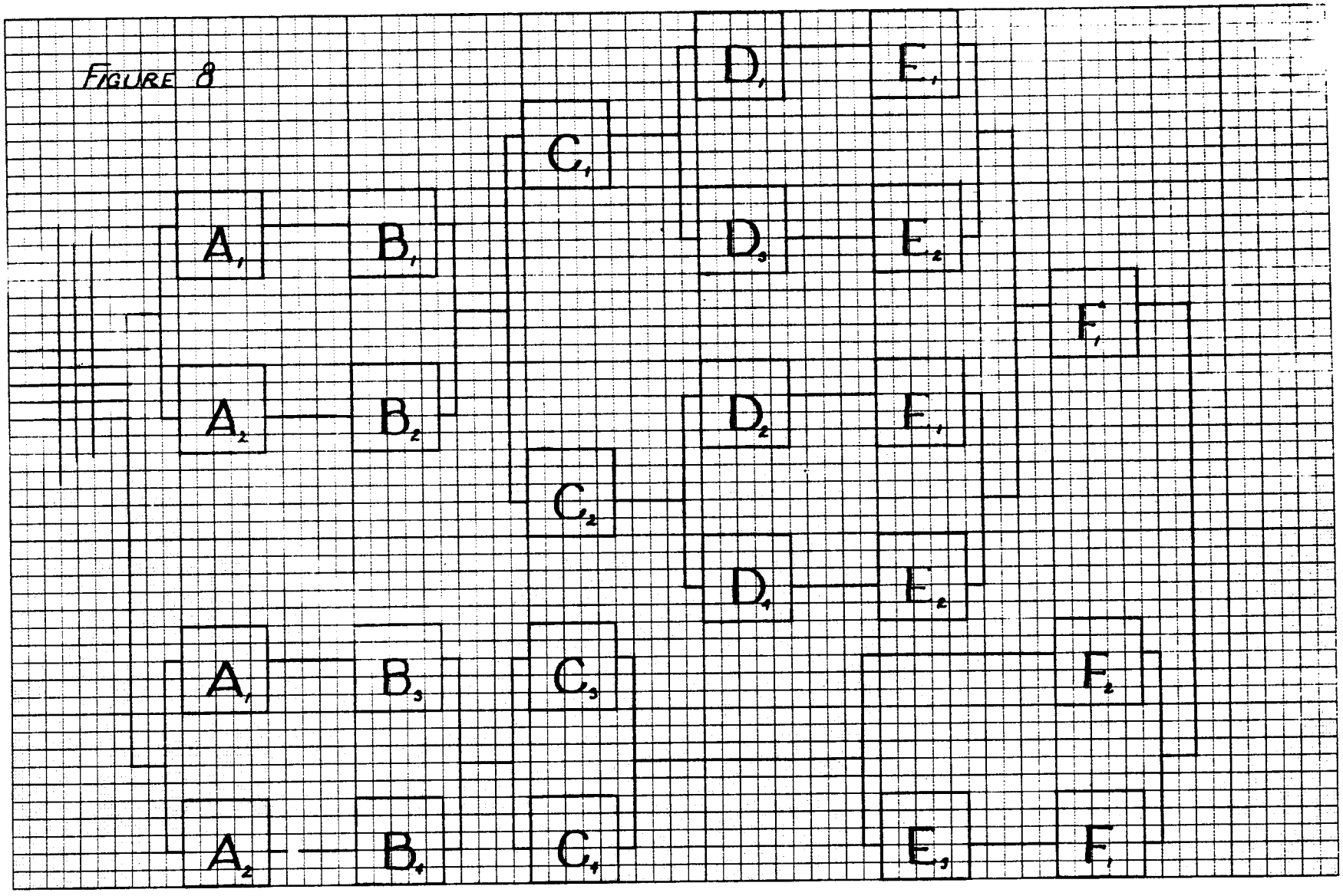

SAMPLE SIZE 100
SAMPLE MEAN .96
SAMPLE SIZE 100
SAMPLE MEAN .91
SAMPLE SIZE 100
SAMPLE MEAN .99

FIGURE 7

This page intentionally left blank.

FIGURE 8



This page intentionally left blank.

BIOASSAY: THE QUANTAL RESPONSE ASSAY

H. C. Batson
University of Illinois College of Medicine
Chicago

I. INTRODUCTION. In many instances of interest in medical and biological research, the properties, activity or potency of certain substances cannot be measured directly by common in vitro chemical or physical methods, but can be measured (quantitated) only in terms of some effect they evoke in a living test subject, - animal, plant or microorganism.

Substances in this category include many hormones, vitamins, pharmacologically and toxicologically active substances, antibiotics, and immunologically active substances, - vaccines, toxins, toxoids, antisera, allergens, etc. Measurement or quantitative assessment of the activity of such substances constitutes the subject matter of biological assay.

Design of bioassay experiments and statistical analysis of the resultant data involve mainly an extension of principles and procedures readily available in standard references on experimental design and statistical analysis with major emphasis on regression analysis and analysis of variance with or without transformation of the data originally recorded in conventional units.

II. TYPES OF BIOASSAYS

1. On a basis of intent: On the basis of intent, bioassays can be classified in one of two main groups, - absolute or comparative.

Absolute assays: Absolute assays involve an attempt to obtain some quantitative measurement that can be expressed in absolute terms, such as a Minimal Lethal Dose (MLD) or Median Effective Dose (ED_{50}, LD_{50}, etc.). Such attempts are based on the assumption or belief that some such absolute value exists and that universally it can be determined with adequate precision. However, the absolute potency of substance X for "the cat" typically depends on just which cat is used and, unfortunately, cats invariably do differ. Laudable though the goals and objectives may be, absolute assays of biologically active substances, with

few (if any) exceptions, have little useful quantitative meaning.

Comparative assays: Although absolute assays seldom if ever yield adequately reproducable results, it generally is possible to achieve experimental quantitation of many biologically active substances through assessment of the substance of interest (unknown) in direct comparison with a reference substance (standard) qualitatively identical or, at least, similar in terms of the response evoked in the test subject of choice. While the absolute potency of either may never be known, the comparative or relative activity of the two may be assessed and the biological activity of the unknown expressed in relation to that of the standard in terms of relative potency, - whether expressed in proportions, percentages or in arbitrarily defined units. By using a common reference or standard substance, various investigators may obtain quantitative results with a degree of comparability adequate for their needs. Such relative potency estimates are subject to uncertainty (experimental error), of course, but ideally this may be kept within manageable proportions. It is this innate element of uncertainty that makes bioassay a candidate for statistical consideration.

2. On the basis of response: On the basis of the response evoked in the test subjects of choice, most bioassays may be categorized into one of the following types:

Direct assays: In these the response in the individual test subject is absolute (live, die; response, non-response; etc.) and critical (thresh-hold) levels of the assayed material are determinate, at least within reasonable limits. Computations mainly involve calculation of means and ratios, and estimation of standard errors or confidence limits of such statistics. Example: the cat assay of digitalis.

Graded response-parallel line assays: In these, the response in the individual is proportional to the dose of test substance administered and the degree of response is experimentally determinable. Typically, the degree of response is a linear function of log-dose and the dosage-response regression lines of "Unknown" and "Standard" will be parallel denoting identity or similarity of action. Statistical analysis involves mainly regres-

sion analysis and analysis of variance. With proper design (balanced or partially balanced factorial assays), analysis can be simplified greatly through the use of coefficients. Example: assay of insulin in the rabbit.

Slope-ratio assays: These include mainly the microbiological assays, a group of rather limited general interest in which the degree of measurable response in the individual probably is absolute, but since masses of test subjects (microorganisms) are dealt with, the total response measured, as density, acid formation, etc., approaches a continuous function. Statistical analysis involves multiple regression and relative potency is estimated from the ratio of the partial regression coefficients. Example: microbiological assay of riboflavin.

Quantal response assays: In these, response in the individual test subject is absolute (frequently, live or die) but the critical dose of test material necessary to evoke the response is not directly determinable. Quantitation is achieved through the use of groups of test subjects and determination of the proportion responding to various dosage levels of "Unknown" and "Standard" test products. Following suitable transformation of the data (probits, angles, etc.,) response typically is a linear function of log dose and statistical analysis is essentially similar to that employed with the graded response-parallel line bioassays. Examples: mouse-protective potency assays of typhoid, pertussis and rabies vaccines.

III. REQUIREMENTS OF A VALID BIOASSAY. The following requirements of a "valid" bioassay have evolved from recommendations originally made by Gaddum (1) with modifications made by Bliss, Finney, and others, and are practically universally accepted by students of bioassay. Perhaps the word "valid" should be replaced by "good" or "acceptable."

1. The assay should involve a direct comparison of an unknown with a standard in identical, concomitant tests.

   a. Ideally, the two products should be of essentially equal potency.

2. There should be a significant progressive relationship between dosage and response.

a. Linear following transformation as required.

b. Highly significant slope.

c. No significant curvature; combined or opposed.

3. Dosage-response regression lines for the two products should be parallel, denoting identity or similarity of action.

4. There should be internal evidence of homogeneity (of the data) establishing validity of statistical analysis and adequacy of the testing situation.

5. Analysis should include an estimate of assay error (uncertainty) calculated directly from the data.

Obviously, not all requirements can be applied to each type of assay. Requirements pertaining to slope do not apply to direct assays; those pertaining to parallelism do not apply to slope-ratio assays, etc. However, all do apply to parallel-line graded response assays and most quantal response assays are of similar design.

IV. REDUCTION OF UNCERTAINTY (ERROR) OF BIOASSAYS.

All experienced bioassayists are aware of the innate uncertainty and poor reproducibility of such assays as a whole. The degree of variability differs markedly with various assays, perhaps being least with slope-ratio assays and greatest with quantal response assays. This variability can be reduced to some extent in a variety of ways including:

1. Perfection of technique: equipment, reagents, etc.

2. Control of environment: constant temperature, humidity, etc.

3. Increased homogeneity of test subjects: selection of strains, sex and size of test animals; use of litter mates, etc.

4. Use of restricted designs: randomized blocks (complete or incomplete), Latin squares, cross-over designs, confounding, etc.

5. Statistical adjustment of data: covariance analysis, adjusting response data on the basis of a pertinent associated measurement.

6. Increasing the number of observations (test subjects), either by using more subjects per assay or, preferably, by independent replication of the assay as a whole.

In many quantal response assays, particularly assays of vaccines, antisera, etc., most of the above conventional approaches accomplish only modest reduction in assay error. Slopes of the dosage-response regression lines characteristically are low, constituting a major source of assay error, and the main direct compensating approach is to increase the number of test subjects.* A major reduction in assay error, however, would require impractically large numbers of subjects. Practical solution to many of these problems probably lies in the development of assay procedures involving new experimental approaches. If some meaningful response or attribute of the individual test subject can be measured as a continuous variable, a graded response-parallel line assay procedure should be possible. Typically, errors of these assays are much less than of quantal response assays. In some situations, "time to death" has shown promise as a meaningful quantitative response metameter.

V. REQUIREMENTS OF AN ADEQUATE STATISTICAL ANALYSIS.

1. The analysis should provide for the acceptance or rejection of the assay results as a whole; - such acceptability based

---

* In simplified probit analysis, a crude approximation of the standard error of M (log-ratio of potency) is given by

$$s_M = \frac{1}{b_c}\sqrt{\frac{2}{N_U} + \frac{2}{N_S}}$$

where b_c = combined (average) slope, and N_U and N_S = the number of test subjects assigned to the unknown and standard, respectively

upon the requirements outlined in part III.

2. The analysis should provide for a reliable, unbiased estimate of relative potency that is independent of dosage throughout the maximum possible range.

3. The analysis should provide for an estimate of assay uncertainty, - preferably expressed as confidence limits of the relative potency, - provided meaningful alternatives for action based upon such resultant estimates can be established.

Of the above requirements, the first is considered by this writer to be the most essential and the one most commonly unrecognized or neglected in routine analysis of bioassay data. Specific computational procedures and illustrative examples for all the main types of bioassays are given in standard reference books such as Burn (2), Bliss (3), and Finney (4, 5).

VI. STATISTICAL ANALYSIS OF QUANTAL RESPONSE BIOASSAY DATA. A surprising number and variety of computational procedures for analysis of quantal response bioassay data have been proposed. In terms of statistical rigor and sophistication, they range from simple "quick-and-dirty" graphic approximations to formal iterative procedures involving a degree of complexity and tedious computational detail which is difficult to justify except, possibly, in the most critically extenuating circumstance.

Most, or perhaps all, of these methods have some advantages or disadvantages dependent upon their contemplated use but any critical comparison is far beyond the scope of this presentation. It is consoling to find, however, that they all lead to closely similar estimates of relative potency (or end-points) when applied to truly good data as defined in Part III. Unfortunately, the simpler approximate methods generally do not provide a basis for discrimination between acceptable and non-acceptable data and when applied unwittingly to truly unreliable data may yield estimates which are seriously misleading.

The more commonly used computational procedures can be classified into four general categories. These general categories, examples of methods included in each, and minimal comments regarding each, are given below:

Class	Examples	Comments
Graphic approximations	Miller-Tainter (6).	Minimal calculations; adequate reliability provided good data; some discriminatory power by inspection.
Calculated approximations	Reed-Muench Behrens (7).	Most widely used and probably least reliable of all methods; limited to estimating 50 % endpoint.
Formal procedures	Probit analysis; Bliss (8), Finney (5). Knudsen-Curtis (9).	Laborious calculations; maximum reliability and discriminatory power.
Compromise methods	Litchfield-Wilcoxon (10).	Generally adequate reliability and discriminatory power; appreciably less calculations than formal methods.

Another method, involving a factorial χ^2 approximation, is proposed by this writer. This should be considered a compromise method and is presented in some detail in part VII of this presentation.

The factorial χ^2 approximation is based essentially on analysis of variance of quantal response data espressed in terms of per cent response and log dose. When used with data from balanced factorial bioassays involving a constant number of test subjects per experimental unit, adequate tests for acceptability of the data, the relative potency estimate and an approximation to confidence limits of the relative potency estimate can be obtained with only moderately extensive calculations. Analysis of the data from numerous factorial quantal response bioassays by this method has yielded results in close agreement with those obtained by formal probit analysis (5) and the Knudsen-Curtis method (9).

VII. FACTORIAL χ^2 ANALYSIS OF QUANTAL RESPONSE BIOASSAY DATA. In a previous report (11) the essential computational details of factorial analysis of attribute (enumeration) data, as developed by Brandt, were presented together with illustrations of applications of the method to selected experiments in industrial chemistry. Two forms of the basic formula were presented. The first "(Formula 1)" being the form for

calculating values of χ^2 for individual degrees of freedom from complete factorial experiments in which the experimental units are of equal size, was given as

$$\chi^2_{[1]} = \frac{N^2}{SxF} \times \frac{T^2}{D}$$

where N = total individuals or observations; S = total successes; F = total failures; T = the total of the sums of products of factorial coefficients and the number of successes in the corresponding experimental units; D = the product of the sums of the squares of the factorial coefficients and the number of individuals per experimental unit; and, the subscript in brackets indicates the degrees of freedom. Either of the outcomes (yes or no, response or non-response, survival or death, etc.) can be designated as success; the other outcome as failure.

In many instances, quantal response bioassay data can be subjected to factorial χ^2 analysis, the major restrictions being that the experimental units are of equal size and that successive doses of the independent variable (i.e., the toxic or protective substance being assayed) differ by a constant interval when expressed in appropriate units of measurement. In most (perhaps all) assays of immunologically active substances, the successive doses (levels of X) should be increased or decreased in a geometric series such as 1, 2, 4, 8, 16; 1, 3, 9, 27; etc., as the differences between the logarithms of successive doses are constant in value. When these restrictions are complied with, factorial coefficients (3) can be used directly in analysis of the data and χ^2 values can be computed by the formula given above. In this manner it is possible to obtain statistical information regarding the validity or adequacy of the data (Part III) and, as shown below, to obtain a direct estimate of relative potency and its approximate confidence limits.

The procedures are illustrated with actual examples of both 2-dose (4-point) and 3-dose (6-point) assays of the mouse protective potency of typhoid vaccine performed by the author at the Army Medical Service Graduate School.* Details of the assay procedure employed have been published previously (12); attention here will be limited primarily to statistical treatment of the data.

---

* Now known as Walter Reed Army Institute of Research.

1. Factorial χ^2 analysis of a 2-dose quantal response bioassay

As a part of a study to determine the reproducability of mouse protection potency assays of typhoid vaccine (13), a series of 6 assays were run on identical aliquots of a reference vaccine. The aliquotes were identified only as A and B and prior to the assay it was decided to calculate their relative potency, B as per cent of A. Data from the sixth trial are reproduced in Table I.

Table I
Two-dose assay of the mouse protective potency of typhoid vaccines
(Survivals/totals)

Vaccine	Vaccine dose (ml)	
	0.015	0.15
A	5/20	13/20
B	2/20	15/20

For factorial χ^2 analysis, these data are rearranged to the form given in Table IA. For purposes of obtaining tests of significance (χ^2) it is of no consequence in which order the vaccines are entered in the table or which comparison groups are assigned + and - coefficients. However, in the estimation of relative potency, slope, etc., computations are more convenient if certain orders are followed. For the comparison between products (designated as comparison a), positive coefficients should be assigned to the "unknown" (vaccine B in this case). Likewise, for the estimation of slope (comparison a), positive coefficients should be assigned to the higher dose level. Assignment of coefficients to the interaction comparison (ab) is uniquely determined as the cross products of coefficients for the first 2 comparisons, of course. This assignment of coefficients is consistent with that employed by Bliss (3) and others.

Table IA
Factorial χ^2 analysis of the data on Table I

Vaccine	B (unknown)		A (standard)								
Dose	Low	High	Low	High							
Success (survivors) / 20	2	15	5	13	Σ+	Σ-	T	T^2	D	T^2/D	χ^{2*}
Comparisons											
a Unknown vs standard	+	+	-	-	17	18	-1	1	80	0.0125	0.05
b Slope (high vs low dose)	-	+	-	+	28	7	21	441	80	5.5125	22.38
ab Departure from parallelism (products x doses)	-	+	+	-	20	15	5	25	80	0.3125	1.27

$^{*}\chi^2 = \frac{N^2}{SxF} \times \frac{T^2}{D} = \frac{80^2}{35 \times 45} \times \frac{T^2}{D} = 4.06 \times \frac{T^2}{D}$.

Evidence of assay validity: All calculations are performed in the manner previously described (11). From comparison a, it is found that the 2 vaccines do not differ appreciably in total effect ($\chi^2[1] = 0.05$). From comparison b it can be seen that there is a highly significant relationship between dosage and response ($\chi^2[1] = 22.38$), and by comparison ab it is determined that there is no significant departure from parallelism exhibited by the dosage response lines for the unknown and standard. No information is available concerning curvature of the dosage response curves. Such can be obtained only when 3 or more dosage levels are employed.

As the assay actually was conducted, the 20 mice in each experimental unit were not handled as a single group but as 4 independent groups of 5 each. These groups were selected, assigned spaces in the test room, immunized and challenged in random order and the number of survivors originally were recorded per group of 5. Thus it is possible to calculate a "within groups" χ^2 with 12 degrees of freedom which can be used as a measure of internal homogeneity (requirement 4). The procedure will be illustrated with data from the next example (Table II).

Estimation of relative potency: It is possible to obtain an estimate of relative potency (RP) from the data and calculations of Table IA by use of the formula for estimating relative potency from

a 2-dose factorial assay as given by Bliss (3).

$$M = \frac{i \times T_a}{T_b}$$

where M = the log ratio of potency; i = the log-dose increment *; and, T_a and T_b are the values in the column headed T for comparisons a and b, respectively. In this assay, the dosage increment was 10-fold, so i = log 10 = 1. T_a = -1 and T_b = 21. Substituting these values in the formula, M is calculated as

$$M = \frac{1 \times -1}{21} = -0.0476.$$

This value is a logarithm and must be converted to the usual form $\bar{1}.9524$. The antilogarithm of $\bar{1}.9524$ is the relative potency which is found to be 0.896; or, in terms of percentage, vaccine B is 89.6 per cent as potent as vaccine A. This estimate is in reasonably close agreement with that obtained by probit analysis, 85.4 per cent.

Approximate confidence limits of relative potency: It also is possible to obtain an approximation of the confidence limits of the relative potency estimate from the data and calculations presented in Table IA. This is most easily done by first determining the approximate confidence interval for M(CI'_M) which for a 2-dose assay is calculated as

$$CI'_M = \frac{1.96 \times 2n\sqrt{N} \times i}{T_b} \quad **$$

where n = individuals per experimental unit; N = 4n or grand total individuals, and i and T_b have the same meaning as before. The

*Logarithms of dosage increments from 2-fold to 10-fold are tabulated in Table 1, Appendix 1 and designated as constants $c_{M.2}$.

**The term "confidence interval" typically is used to denote the entire range included between lower and upper confidence limits. The quantity approximated by CI'_M, as used here, is one-half the entire range expressed in logarithmic units. Derivation of this approximation is given in Appendix II to this paper.

95 per cent confidence limits of M then are determined as

$$M \pm CI'_M$$

and the 95 per cent confidence limits of the relative potency ($95\%\ CL_{RP}$) are found as the antilogarithms of these 2 values.

$$95\%\ CL_{RP} = \text{antilogarithms of } M\text{-}CI'_M \text{ and } M\text{+}CI'_M.$$

These limits will be in the form of ratios which can be converted to percentage by multiplying by 100. For the illustrative problem dealt with here (Tables I and IA)

$$CI'_M = \frac{1.96 \times 40/\sqrt{80} \times \log 10}{21} = 0.4174$$

Then

$$95\%\ CL_M = -0.0476 + 0.4174 = -0.4650 \text{ and } 0.3698$$

$$= 1.5350 \text{ and } 0.3698$$

Taking antilogarithms

$$95\%\ CL_{RP} = 0.34 \text{ and } 2.34$$

or

$$34\% \text{ and } 234\%.$$

Thus, the best estimate of relative potency (B as per cent of A) is 89.6 per cent and the odds are approximately 19 out of 20 that the true potency is between 34 and 234 per cent.

For a factorial assay of set design, where i and n are constant, assay to assay, the foregoing calculations can be simplified as all elements in the formula for CI'_M will be the same

except for T_b. Thus constants for 2-dose assays ($C_{I.2}$) and 3-dose assays ($C_{I.3}$) for fold-increments of dosage from 2 to 10, and for values of n from 10 to 20, have been calculated and are presented in Appendix I, Tables 2 and 3.

It must be emphasized that this estimate of the confidence limits of the relative potency is only an approximation. Yet the results obtained were in reasonably close agreement with those obtained by probit analysis, 31.6 and 230.4 per cent.

2. Factorial χ^2 analysis of a 3-dose quantal response bioassay. Factorial χ^2 analysis of a 3-dose quantal response assay for determining the validity of the assay and the estimation of relative potency and approximate 95% confidence limits of the potency estimate, are illustrated with data from another typhoid vaccine mouse protection potency test performed at the Army Medical Service Graduate School. The vaccines tested were a routine production lot (unknown) and a reference standard. Results of the assay are summarized in Table II, and are arranged in the form suitable for factorial χ^2 analysis in Table IIA.

Table II

Three-dose assay of the mouse protective potency of an unknown typhoid vaccine in respect to a standard

(Survivors/totals)

Vaccine	Vaccine dose (ml)		
	0.02	0.08	0.32
Unknown	1/10	5/10	8/10
	1/10	7/10	9/10
Standard	2/10	4/10	8/10
	1/10	5/10	7/10

Table IIA
Factorial χ^2 analysis of the data of Table II

Vaccine	Unknown			Standard									
Dose	D_1	D_2	D_3	D_1	D_2	D_3							
Success (survivors)	1	5	8	2	4	8							
10	1	7	9	1	5	7							
Successes/ 20	2	12	17	3	9	15	E+	Σ-	T	T^2	D	T^2/D	χ^{2*}
Comparisons													
a Unknown vs standard	+	+	+	-	-	-	31	27	4	16	120	0.13	0.52
b Slope	-	0	+	-	0	+	32	5	27	729	80	9.11	36.44
ab Parallelism	-	0	+	+	0	-	20	17	3	9	80	0.11	0.44
c Combined curvature	+	-2	+	+	-2	+	37	42	5	25	240	0.10	0.40
ac Opposed curvature	+	-2	+	-	+2	-	37	42	5	25	240	0.10	0.40

$$^{*}\chi^2 = \frac{N^2}{S \times F} \times \frac{T^2}{D} = \frac{120^2}{58 \times 62} \times \frac{T^2}{D} = 4.00 \times \frac{T^2}{D}$$

Between groups within experimental units:

$$\chi^2[6] = 4.00 \times \frac{(1-1)^2 + (7-5)^2 + (9-8)^2 + (2-1)^2 + (5-4)^2 + (8-7)^2}{20} = 4.00 \times \frac{8}{20}$$

$$= 1.60.$$

There is little need for comment regarding the computational procedure employed. Factorial coefficients were assigned in conventional order (3) and χ^2 values for each comparison were computed in the manner previously described. Calculation of χ^2 "between groups within experimental units" was accomplished by summation of all T^2/D values between pairs of groups of 10 each and multiplying the total by the constant $\frac{N^2}{S \times F}$.

Evidence of validity

There was no evidence of significant differences between the pairs of groups within experimental units ($\chi^2[6] = 1.60$). This yields assurance that the randomization procedures employed during the assay were adequate to prevent appreciable bias due to technical and environmental factors. Since 3 dosage levels of vaccine were employed, it was possible to gain information regarding curva-

ture of the dosage response lines, both combined and in opposition. There was no evidence of systematic departure from linearity. Thus, all requirements for assay validity (Part III) were satisfied.

Estimation of relative potency

The relative potency of the unknown in respect to the standard was estimated by the formula given by Bliss (3) for calculating M in 3-dose factorial assays

$$M = \frac{4 \times i \times T_a}{3 \times T_b} \quad *$$

The dosage increment employed in this assay was 4-fold, so i= log 4 = O. 6O21. Substituting calculated values of T_a and T_b into the formula, M was calculated as

$$M = \frac{4 \times O.6021 \times 4}{3 \times 27} = O.1189$$

and the relative potency = 100 x antilog O. 1189 = 131. 5 per cent.

Approximate confidence limits of relative potency

The formula for estimating the approximate confidence interval of M in a 3-dose assay differs from that for 2-dose assay only in that 4n must be substituted for 2n. Thus, for a 3-dose factorial assay.

$$CI'_M = \frac{1,96 \times 4n/\sqrt{N} \times i}{T_b}$$

* Values of $\frac{4 \times i}{3}$ dosage increments of 2-fold through 10-fold have been calculated and are given as constants $c_{M.3}$ in Table 1, Appendix I. M is determined by multiplying the ratio T_a/T_b by the appropriate value of $c_{M.3}$ (O. 8020 in this example).

For the data dealt with here (Tables II and IIA), n = 20, N = 120, and i = log 4 = 0.6021. Then

$$CI'_M = \frac{1.96 \times 80/\sqrt{120 \times 0.6021}}{27}^{*}$$

$$= 0.3192.$$

The confidence limits of M are found as

$$95\% \, CL_M = M \pm CI'_M$$

$$= 0.1189 \pm 0.3192 = -0.2003 \text{ and } 0.4381$$

$$= \bar{1}.7997 \text{ and } 0.4381$$

Then the 95 per cent confidence limits of the relative potency are obtained as the antilogarithms of these values.

$$95\% \, CL_{RP} = 0.63 \text{ and } 2.74$$

or $$= 63 \text{ and } 274 \text{ per cent}$$

These data also were analyzed by the probit analysis. The relative potency estimate was 132.2 per cent and the 95 per cent confidence limits were 64.2 per cent and 272.2 per cent.

3. Resumé of computational procedure. Chi square analysis of quantal response factorial assays yielding (1) statistical evidence regarding reliability of the data, (2) an estimate of relative potency, and (3) approximate confidence limits of the relative potency, involves a series of 7 main steps.

1. Arrange the data on a work sheet of the form used in Tables IA and IIA.

2. Assign the factorial coefficients in accordance with the actual design of the experiment. Compute N^2/SxF from the grand

*Constants $c_{I.3}$ for estimating values of CI'_M in 3-dose factorial assays for dosage increments of 2-fold through 10-fold and for values of n from 10 through 20, have been calculated and are given in Table 3 of the appendix. For this problem, $c_{I.3} = 8.6183$. This divided by 27 (T_b) = 0.3192, the same as calculated above.

totals and then $\Sigma +$, $\Sigma -$, T, T^2, D, T^2/D and χ^2 for each comparison (row). Also, if data on subgroups within experimental units are available, calculate the "between groups" χ^2 (cf. Table IIA). From the various values of χ^2 determine if there is sufficient evidence of validity to justify estimation of potency.

3. If justified, compute the ratio T_a/T_b and calculate M as:

   a. Two dose assay: $M = i \times T_a/T_b$. Values of i are given as the constants $c_{M.2}$ in Table 1, Appendix I.

   b. Three-dose assay: $M = \frac{4 \times i}{3} \times T_a/T_b$. Values of $\frac{4 \times i}{3}$ are given as the constants $c_{M.3}$ in Table 1, Appendix I.

4. Determine the relative potency (RP) as a ratio or percentage as antilog M, or as 100 x antilog M, respectively.

5. Compute CI'_M as:

   a. Two-dose assay:

   $$CI'_M = \frac{1.96 \times 2n\sqrt{N} \times i}{T_b}$$

   or, using constants $c_{I.2}$ from Table 2, Appendix I:

   $$CI'_M = \frac{c_{I.2}}{T_b}.$$

   b. Three-dose assay:

   $$CI'_M = \frac{1.96 \times 4n/\sqrt{N} \times i}{T_b}$$

   or, using constants $c_{I.3}$ from Table 3, Appendix I:

   $$CI'_M = \frac{c_{I.3}}{T_b}.$$

6. Calculate the 95 per cent conficence limits of M as

   $$95\%\ CL_M = M \pm CI'_M.$$

7. Determine the 95 per cent confidence limits of the relative potency as

$$95\%\ CL_{RP} = \text{antilog } M\text{-}CI'_M \text{ and antilog } M + CI'_M.$$

If it is desired to express the limits as percentages, multiply each value by 100.

APPENDIX I

Table 1

Values of $c_{M.2}$ and $c_{M.3}$ for obtaining estimates of M, the log ratio of potency, from 2-dose and 3-dose factorial assays

$$(M = c_{M.i} \times T_a/T_b)\ *$$

Fold-increment in dosage	$c_{M.2}$ (2-dose assays)	$c_{M.3}$ (3-dose assays)
2	O. 3O1O	O. 4O13
3	O. 4771	O. 6361
4	O. 6O21	O. 8O28
5	O. 699O	O. 932O
6	O. 7782	1. O376
7	O. 8451	1. 1268
8	O. 9O31	1. 2O41
9	O. 9542	1. 2722
1O	1. OOOO	1. 3333

*Relative potency = antilog M.

Relative potency in % = 1OO x antilog M.

This page intentionally left blank.

APPENDIX I
Table 2
Values of $c_{I.2}$ *

Constants for calculating CI'_M, the confidence interval of M, for 2-dose factorial assays. Select the values of $c_{I.2}$ determined by the dosage increment (row) and group size (column).

$$\left(CI'_M = \frac{c_{I.2}}{T_b}\right)$$

Fold-increment in dosage	n (individuals per group)										
	10	11	12	13	14	15	16	17	18	19	20
2	1.8656	1.9567	2.0437	2.1271	2.2074	2.2849	2.3598	2.4325	2.5030	2.5716	2.6384
3	2.9571	3.1014	3.2393	3.3716	3.4989	3.6217	3.7405	3.8556	3.9674	4.0761	4.1820
4	3.7318	3.9104	4.0880	4.2550	4.4156	4.5705	4.8658	4.7205	5.0068	5.1440	5.2776
5	4.3324	4.5439	4.7459	4.9398	5.1263	5.3061	5.4802	5.6488	5.8126	5.9718	6.1270
6	4.8233	5.0588	5.2837	5.4995	5.7071	5.9073	6.1011	6.2889	6.4712	6.6485	6.8212
7	5.2379	5.4937	5.7379	5.9722	6.1977	6.4152	6.6256	6.8295	7.0275	7.2200	7.4076
8	5.5974	5.8707	6.1317	6.3821	6.6231	6.8554	7.0803	7.2982	7.5098	7.7155	7.9160
9	5.9141	6.2029	6.4786	6.7432	6.9978	7.2433	7.4809	7.7112	7.9347	8.1521	8.3639
10	6.1980	6.5006	6.7896	7.0669	7.3337	7.5910	7.8400	8.0813	8.3156	8.5434	8.7654

*Calculated as $\frac{1.96 \times 2n \times i}{N}$

This page intentionally left blank.

APPENDIX I

Table 3

Values of $c_{I.3}$ *

Constants for calculating CI'_M, the confidence interval of M, for 3-dose factorial assays.

Select the values of $c_{I.3}$ determined by the dosage increment (row) and group size (column).

$$(CI'_M = \frac{c_{I.3}}{T_b})$$

Fold-increment in dosage	n (individuals per group)										
	10	11	12	13	14	15	16	17	18	19	20
2	3.0465	3.1952	3.3373	3.4736	3.6047	3.7313	3.8536	3.9722	4.0874	4.2013	4.3085
3	4.8289	5.0646	5.2898	5.5058	5.7136	5.9142	6.1081	6.2961	6.4787	6.6593	6.8291
4	6.0941	6.3915	6.6757	6.9483	7.2106	7.4638	7.7084	7.9457	8.1761	8.4041	8.6183
5	7.0749	7.4202	7.7501	8.0665	8.3710	8.6649	8.9490	9.2245	9.4919	9.7566	10.0053
6	7.8765	8.2609	8.6282	8.9805	9.3195	9.6467	9.9630	10.2697	10.5674	10.8620	11.1390
7	8.5536	8.9711	9.3670	9.7525	10.1207	10.4760	10.8195	11.1525	11.4759	11.7958	12.0966
8	9.1406	9.5868	10.0130	10.4219	10.8153	11.1950	11.5620	11.9179	12.2635	12.6054	12.9268
9	9.6578	10.1292	10.5796	11.0116	11.4272	11.8285	12.2162	12.5923	12.9574	13.3186	13.6582
10	10.1214	10.6154	11.0874	11.5401	11.9757	12.3962	12.8026	13.1967	13.5793	13.9579	14.3138

* Calculated as $\frac{1.96 \times 4n \times i}{N}$

This page intentionally left blank.

APPENDIX II

Approximation of the Confidence Interval of M

The standard error of M (s_M) from a balanced factorial bioassay is given by Bliss (3) as:

$$s_M = \frac{s}{b_c}\sqrt{\frac{4}{N}\left[1 + \frac{D^2}{B^2 - s^2 t^2}\right]}$$

where D^2 = mean square between products; B^2 = mean square for combined slope; s^2 = mean square; t = Student's statistic; N = total number of test subjects (possible responses); and b_c is the combined or average slope of the dose-response regression line.

For a 2-dose (4-point) assay, bc is estimated as $\frac{Tb}{2 \times i \times n}$; for a 3-dose (6-point) assay, as $\frac{Tb}{4 \times i \times n}$. In these, Tb is found as shown in Tables IA and IIA, i is the log ratio of dosage increment, and n is the number of test subjects per experimental group.

In a good bioassay (statistically acceptable), D^2 will be small and B^2 will be large. Thus, the quantity enclosed in brackets approaches unity and can be ignored. In the binomial, the variance (s^2) has a maximum value of O. 25 and s has a maximum value of O. 5. In a balanced assay of fixed design, N will be 4n or 6n for a 2-dose and 3-dose assay, respectively. Substituting the appropriate formula for b_c as given above, and introducing $t_\infty = 1.96$, the confidence intervals of M can be reduced to the following approximations:

$$\text{2-dose assay: } CI_M = \frac{(1.96 \times 2n \times i)/\sqrt{N}}{T_b}$$

$$\text{3-dose assay: } CI_M = \frac{(1.96 \times 4n \times i)/\sqrt{N}}{T_b}$$

These approximations were used for calculating the constants presented in Tables 2 and 3 of Appendix I.

REFERENCES

1. Emmens, C. W., Principles of Biological Assay, London, Chapman and Hall, Ltd., 1948.

2. Burn, J. H., Finney, D. J., and Goodwin, L. G., Biological Standardization. Oxford Univ. Press, New York, ed. 2, 1950.

3. Bliss, C. I., The Statistical Method in Biological Assay. Hafner Publ. Co., New York, 1952.

4. Finney, D. J., Statistical Method in Biological Assay. Hafner Publ. Co., New York, 1952.

5. Finney, D. J., Probit Analysis. Cambridge Univ. Press, London, 1947.

6. Miller, L. C., and Tainter, M. L., Estimation of the E. D. 50 and its error by means of logarithmic - probit graph paper. Proc. Soc. Exp. Biol. and Med., 57: 261-264, 1944.

7. Reed, L. J., and Muench, H., A simple method of estimating fifty per cent endpoints. Am. J. Hygiene, 27: 493-497, 1938.

8. Bliss, C. I., The determination of dosage-mortality curves from small numbers. Quart. J. Pharm., 11: 192-216, 1938.

9. Knudsen, L. F., and Curtis, J. M., The use of the angular transformation in biological assay. J. Am. Stat. Assoc., 42: 282-296, 1947.

10. Litchfield, J. T., and Wilcoxon, F., A simplified method of evaluating dose-effect experiments. J. of Pharm. and Expt. Therap., 96: 99-113, 1949.

11. Batson, H. C., Applications of factorial χ^2 analysis to experiments in chemistry. Transactions, 10th Annual Meeting, ASQC, p. 9-23, 1956.

12. Batson, H. C., Brown, M., and Oberstein, M., The mouse-protective potency assay of typhoid vaccine as performed at the Army Medical Service Graduate School. Pub. Health Rep. 66: 789-806, 1951.

13. Batson, H. C., Brown, M., and Oberstein, M., An adaptation of quality control chart methods to bacterial vaccine potency testing. J. Bact. 61: 407-419, 1951.

This page intentionally left blank.

ATTENDEES TO THE EIGHTH CONFERENCE ON THE DESIGN OF EXPERIMENTS IN ARMY RESEARCH, DEVELOPMENT AND TESTING

24-26 Oct. 1962

Abelson, Robert P.	Yale University, New Haven, Conn.
Addelman, Sidney	Research Triangle Institute, Durham, N.C.
Altman, Peter S.	U. S. Naval Weapons Lab, Dahlgren, Va.
Atkinson, John C.	Army Chem. Center, Md.
Atwood, Earl	WRAIR, Washington, D. C.
Baird, H. Robert	National Institutes of Health, (NIH) Bethesda, Md.
Bales, Robert A.	Tech. Oper., Inc. Ft. Belvoir, Va.
Banton, Clarence W.	Army Tank Automotive Center, Detroit Arsenal
Barker, Dana W.	Combat Dev. Comd., Ft. Belvoir, Va.
Barker, Lewellys	NIH, Bethesda, Md.
Batson, Herbert C.	University of Illinois, Urbana, Ill.
Beckmann, Edward M.	Rock Island Arsenal, Rock Island, Ill.
Benson, Jacob	Ft. Monmouth, New Jersey
Benson, Karl L.	Office of the Surgeon General, Wash., D.C.
Bombara, Elwood L.	Redstone Arsenal, Huntsville, Ala.
Boyd, Donald M.	Res. Analysis Corp., Bethesda, Md.
Boyle, Douglas G.	Hercules Powder Co., Magna, Utah
Brown, Ralph E.	Frankford Arsenal, Philadelphia, Pa.
Brown, William A.	Dugway Proving Ground, Dugway, Utah
Bruno, O. P.	Ballistic Res. Labs, Aberdeen Proving Ground, Maryland
Bulfinch, Alonzo	Picatinny Arsenal, Dover, N. J.
Burdick, Donald S.	Duke University, Durham, N. C.
Cabaniss, Jelks H., Jr.	Department of Math, U. S. Military Academy, West Point, N. Y.
Carter, Frederick L., Jr.	USA Chem. Labs., Ft. Detrick, Md.
Chernoff, Herman	Stanford University, Stanford Calif.
Coberly, Robert H.	Rock Island Arsenal, Rock Island, Ill.
Cohen, A. Clifford	University of Georgia, Athens, Ga.
Collins, Elmer F.	Naval Receiving Station Annex
Cook, Kenneth G.	Applied Psychology Corporation, Arlington, Va.

Cornfield, Jerome	NIH, Bethesda, Md.
Cox, Claire B.	NIH, Bethesda, Md.
Cox, Edwin L.	Agr. Res. Service, Beltsville, Md.
Cox, Paul C.	White Sands Missile Range, N. M.
Craw, Alexander R.	USA Bio. Labs., Ft. Detrick, Md.
Creecy, Donald B.	National Security Agency, Ft. Meade, Md.
Culbertson, George E.	Hercules Powder Co., Radford Arsenal, Va.
Daggit, Edward A.	U. S. Military Academy, West Point, N. Y.
DeArmon, Ira A.	Edgewood Arsenal, Md.
Dobrindt, Gerard T.	Aberdeen Proving Ground, Md.
Dowling, Jasper B.	Picatinny Arsenal, Dover, N. J.
Dressel, Francis D.	Army Research Office, Durham, N. C.
Dorn, H. F.	NIH, Bethesda, Md.
Eckard, Cecil O.	Deseret Test Center, Ft. Douglas, Utah
Einbinder, Seymour K.	Picatinny Arsenal, Dover, N. J.
Eisenhart, Churchill	National Bureau of Standards, Washington, D. C.
Eisler, Stanley L.	Rock Island Arsenal, Ill.
Eissner, Robert M.	Aberdeen Proving Grounds, Md.
Ellner, Henry	Army Chemical Center, Md.
Ewart, Wade H.	Redstone Arsenal, Huntsville, Ala.
Federer, Walter T.	Math Res. Center, University of Wisconsin, Madison, Wisconsin
Fiess, Edward R.	U.S. Army Electronic R&D Lab., Ft. Monmouth, N. J.
Finkel, Peter W.	Res. Analysis Corp., Bethesda, Md.
Foster, Walter D.	Chemical, Biological, & Radiological Laboratories, Ft. Detrick, Md.
Frishman, Fred	Army Research Office, Washington 25, D. C.
Frick, James W.	Cornell Aero. Lab., Arlington, Va.
Fulton, Charles L.	Frankford Arsenal, Philadelphia, Pa.
Galbraith, Alan S.	Army Research Office, Durham, N. C.
Gardner, Roberta A.	NIH, Bethesda, Md.
Gauld, Ross L.	Walter Reed Army Institute of Research, Washington, D. C.

Gebhard, R.	Human Resources Research Office Alexandria, Va.
Geisser, Seymour	NIH, Bethesda, Md
Grant, Roy L.	US Bureau of Mines, Pittsburgh, Pa.
Greenberg, Bernard G.	University of North Carolina Chapel Hill, N. C.
Greenhouse, Samuel W.	NIH, Bethesda, Md.
Greenstone, Reynold	Operations Res., Inc., Silver Springs, Md.
Greenwood, Joseph A.	Armed Forces Radiobiol. Research Institute, Bethesda, Md.
Grubbs, Frank E.	Aberdeen Proving Ground, Md.
Haines, Bertram W.	Ft. Detrick, Md.
Hamilton, Eugene L.	Office of the Surgeon General, Washington, D. C.
Hamman, J. Wayne	Armed Forces Food & Cont. Inst., Chicago, Ill.
Harshbarger, Boyd	Virginia Polytechnic Institute, Blacksburg, Va.
Hartley, Herman O.	Iowa State University, Ames, Ia.
Heermann, Emil F.	U. S. Army Personnel Research Office, Washington, D. C.
Heifetz, Samuel	Aberdeen Proving Ground, Md.
Helmbold, Robert L.	Burlington, Mass.
Hill, Floyd I.	Cornell Aero. Lab., Arlington, Va.
Hixson, Eugene E.	National Aeronautics & Space Administration, Greenbelt, Md.
Hoppes, Harrison N.	Res. Analysis Corp., Bethesda, Md.
Horner, Theodore W.	Booz-Allen Appl. Research, Bethesda, Md.
Hundley, Douglas V.	Army Chemical Center, Md.
Hutton, Matthew I.	Edgewood Arsenal, Md.
Inselmann, Edmund H.	Frankford Arsenal, Philadelphia, Pa.
Isaac, Gerhard J.	Med. Res. & Nutr. Lab., Denver, Colo.
Jebe, Emil H.	University of Michigan, Ann Arbor, Mich.
Jessup, Gordon L., Jr.	USA Bio. Labs., Ft. Detrick, Md.
Johnson, Cecil D.	USA Personnel Research Office, Washington, D. C.

Johnson, Dennis E.	Arthur D. Little, Inc., Cambridge, Mass.
Johnson, Wilbur V.	Res. Analysis Corp., Bethesda, Md.
Karp, Abraham E.	Department of Army, Bethesda, Md.
King, Donald W.	Westat Research Analysts
Kirby, William H. Jr.	Aberdeen Proving Ground, Md.
Kornack, Wallace R.	Atomic Energy Commission, Washington, D. C.
Krane, Scott A.	C-E-I-R, Inc., Dugway, Utah
Krimmer, Manfred W.	Picatinny Arsenal, Dover, N. J.
Ku, Hsien H.	National Bureau of Standards, Washington, D. C.
Kupperman, Morton	National Space Agency, Ft. Meade, Md.
Kurkjian, Badrig M.	Harry Diamond Laboratories, Washington, D. C.
Langlie, H. Joseph	Ford Aeronutronic, Newport Beach, Calif.
Lavin, George I.	Aberdeen Proving Ground, Md.
LeMaster, Vaughn	Ammo. Proc. & Supply, Joliet, Ill.
Levin, Sheldon G.	Harry Diamond Laboratories, Washington, D. C.
Littrell, Robert	Res. Analysis Corp., Bethesda, Md.
Love, Benjamin T.	RCA Service Co., Tuscon, Ariz.
Lucas, Henry L.	North Carolina State College, Raleigh, N. C.
Maisel, Herbert	U. S. Army Strategy & Tactics Analysis Group, Washington, D. C.
Mall, Adolph W.	Frankford Arsenal, Philadelphia, Pa.
Mallios, William S.	Booz-Allen Appl. Research, Inc., Bethesda, Md.
Maloney, Clifford J.	NIH, Bethesda, Md.
Mandelson, Joseph	Army Chemical Center, Md.
Marshall, Jack A.	University of Chicago, Chicago, Ill.
Marthens, Arthur S.	BuShips, Washington, D. C.
Martin, Cyrus	Corps of Engineers, Ft. Belvoir, Va.
Masaitis, Ceslovas	Aberdeen Proving Ground, Md.
Mauss, Bess Day	Washington, D. C.
Metcalfe, Samuel N.	Ft. Detrick, Md.
Mitchell, Thomas D.	Naval Avionics Facility, Indianapolis, Ind.
Mann, G. V.	Vanderbilt University, Nashville, Tenn.

Morrison, Donald F.	NIH, Bethesda, Md.
Mosimann, James E.	NIH, Bethesda, Md.
McIntosh, William B.	U. S. Army Electronic Proving Ground, Ft. Huachuca, Ariz.
Natrella, Mary G., Mrs.	National Bureau of Standards, Washington, D. C.
O'Brien, Richard J.	Rock Island Aresenal, Ill.
Okano, Fred	National Aeronautics & Space Administration, Washington, D. C.
Olkin, Ingram	Stanford University, Stanford, Calif.
O'Malley, Joseph P.	NIH, Bethesda, Md.
Orleans, Beatrice S.	BuShips, Washington, D. C.
Owen, Donald B.	Sandia Corp., Albuquerque, N. M.
Pabst, W. R.	Bureau of Naval Weapons, Washington, D. C.
Panos, Robert V.	C-E-I-R, Inc., Dugway, Utah
Parks, Albert B.	Harry Diamond Laboratories, Washington, D. C.
Parnell, Benjamin L.	Walter Reed Army Medical Center Washington, D. C.
Parrish, Gene B.	Army Research Office, Durham, N. C.
Pearson, Egon S.	London University, London, England
Perry, Virginia W.	QM Research & Engineering, Field Evaluation Agency, Ft. Lee, Va.
Pierce, William L.	Res. Analysis Corp., Bethesda, Md.
Pradko, Fred	Army Tank-Auto Comd., Center Line, Mich.
Prairie, Richard R.	Sandia Corp., Albuquerque, N. M.
Purtell, John P.	Watervliet Arsenal, Watervliet, N. Y.
Raisbeck, Leo R.	Rock Island Arsenal, Ill.
Rhian, Morris A.	Edgewood Arsenal, Md.
Rider, Paul R.	Aero Research Lab., Wright-Patterson AFB
Riggs, Charles W.	USA Bio. Labs., Ft. Detrick, Md.
Romano, Albert	National Bureau of Standards Washington, D. C.
Schneiderman, Marvin A.	National Cancer Institute, Bethesda, Md.
Smith, N. M.	Res. Analysis Corp., Bethesda, Md.

Schnell, Raymond B.	Chemical Res. & Dev. Lab., Edgewood, Md.
Schoolman, Harold M.	Veterans Administration Hospital, Hines, Ill.
Schroeder, Bruno W.	U. S. Army Electronic Proving Ground, Ft. Huachuca, Ariz.
Schwarz, Richard A.	U. S. Atomic Energy Commission, Germantown, Md.
Shaw, Thomas R.	Operations Research, Inc., Silver Spring, Md.
Sheehy, Robert W.	Dept. Biologics Research Walter Reed Army Institute of Research, Washington, D. C.
Sloane, Harry S.	Dugway Proving Ground, Dugway, Utah
Sobelman, Sidney	Picatinny Arsenal, Dover, N. J.
Speckman, Janace A.	Bureau of Standards, Washington, D. C.
Stallmann, Friedemann W.	Veterans Administration, Washington, D. C.
Stearns, Edward R.	Army Chemical Center, Md.
Steck, George P.	Sandia Corporation, N. M.
Steedman, Joseph E.	Aberdeen Proving Ground, Md.
Stephanides, Agatha, Mrs.	U.S. Army Strategy & Tactics Analysis Group, Bethesda, Md.
Taft, Everett	University of Michigan, Ann Arbor, Mich.
Tarlow, Irving	QM Res. & Eng. Cmd., Natick, Mass.
Thompson, W. A.	University of Delaware, Newark, Md.
*Wadley, Francis M.	Chem. Corps Biol. Lab, Ft. Detrick, Md.
Wallach, Harold C.	Census Bureau, Suitland, Md.
Walner, Arthur H.	N.Y. Naval Shipyard, Brooklyn, N.Y.
Weinstein, Joseph	Elect. Res. & Dev. Laboratories, Ft. Monmouth, N. J.
Weintraub, Gertrude	Picatinny Arsenal, Dover, N. J.
White, John E.	Bridgewater College, Bridgewater, Va.
Wilburn, Nicholas T.	Ft. Monmouth, N. J.
Wilks, S. S.	Princeton University, N. J.
Wolman, William	National Aeronautics & Space Administration, Washington, D. C.
Wentworth, Paul	WRAIR, Washington, D. C.
Youden, William J.	National Bureau of Standards, Washington, D. C.
Vivona, Stefano	WRAIR, Washington, D. C.

Zelen, Marvin — Math Research Center, University of Wisconsin, Madison, Wis.

Zundel, Ivan B. — C-E-I-R, Inc., Dugway, Utah

This page intentionally left blank.