TEQUILAP: TEN QUANTITATIVE ILLUSIONS OF ADMINISTRATIVE PRACTICE*

Clifford J. Maloney

In 1949 I had occasion to install a punched card tabulator for the purpose of machine calculation of analyses of variance arising in a research, development, and testing program. At that time no need was felt to impose extraordinary restrictions on the procurement and utilization of such equipment. About the time of the beginning of the Korean War, however, higher authority, under the impression apparently that punched cards were employed only in Comptroller functions and that all such machinery was rented, instituted a requirement for monthly reports of per cent utilization, with a somewhat informal understanding that good management would secure a level of utilization of each piece of equipment at least as high as 50% and that 80% would be much more appropriate. Having encountered what has since come to be called queuing theory in Thornton Fry's text on probability theory many years earlier, I had a summer worker in 1955 make an application of these considerations to the congestion delays which would result from any given level of per cent utilization. These are shown in Figure 1. This study has appeared in a paper given at the Second Statistical Engineering Symposium at Edgewood Arsenal in April of 1956. However, my efforts and those of others--a few of which have come to my attention--to point out the costs as well as the benefits as per cent utilization increases had an absolutely zero effect on "administrative practice." Two major conclusions from this experience and many others, before and since, were, however, made clear to me. The first conclusion is that decision making is an emotional, not a rational, operation. People often bolster their decisions by arguments--some of them rational--but seldom reverse the process. I am sorry to say that so far as I can see this holds as much for logicians and mathematicians as for anyone else. This is of course what is meant by that well known saying: "I've already made up my mind; don't bother me with the facts." The second conclusion deals with the arguments by which it is customary to rationalize emotional decisions. Even where the "reasoning" can be accepted as not totally irrelevant, it will be based, not invariably but very often, on unwarranted but unquestioned assumptions. My example

*The views expressed herein are those of the author and are not to be ascribed to any other agency or individual.

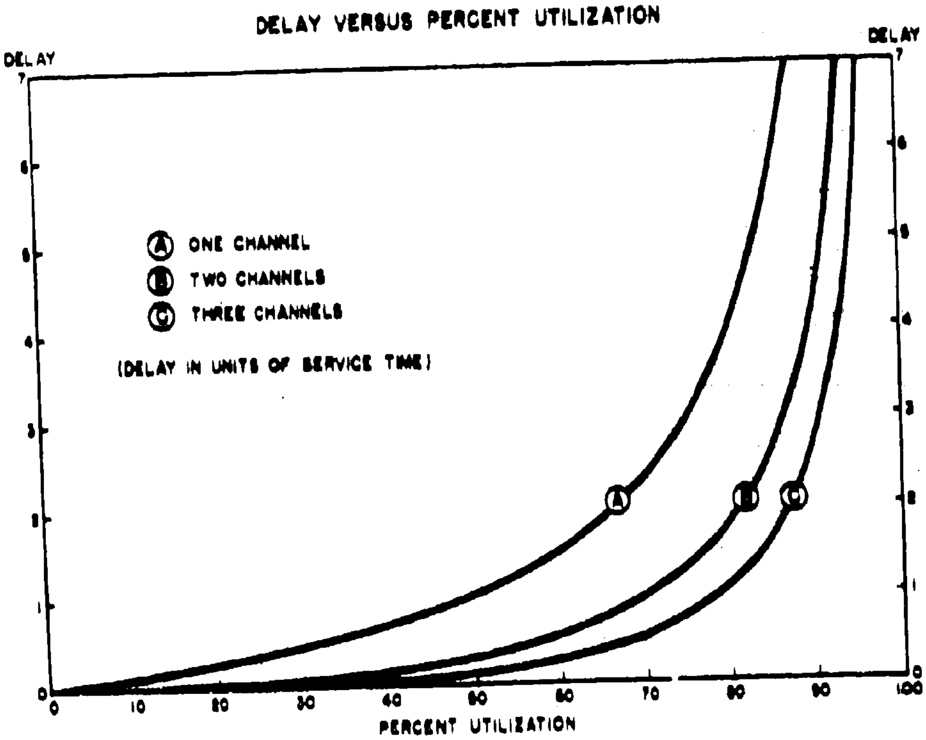

DELAY VERSUS PERCENT UTILIZATION
DELAY
DELAY
A ONE CHANNEL
B TWO CHANNELS
C THREE CHANNELS
(DELAY IN UNITS OF SERVICE TIME)
PERCENT UTILIZATION

Figure 1

of per cent utilization as a measure of punched card installation efficiency illustrates this. One assumption was that the equipment was rented and not owned. Another, that efficiency was not merely a function of utilization but actually a monotone function of it.

I cannot claim originality in this insight. In a commencement address at Yale University, June 11, 1962, the President of the United States said: "For the great enemy of the truth is very often not the lie--deliberate contrived, and dishonest, but the myth--presistent, persuasive, and unrealistic. Too often we hold fast to the clichés of our forebears. We subject all facts to a prefabricated set of interpretations. We enjoy the comfort of opinion without the discomfort of thought . . . . Mythology distracts us everywhere--in government as in business, in politics as in economics, in foreign affairs as in domestic policy." The former President's indictment is much stronger and more inclusive than mine.

Allyn Kimball* has defined "errors of the third kind" as giving correct answers to the wrong questions. The assumptions of the question become postulates of the answer. He observes that a first step in finding useful answers is to query the question. The originator of the cognate insight that most of what in common life passes for argument consists of more or less accurate deduction from wrong premises is lost in the mists of time. But perhaps there is room for me to assert some small claim to originality in the recognition that many of these false premises spring from an inability or an unwillingness to think in quantitative terms, to see the clarifying role of an appreciation of their quantitative nature, and to see that some at least of these false deductions cannot be resolved by logic alone; as their very nature is inherently quantitative.

The application of mathematical principles in administrative practice which I illustrated in my first example, specifically an application of probability, is relatively sophisticated. Further, the problem which gave rise to it, while important, was rather limited in scope. Most of what administrators do day in and day out consists of more homely if more important actions, though it is true that queuing theory has many applications there

---

* Allyn Kimball, "Errors of the Third Kind in Statistical Consulting," JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, June 1957, p. 133.

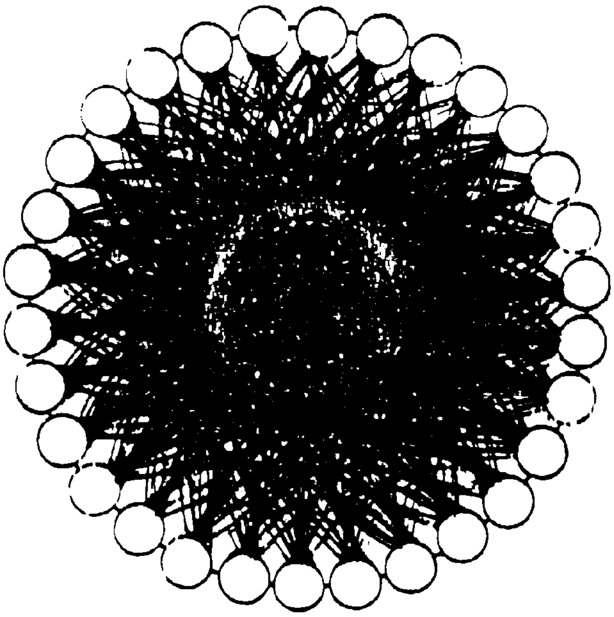
also. Dr. Edward F. R. Hearle* in an article "How Useful Are 'Scientific' Tools of Management?" enumerates them as: "linear and dynamic programming, queuing theory, game theory, simulation, and monte carlo, to name a few." The general tenor of his appreciation of these tools and, hence, of quantitative thinking in management is believed expressed in his sentence: "Furthermore these tools do not deal with some of the more exciting parts of the total management process . . . . ." I take exactly the opposite view to the one that I believe Dr. Hearle is espousing; that quantitative thinking is not (as he asserts) limited to formal manipulation of numerical quantities but is very useful where the mere recognition of a variation from instance to instance of a given type is involved.**

The serviceability of adhering rigidly to the "channels" of an orginization chart can be judged by indicating the lines of contact that exist in the absence of "organization." (Figure 2). The organization chart of Figure 3 was selected, not to suggest that administration goes around in circles, but because the chart is round and, hence, emphasizes the contrast with Figure 2. The organization replaces an unorganized conglomerate. Now, when the production department gets ready to fill an order, they send the goods and the invoice to the President's office, and his secretary passes them on to the shipping and accounting departments. At least no organization chart of which I am aware gives any guidance to the contrary. This may be the reason that organization charts and the "authority lattices" behind them have such a low reputation outside of organization and management departments. There is a great deal of discussion and some useful research which distinguishes between "formal" and "informal" organizational relationships, but a recognition that a great gulf so frequently exists

---

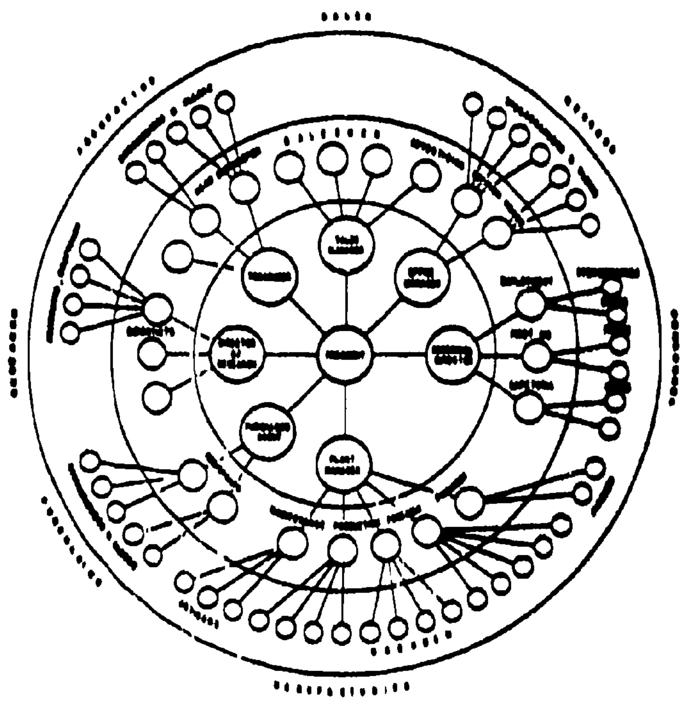
*Dr. Edward F. R. Hearle, "How Useful Are 'Scientific' Tools of Management?" PUBLIC ADMINISTRATION REVIEW, Autumn 1961, pp. 206-209.

**In making his statement Dr. Hearle had the type of mathematical tools which he had enumerated, and others like them, in mind. So my challenge to him relates to the inference which can fairly be drawn from his remarks rather than to any direct statement which he makes.



DISORGANIZATION CHART

Figure 2



Organization Chart *

Figure 3

*From COOPERATION AND CONFLICT IN INDUSTRY
by F. Alexander Magoun (Harper & Brothers, 1960).

between these structures is itself a witness of an inadequate formal structure in the organization--as is the fact that the disparity is so widespread a testimony to the unrealistic (and unhelpful) nature of current theories of organization.

The comparatively sterile status of organisation theory may be due to the inherent difficulty of the subject, the low ability of those who work in the field, or to the absence of one or a few essential concepts not yet sufficiently clearly delineated. The author is encouraged to propose that the latter may be an essential feature, and that the principles to be discussed may be included among the missing essential elements. Few of these points are entirely original, yet essentially none are clearly and widely appreciated.

Reflection of many years, exponentially enhanced in intensity in recent months, has led me to subsume the most significant examples of quantitative illusions in administrative practice which have come to my attention under 10 headings, partly because I have 10 fingers, and partly because this produced the acronymic title. This paper will consist of a listing of these 10 principles with a few exemplifications--not to convince anyone of the truth of my position on the examples; it would be impossible to discuss more than one or two in the little space allotted to me--but to demonstrate the clarification let into many administrative forms of action which otherwise must remain, as they were formerly to me and must still be to you, an enigmatic mystery. If the reader feels that my examples are inadequate or wrong, he is invited to refer to others from his own experience. Only if he feels that none or few can be found does he have a quarrel with my principles per se.

My first "quantitative illusion of administrative practice," (Figure 4), is entitled "Peas in a Pod," to emphasise that (a) the assumption that is generally made is that "if the name's the same, the thing's the same"--"as alike as peas in a pod," and (b) that the assumption is wrong. I see this as the exact opposite of Professor Hearle's (implied) position cited earlier. We can gain greatly by merely recognising this fallacy, without in any way being able to quantify it. This is at once the most important and the most fundamental of my 10 illusions. The fallacy lies in the denial of the reality and significance if quantitation in situations where it is real and important. Sometimes it takes the slightly more subtle form of acknowledging: yes, the several members of any one class do

PEAS IN A POD

Figure 4

differ and perhaps differ widely, but it is administratively impossible to allow for every possible variation; therefore, we will allow for none. The same punched card computing installation provided a glaring example of this "reasoning." We needed a card punch operator. The position analyst referred to a job standard which explained that card punch operators work from edited repetitions item records which are punched mechanically with no exercise of individual intelligence or ingenuity. The position analyst had no need to look at the facts when he had a book to tell him that (a) this position was identical with all other positions called by the same title, and (b) one such position filled the book specifications. Another exemplification of the illusion which receives a great deal of public attention is lowest bid procurement. Since all items (even those not yet invented) and all services can be exactly described in the invitation to bid, all are equivalent; and hence the purchase price is the one remaining variable. Of course, there is a contrary cliché: "you get what you pay for." There is no requirement in "administrative practice" that the system of clichés be consistent.*

My second illusion (Figure 5) attempts to deal with the assumption that all the good qualities reside in one product or one course of action, and, by inescapable logic, all the bad lie in any alternative--though if there are several, these bad qualities may be distributed among them. I am sorry to say that I was supplied with a perfect example of this

*It is a truly remarkable thing that philosophers, since the time of Plato, have been concerned with the problem of "nominalism" versus "realism," which, however, important theoretically, seems not to constitute a stumbling block in day to day relationships; whereas this first illusion is at once the most pervasive and most pernicious logical fallacy entering not just into almost every discussion between friend and foe, between advocate and adversary, but between even so intimately related and favorably disposed groups as members of one family. It was the essence of the "hyphenated-American" dispute, the merits and the abuses of political party labels and party loyalties, methods versus subject matter in education and of occupational jurisdictional disputes, whether within one organization or between competing parties or groups.

"Maybe she'd catch up if you didn't put so much stress on the importance of accuracy over speed."

PROFIT AND LOSS*

Figure 5

*Used by special permission of KING FEATURES SYNDICATE.

illusion in the 16 October 1964 issue of SCIENCE. Two of this country's most illustrious scientists explained their choice of candidate for President. I have examined these statements with some care. Neither protagonist could find a fault worth mentioning in his own choice nor a good quality in the latter's opponent. This action constitutes a conformance to the practice in disputation.* But it is the fact that the practice prevails in the day to day administrative process that concerns us. Why does it? The complete explanation must lie in psychology. A plausible treatment of just this phenomenon has been contributed by Professor Leon Festinger of Stanford University.** Of course, Professor Festinger is not responsible for my understanding or use of his theory. In essence, the mind demands harmony. Yet all real things and all real courses of action involve advantages and disadvantages. The only achievable harmony is a quantitative one--a balancing of opposing forces. But this type of harmony is uncongenial to many minds.

That, if one embraces one course of action or one belief, he must impute all virture to the chosen, and all evil to the rejected, was long ago recognized as a fundamental error by Georg Wilhelm Hegel, last of the global philosophers, who saw progress of social organization in the reconciliation of the thesis and the antithesis into a synthesis that removed the conflict by absorption of the thesis and antithesis as elements in a higher concept. Why Hegel never "caught on" in administrative practice I cannot say. But it is possible that, since his view was essentially qualitative and not quantitative, hence didn't in fact apply in many instances, his solution tended to be neglected even when perfectly applicable. An entirely analogous situation is known to have delayed acceptance of the contagion theory of disease for centuries.

---

*One of the authors was good enough to acknowledge receipt of an early draft of this paper with the statement that the allocation between the two discussions was a deliberate attempt to conform to the anticipated expectations of the readership of SCIENCE. I do not have the boldness to point out that insofar as the anticipation is correct, my strictures apply then to the readership if not the disputants.

**Leon Festinger, "A Theory of Cognitive Dissonance," (Evanston, Illinois, Row, Peterson, 1957) pp. 260-285.

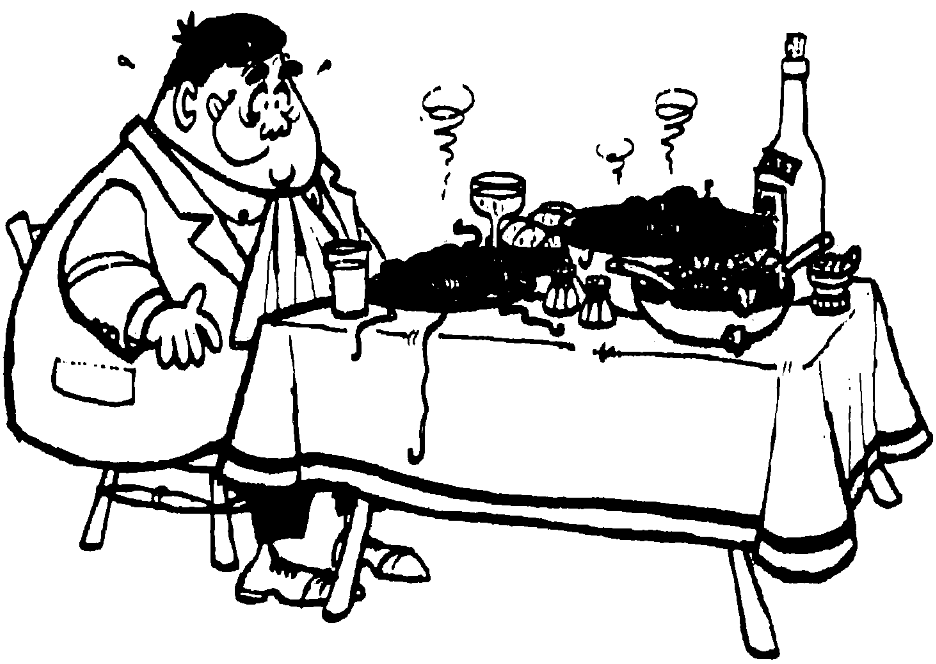
Those who escape the first illusion and recognize that quantitative variation pervades most, if not all, of life are candidates for the notion that, if a little is good, more is better. (Figure 6). The "spartan" philosopher would put that from the standpoint that, if moderation is good, abstinence is ideal. I have seen this illusion active in the question of where to put the statistician in an organization. The same holds for engineers, stenographers, air support, artillery, computers and machine tools. If the drawbacks of the widest possible organizational scattering are recognized, then complete centralization appears inescapable--and vice versa to most administrators, organization specialists, operations analysts and other people who move productive workers around. The most extreme proposal that has come to my attention is to centralize all computers in government. The one facet of all these matters that is of concern here is the human tendency, once embarked on a path, to assume that that path leads upwards (or downwards) indefinitely. The visible existence of "side effects" in therapeutics has compelled an avoidance of this illusion with full virulence in medical practice--but, as recent history shows, the tendency certainly exists.

The recognition of the existence--though I think not the nature--of this tendency has led those who prefer the status quo to take refuge in warnings against "the foot in the door," "the opening wedge," "the breach in the dike," "the camel's nose under the tent," with the consequence that anyone who fears extremism acts like and is characterized as an obstructionist. A recognition that the action or state rebelled against would lose its terrors (so that its advantages could be secured) if illustion three could be eliminated might remove much acrimonious social debate.

Forty years ago there was a principle in psychology, that I have never heard mentioned since, which involves a modified form of this fallacy and provides a partial explanation for its existence. I can best exhibit its nature by recalling to mind the once popular medical treatment of bleeding. General Washington was bled in his last illness and the practice continued, though with declining popularity, to the Civil War, and even lingered till World War I. Suppose a patient is "treated" by bleeding but succumbs. Three conclusions are possible. The bleeding either was (1) deleturious, (2) indifferent, Those who escape the first illusion and recognize that quantitative variation pervades most, if not all, of life are candidates for the notion that, if a little is good, more is better. (Figure 6). The "spartan" philosopher would put that from the standpoint that, if moderation is good, abstinence is ideal. I have seen this illusion active in the question of where to put the statistician in an organization. The same holds for engineers, stenographers, air support, artillery, computers and machine tools. If the drawbacks of the widest possible organizational scattering are recognized, then complete centralization appears inescapable--and vice versa to most administrators, organization specialists, operations analysts and other people who move productive workers around. The most extreme proposal that has come to my attention is to centralize all computers in government. The one facet of all these matters that is of concern here is the human tendency, once embarked on a path, to assume that that path leads upwards (or downwards) indefinitely. The visible existence of "side effects" in therapeutics has compelled an avoidance of this illusion with full virulence in medical practice--but, as recent history shows, the tendency certainly exists.

The recognition of the existence--though I think not the nature--of this tendency has led those who prefer the status quo to take refuge in warnings against "the foot in the door," "the opening wedge," "the breach in the dike," "the camel's nose under the tent," with the consequence that anyone who fears extremism acts like and is characterized as an obstructionist. A recognition that the action or state rebelled against would lose its terrors (so that its advantages could be secured) if illusion three could be eliminated might remove much acrimonious social debate.

Forty years ago there was a principle in psychology, that I have never heard mentioned since, which involves a modified form of this fallacy and provides a partial explanation for its existence. I can best exhibit its nature by recalling to mind the once popular medical treatment of bleeding. General Washington was bled in his last illness and the practice continued, though with declining popularity, to the Civil War, and even lingered till World War I. Suppose a patient is "treated" by bleeding but succumbs. Three conclusions are possible. The bleeding either was (1) deleterious, (2) indifferent,



SO MUCH THE BETTER

Figure 6

or (3) helped but was inadequate. In the third case, the remedy is even more heroic bleeding. This latter conclusion is associated most strongly with the name of Dr. Benjamin Rush, Professor in the first medical school in America.

In the psychological context this process--the persistence in a wrong course of action under the misapprehension, entirely sincere, that, while the strength was weak, the sense was right--was called the beta hypothesis. It seems most curious that such a fecund insight should drop from view.

But that it is more important to run in the right direction than to gain great yardage was redemonstrated in a football game Sunday, 25 October 1964, by Minnesota end Jim Marshall. A gyroscope persists on the course of its setting despite contrary forces. But does that make the setting right? Methods for determining the "direction of choice" have long been known in the science of statics and have, more recently, been investigated in economics. The glaring fact that scholars cannot agree on the direction of the consequences of an economic action is a greater blow to the status of the science of economics than the uncertainty of the magnitude of the effect in those few cases where there is agreement on the direction.

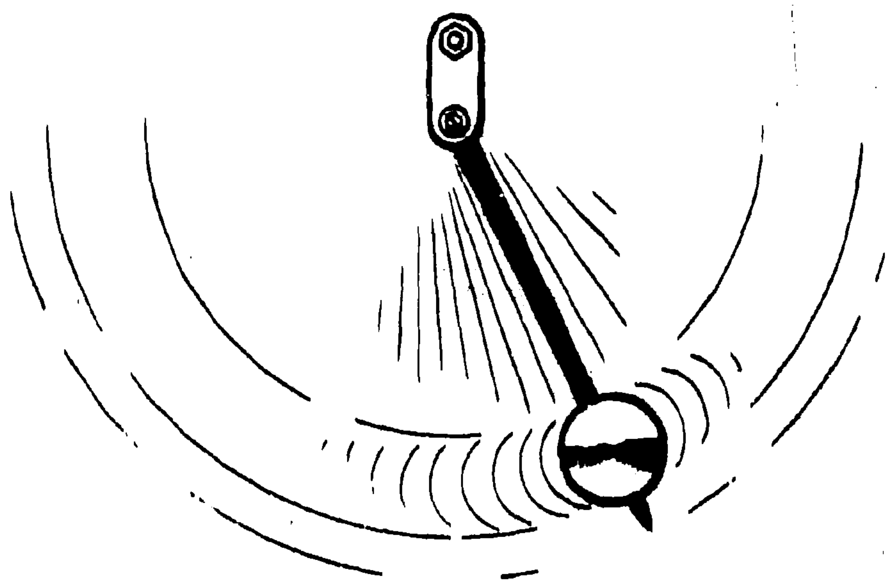
Innovators since the beginning of time have regarded all who saw merit in the old as obstructionists. That they may merely be victims of the beta virus has, so far as I know, never been entertained. Galileo's experience is the most famous, if not the most meritorious, example, but the theme is commonplace. It is a reliable story to tell in the movies, on TV, and in the pulp press. In real life it is as common in criminal prosecution, apparently, as in scientific innovation.

In the common view, stubborn insistence on one view and a refusal to even hear the evidence for another is a sign of malevolence on the part of the defenders of the status quo. In the psychological beta hypothesis it was a particular form of mental illness, possibly mild in character and little disabling. In a seemingly neglected paper in SCIENCE, the philosopher, Michael Polanyi, citing his own severe injury from exactly this process nevertheless defends it as essential to progress in science. He says: "there must be at all times a predominantly accepted scientific view of the nature of things in the light of which research is jointly conducted by members of the community of scientists. A strong presumption that any evidence which contradicts this view is invalid must prevail. Such evidence has to be disregarded even if it cannot be accounted for, in the hope that it will eventually turn out to be false or irrelevant."

This thesis of Professor Polanyi seems to be receiving just the treatment which he argues it must. Despite his evident eminence, both as scientist and as philosopher, he appears to be suffering from an acute case of the second (Profit and Loss) illusion, complicated by the presence of the beta virus. He sees that science, functioning as it does, advances. He does not see how it could do so were it to purify itself.

It is clear that the seriousness of the third illusion is unmistakable, once its reality is granted, but when combined with the second illusion (that one of alternative courses of action, degrees of centralization, size of computer, has all the virtures and none of the faults--or that one must at least act that way) explains, I believe, the peculiar property of progress in administration. (Figure 7) At one extreme, (form of organization, method of scheduling work, approval channels, work procedure, etc.) the adherents of the other will have ample proof of the inadequacy of the chosen solution--and they will be right. Fears of the "opening wedge" effect will, however, maintain the existing status quo until proponents of the opposite extreme gain sufficient ascendency to overcome these fears, when a shift towards a balanced solution will set in. Now, just when the optimum position is reached, the third illusion will add momentum to the swing until the opposite extreme is attained.

In consequence of the interaction of these two illusions, "Profit and Loss" and "So Much the Better," progress in administrative practice--and many other facets of human behavior--consists in discarding yesterday's procedure and adopting that of the day before, with the assumption and, indeed, the claim that the new is novel. There is no risk; for few will have survived from the earlier period, and they can be suppressed. The number who, not personally surviving, will have read history will be of measure zero.



ADMINISTRATIVE PROGRESS

Figure 7

Professor Jay W. Forrester of M.I.T. argues* that this result is in fact a necessary and not accidential characteristic of administrative progress under the existing circumstances. He writes: "In the past, management methods have been learned primarily through personal experience. The developing manager rotates through numerous assignments. Management schools repeat the folklore and the experiences of practicing managers. This experience is used as a basis for generalizing, so that past experiences can become a basis for anticipating the nature of new situations."

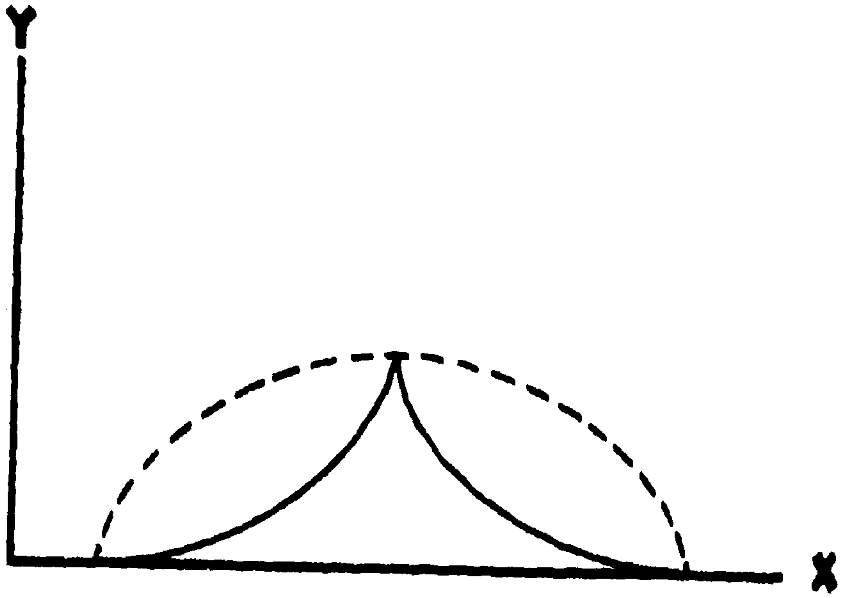
". . . we have here at the M.I.T. School of Industrial Management been developing an approach to management policy design which we call 'industrial dynamics.' It is intended to be a new way to understand how corporate structure and policy produce the different characteristics which one sees in business enterprises. . . . Most managers are surprises to learn that those practices which they know they are following are sufficient, when assembled in a system model, to cause the major difficulties which they have been experiencing."

Professor Forrester is too polite to do so, but I cannot resist observing that the jibe: "He has never met a payroll" is a pistol pointed backward.

Figure 8 shows another illusion widely prevalent but not recognised as quantitative. It is closely akin to illusion one, but differs in that here the underlying phenomena is recognised as quantitative, e.g., different grade levels for different duties, adjustments in the general level of compensation, fringe benefits, quality of tools or equipment, work space, and so on, but it is assumed that all maxima are cusps; that if some precise (even if unknown) value is optimum, then that even slight departures result in great waste of resources. My own chief experience in this field had to do with the erection of a building to house a computer and associated staff. We were asked to build for the future, but to justify in great detail just how every square foot would be assigned.

---

*Jay W. Forrester, "Dynamics of Corporate Growth." Paper delivered at a conference on "Management Strategy for Corporate Growth in New England," held at M.I.T. November 12, 1963. This brilliant paper deserves reading in its entirety for its positive approach to rescuing administrative practice from the grip of injurious if plausible "folklore."



THAT'S ENUF, CUSP IT

Figure 6

The interests of the government would be served if we were adequately housed, but, though the building would have a life expectancy measured in decades, it seemed to be assumed that any attention to contingencies beyond the needs of the instant, (e.g., an anticipation of the tortuous delays in the path to actuality) whether of our unit or of other units on the Post, could only be "wasteful." Opposition to grade escalation; to better than "necessary" research facilities, military aircraft, weapons, uniform belt buckles, or missile boosters fall in this category. If I say that I think this line of "thought" led to the Russian scoop in space, the importance of this illusion, if not its actuality, will be obvious.

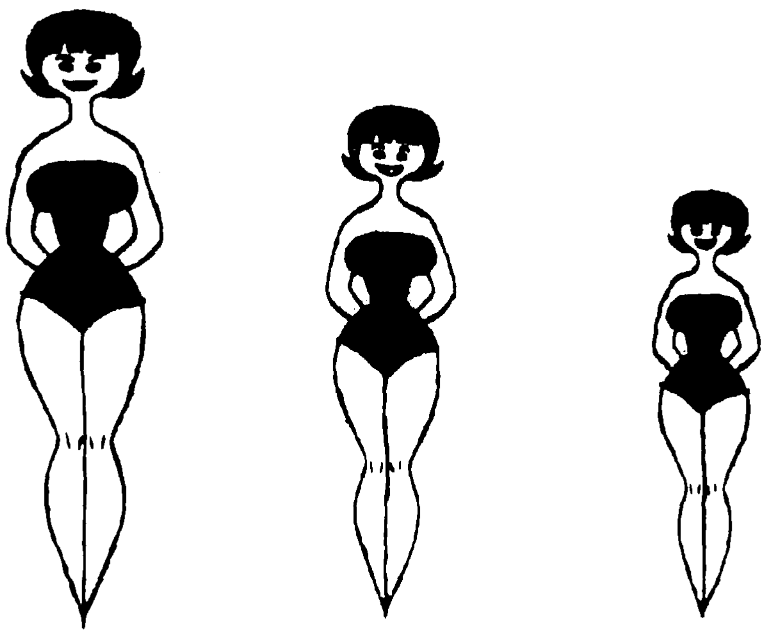
In his treatment of "The Economics of American Medicine" Seymour E. Harris,* it seems to me, suffers severely from this illusion when he refers to the "waste" of better than "necessary" medical attention or hospitalization.

Illusion five (Figure 9) is more conventional. At first blush one would assume that the factor which varies between the three figures is one of height only, despite the fact that we all know that the human figure is a three-dimensional object; hips, waist, and of course, height. Perhaps others. I include this illusion only to pay homage to a factor well recognized in all circles, mathematical and anti-mathematical, the multi-dimensionality of real life problems. The reader is asked to stretch a point in the interest of economy of illustration and view the figures as also suggesting the non-linear (curvilinear) character of most realistic situations. Again, the availability of the computer could push these difficulties back a step or two, were the nonquantitatively oriented administrator aware of the potentialities.

This illusion is unique in that it occurs with great frequency in two opposite forms. The tendency to "oversimplify a problem" is widely indulged--but widely condemned. Indeed the opposition to automation, to "thinking by computer," while reaching a crescendo in modern times, has had its Cassandras in all ages.

---

*Seymour E. Harris, "The Economics of American Medicine" (New York, Macmillan, 1964).



A ONE DIMENSIONAL VIEW

Figure 9

These anti-mathematicians assert that no amount of complicating the model can hope to mirror reality and that the administrator deals with reality. There is a complementary consequence. The administrator, like the observer, "never knows what he is talking about or whether what he concludes is true," and without recognizing it, he has conceded as much by the above argument. Philosophers of the scientific method have repeatedly pointed out the contrast between the rigor of conclusions derived from contrived but designed experiments and from passive observation of a complex, if real, world.

A particularly striking example of this phenomenon appeared in a recent column of Walter Lippmann, concerning the place in history of Herbert Hoover. Lippmann says: "we avoided such a crash [the Great Depression] after the Second World War because we had so well learned the lessions of the First World War." Clearly, we didn't learn the larger lesson of how to avoid war. And is our continuing prosperity a reward for a lesson well learned or a penalty for a cold war still in progress?

A widely publicized illustration of this phenomena has arisen in the controversy over smoking and disease. The evidence in the human is observational and not experimental. The same administrators who most adamantly cling to this objection to an unwanted conclusion, ,ver hesitate to advance the sales force on the basis of their "proven record." That these two attitudes are inconsistent is my only point.

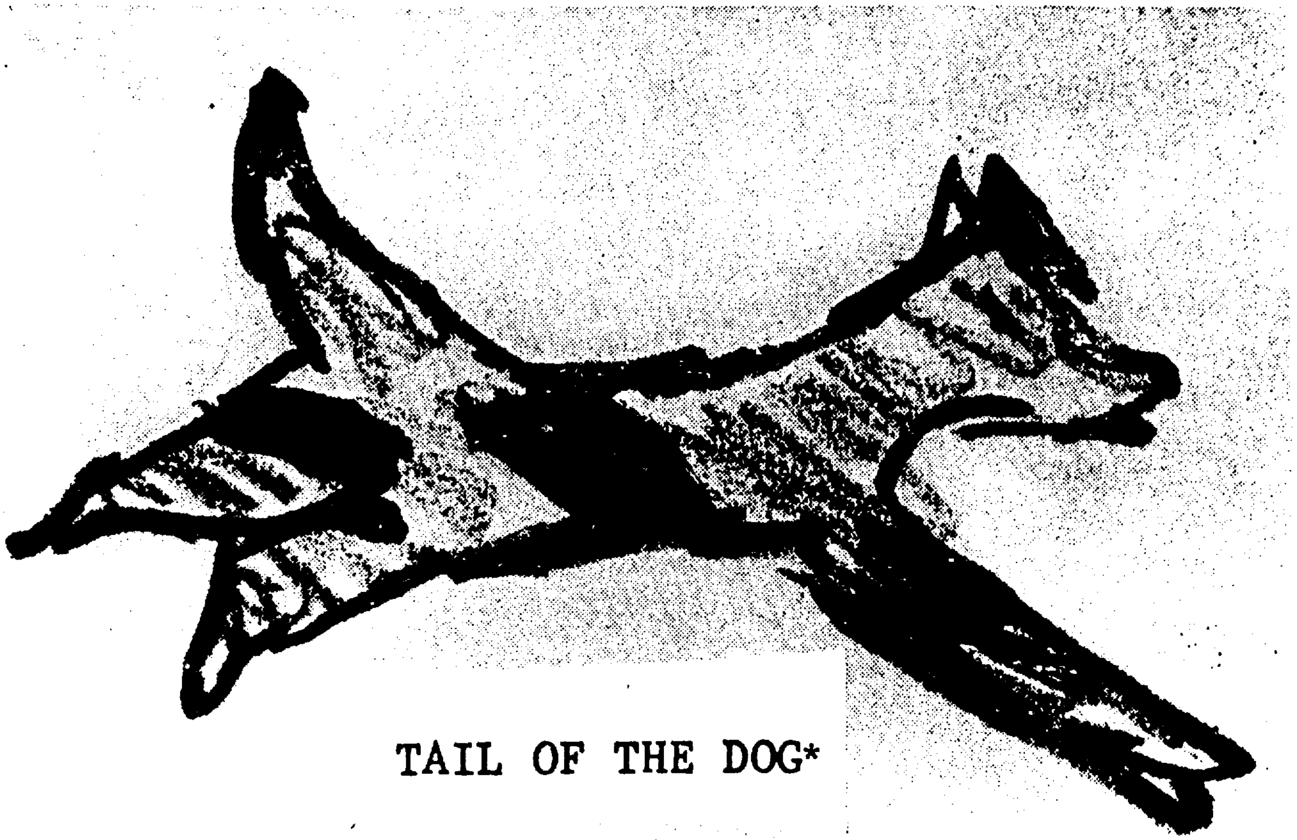
All soothsayer type forecasting of the future, including forecasts of election outcomes, depends for its acceptance on the fact that the public suffers from the misapprehension that one success is a guaranty of superhuman powers. Business chooses its tycoons, and government its "bright young men" by this technique. I hasten to add that it guarantees the wrong choice no better than it does the right. It may be called the one step (two step, k-step) beta decision rule.

The complexity that defeats description denies certainty. If the administrator must act today, guided only or principally by his "intuitive grasp" of the "total situation" and not on the basis of previously enunciated rules thought to embrace the situation at hand, then he cannot carry over any lesson for tomorrow whether he is successful or a failure today.

That experience confirms error as often as truth is hardly a novel observation. This is in fact the beta hypothesis. One does indeed learn by experience, but at present only by an extravagant volume of repetition, which then forces an abstraction into the consciousness of the "man of action." Once verbalized, experience can confirm or refute the hypothesis. Again, the beta hypothesis asserts that experience will lead to the explicit formulation of wrong inferences quite as much as right. But just as bleeding as a medical treatment finally gave way to overwhelming contrary evidence, so must any belief in conflict with reality, if only there is sufficient experience since, ultimately, despite the influence of the forces described in the beta hypothesis, there will be an attrition of incorrect deductions and an enrichment of correct ones. There is a statistical technique to deal with such cases--Sequential Analysis--but its applicability in administrative contexts must so far have occurred to only one individual.

True, the practical administrator would never assemble the number of observations suggested by Sequential Analysis before making his decision--except in matters of trivial moment. He is thereby protected by (and confirmed in) his belief that his poor batting average belies the very quantitative reasoning he ignores!

Illusion six (Figure 10) is the most pervasive of all, for we each see the world only from our own point of view. In Adam Smith's day it was argued that the interplay of universal self interest would produce a general maximum of social well being through the action of an "unseen hand." It has since become clear that the hand is either unsteady or unfriendly. But the same illusion has taken refuge in the cry "let the experts decide." Curiously, it is in the military field where the dictum of Clemenceau: "war is much too important to leave to the generals" is most known and most accepted. Clemenceau should have realized that if he didn't say that road building is too important to leave to the engineers and management too important to leave to the managers that no successor would arise capable of that generalization. The experts should be left to their own devices--when they are concerned with matters of interest only to themselves. In particular, "research managers" are only effective when they adhere to Jeffersonian principles of government (not his actions). What happens when diverse activities impinge--each the field of a different group of "experts"--is the subject of illusion seven.



TAIL OF THE DOG*

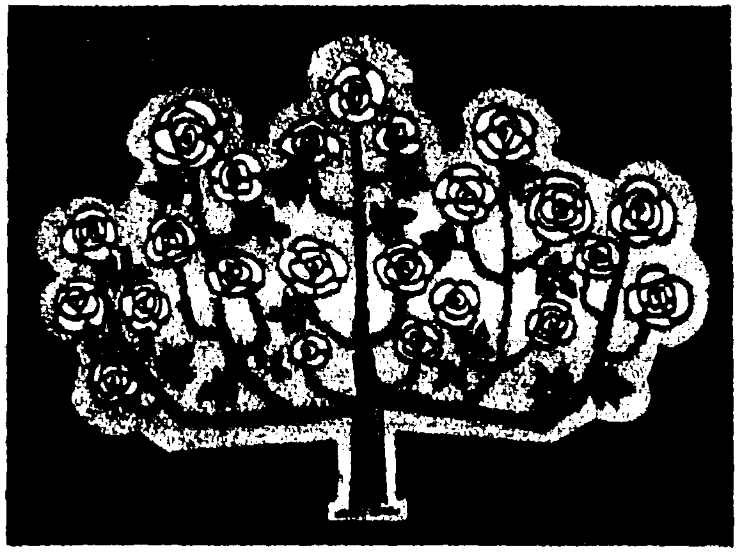
Figure 10

The "blight of the expert" takes two forms. In the previous paragraph I discussed the effect of letting the expert "do his own job" where it, however, impinges on others and/or on their capacity to do theirs. The second form is to extend the expertise from the field in which it is earned into other distinct fields where it is not. This illusion is related to but quite distinct from illusion one. In "Peas in a Pod," the assumption is that, for example, every engineer is interchangeable with every other engineer. This second form of the "Tail of the Dog" illusion is that any engineer (physician, military officer, educator, or bricklayer) and only an engineer (physician, military officer, educator or bricklayer) can perform any other function regardless of the skills involved in an engineering (medical, military, educational, or construction) organization. The example that caught my eye here is: * "I think all good statisticians agree--and I define a good statistician to be one who agrees--that statistics is not mathematics. On the other hand, it happens to be a peculiar subject of its own, which mathematicians when they do take the trouble, can teach much better than non-mathematicians."

Illusion seven (Figure 11) is intended to conjure up in the viewer's mind not just the genus Rosa but the whole field of classification. Classification is associated most securely in the popular mind with the fields of zoology and botany--but is somewhat less frequently recognized as a universal tool of science. Languages, rocks, and forms of tribal kinship relations are classified. The most famous classification in mathematics is that of Klein, but others continue to arise in every quarter. Second only to biology the public comes in contact with classification at the public library. Librarians, at least in America, will stress that a universal classification is not obtainable--but administrative practice has not yet learned this lesson. The form in which this illusion enters here is in the search for the holy grail of the perfect organization.

---

*John G. Kemeny in "New Directions in Mathematics," proceedings of conference arranged by John G. Kemeny and Robin Robinson. Edited by Robert W. Ritchie. (Englewood Cliffs, New Jersey, Prentice-Hall, 1963). I strongly suspect that Professor Kemeny made the above remarks with tongue in cheek and has succeeded in pulling my leg. Why else would he commit such a transparent logical non sequitur?



A Rose is A Rose is A Rose

Figure 11

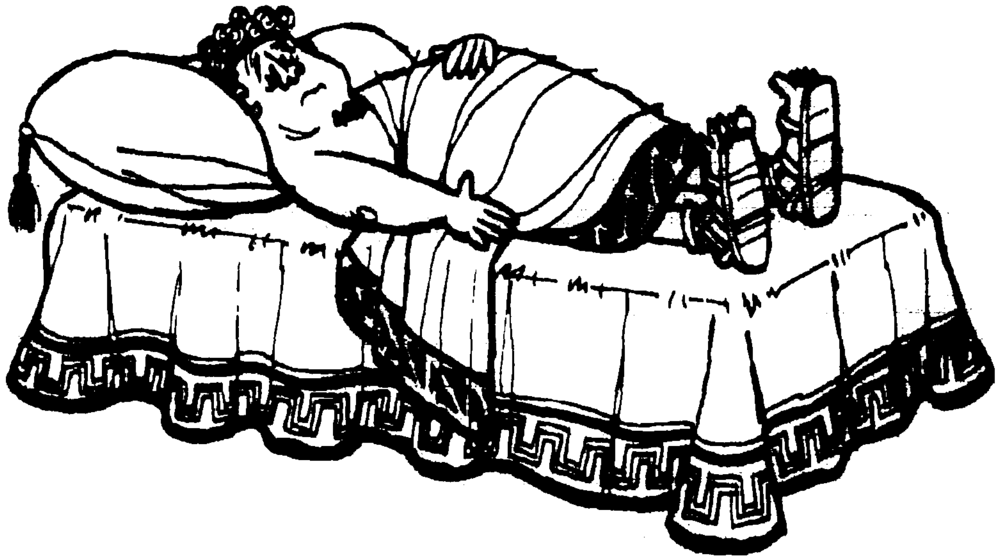
That a satisfactory organizational classification for the distribution of responsibilities has not yet been achieved is shown by the fact that we are forever reorganizing. What I wish to call attention to here is that this process is never interrupted to check on the existence theorem, nor to seek a fruitful setting for an examination of the problem. Organisation is the replacement of the disorganization chart by the organization chart. Bureaucracy is the conversion of the paths in the disorganization chart from paths of persuasion to paths of coercion. One testimonial to the fact and to the consequences was the establishment by the Army of the "Program Managers."*

In the Navy's Polaris Program, (the one instance success which convinced the Army) according to an article in the Civil Service Journal for July-September 1964, Admiral Raborn was authorized by letter to get "whatever people and whatever cooperation he required from any of the Navy's bureaus and offices." Yet "Admiral Burke admonished him that if the letter ever had to be used to force cooperation, the project would fail." Notice that Admiral Raborn went out of channels, he used the organization of the disorganization chart, he used these lines as lines of persuasion--not as lines of control. This is where organization began!

Procrustes' bed (Figure 12) was the ancients' way of putting the principle of what we now know as "the organization man." In America every man is an individual--"just don't step out of line." I wish to unite this principle with another not perhaps widely known but at least not due to me. It has been remarked that never in history (of course; seldom in history) have the professionals invented a radical departure destined to lose them their jobs.

In an article, "Revisionist Theory of Leadership" by Professor Warren G. Bennis in the Harvard Business Review for January-February 1961, Volume 39, pages 26 ff, the author on page 148 observes: "along these lines, Samuel Goldwyn was reputed to have said to his staff one day, 'I want you all to tell me what's wrong with our operation--even if it means losing your job!'" I thought in Hollywood telling the boss what he was doing wrong was a sure way to lose one's job. I'm sure that

*The beta virus strikes again! It is something of an achievement, surely, that bureaucracy is to be cured by increased bureaucracy.



PROCRUSTES' BED

Figure 12

Professor Bennis, like me, thinks this story is aprocrophyl--but it at least lists one impediment to people with an assigned mission ever making very radical changes in how its done.

The National Road starts at Braddock's Rock near the Lincoln Memorial and runs through Frederick, Maryland, and then on west. It has been said that the waggoners on the road didn't develop the canal system, the latter didn't develop the railroad. No railroader developed the automobile or the airplane--indeed, didn't even develop the diesel locomotive. But administrators have not yet drawn the obvious conclusion. If a department is charged with a certain responsiblity, then that department will never introduce radical changes. If anyone does, someone else will do so, But so soon as he does, he will have the organisation manual thrown at him. Anything he may have done will be destroyed and the task will be transferred from one who cares and knows to one who fears and opposes. At least that was my experience, and of several of my former colleagues.

This consequence is particularly unfortunate, since if my claim that the "research manager is best who manages the least" is correct, then this is the environment which will yield the largest number of outstanding results. But the consequence of illusion seven (A Rose Is a Rose Is a Rose) will be that such a manager will get a low rating and be accused of permitting "excessive duplication." Even here I can claim no priority. Albert Hirschman and Charles Lindbloom in an article "Economic Development Research and Development Policy Making: Some Converging Views," in BEHAVIORAL SCIENCE, Volume 7, 1962, pp. 211-222, explicitly recognize the benefits of a little "play" in the tight constraints placed on research and development efforts.

Illusion seven (A Rose Is a Rose) and eight (Procrustes' Bed) together are at the bottom of much of the ferment over "management" of research. It seems so simple to divide up the field of action and portion out support, responsibility, and facilities to each. We may even take a leaf from the military field commander, and recognizing that liaison between adjacent commands is the weak point in a battle line, we can be tolerant of a certain amount of interpenetration at the peripheries. But innovation arises from the darnedest sources. If it arises in the civilian economy or at least from non-governmental sources, we need merely create an Office of Scientific Research and

Professor Bennis, like me, thinks this story is aprocrophyl--but it at least lists one impediment to people with an assigned mission ever making very radical changes in how its done.

The National Road starts at Braddock's Rock near the Lincoln Memorial and runs through Frederick, Maryland, and then on west. It has been said that the waggoners on the road didn't develop the canal system, the latter didn't develop the railroad. No railroader developed the automobile or the airplane--indeed, didn't even develop the diesel locomotive. But administrators have not yet drawn the obvious conclusion. If a department is charged with a certain responsiblity, then that department will never introduce radical changes. If anyone does, someone else will do so, But so soon as he does, he will have the organization manual thrown at him. Anything he may have done will be destroyed and the task will be transferred from one who cares and knows to one who fears and opposes. At least that was my experience, and of several of my former colleagues.

This consequence is particularly unfortunate, since if my claim that the "research manager is best who manages the least" is correct, then this is the environment which will yield the largest number of outstanding results. But the consequence of illusion seven (A Rose Is a Rose Is a Rose) will be that such a manager will get a low rating and be accused of permitting "excessive duplication." Even here I can claim no priority. Albert Hirschman and Charles Lindbloom in an article "Economic Development Research and Development Policy Making: Some Converging Views," in BEHAVIORAL SCIENCE, Volume 7, 1962, pp. 211-222, explicitly recognize the benefits of a little "play" in the tight constraints placed onresearch and development efforts.

Illusion seven (A Rose Is a Rose) and eight (Procrustes' Bed) together are at the bottom of much of the ferment over "management" of research. It seems so simple to divide up the field of action and portion out support, responsibility, and facilities to each. We may even take a leaf from the military field commander, and recognizing that liaison between adjacent commands is the weak point in a battle line, we can be tolerant of a certain amount of interpenetration at the peripheries. But innovation arises from the darnedest sources. If it arises in the civilian economy or at least from non-governmental sources, we need merely create an Office of Scientific Research and

SOUND AND SUBSTANCE

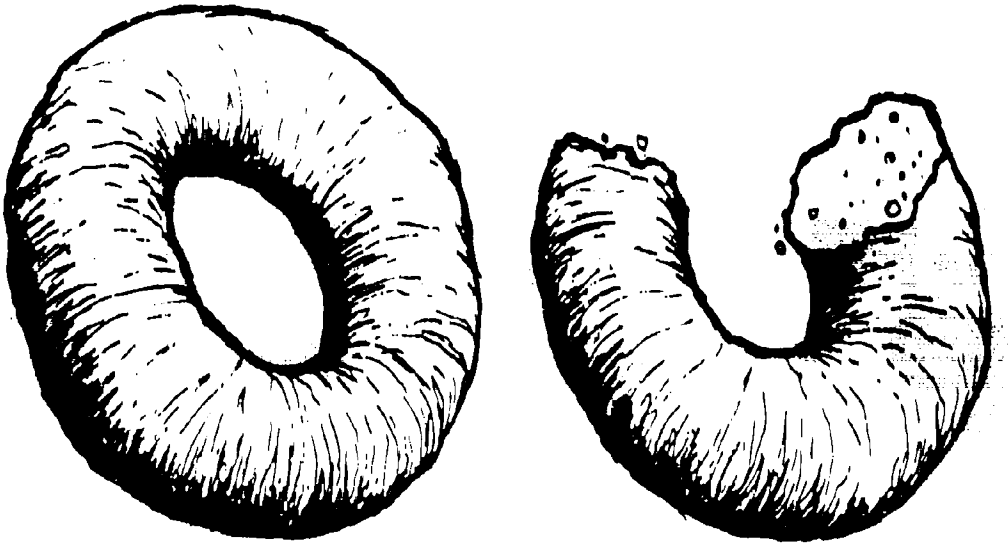
Figure 13

The financial pages of the daily papers have more than once announced the appointment of some outstanding administrator as president of a company--sometimes to rescue it from financial difficulty. The aspect I wish to draw attention to here is that in a number of such instances we learn that the former wizard has now been replaced by a new one. The criteria by which administrative competence is to be measured seem not yet always to have separated sound from substance. That is is more important to look right than to be right is the foundation of those two esteemed aspects of business-advertising and selling. It is also the foundation stone of politics (at least of the office variety). No author of fiction or movie director has the slightest uncertainty as to the great separation between appearance and reality--or as to which is better paid.

My final illusion is perhaps the most subtle. That the whole is more than the sum of the parts (Figure 14) has been most vehemently asserted by those who most abhor quantitative thinking. But the claim has been more often used as a bar to thinking than as a basis. It is the source of the claim "your vote does count"--though I think not for the reason usually given. The true reason is that, if the public would agree, it could rule, and agreement does not require that we assemble an overwhelming force, but merely that we inform ourselves. Of course, it does little good for just one voter or a few voters to do so--the answer lies in the whole (nearly) doing so. Hence it doesn't occur. Hence the principle.

This fallacy has been at the root of the argument over international trade since the Industrial Revolution. Adam Smith wrote his book to set the argument right, and while he did convince the economist, he didn't convince anyone who counted. Our individual fortunes are controlled primarily by our fate as producers, but our collective fortunes are equally involved as consumers. Since the ordinary person considers his individual (his producer) function first and foremost he penalizes himself and his fellow citizens by supporting protectionism. The most widely recognized recent large scale instance of this illusion has been exposed by John Maynard Keynes in connection with governmental fiscal and monetary policy. Fortunately, some of his resulting conclusions can be supported on valid grounds.

Public support of law enforcement "crackdowns" or of any other punitive or control measure, e.g., traffic crackdowns, compulsory



WHOLE AND PART

Figure 14

vehicle inspection, depends in large part on an assumption that it is always "the other guy" who will be affected. The risk of personally being victimized is small. But it is not zero. And the consequences (where the action is misguided) injures society as a whole--over and above the fact that, of course, while wrong actions are in progress, right actions must wait.

An exemplification closer to our topic and of vital importance lies in the field of experience. No one's own experience is "whole" till completed, whether of an individual, a firm, or a nation. Hence one cannot profit by experience as a whole while he can yet experience it. If the whole differs from the sum of the parts, it will have no effect. But the very meaning of a "rare" event is that a brief experience, possibly even that of a life time (whether of a person or of a nation) is not long enough. It has been claimed that while a fool profits by no one's experience and an ordinary man only by his own, a wise man profits by everyone's. Yet even a wise man cannot profit by experience he doesn't know about. And there are many fields in which experience is so rare that it is next to impossible to assemble it (i.e. complete it). The Constitution says the right of the people to keep and bear arms shall not be abridged. This was based on experience. But no living man can have had their experiences, and how to apply them to modern conditions is not obvious. Every teen-ager begins with much confidence and little experience and ends with more experience than confidence. That gangsters, racing drivers, entrepreneurs, dictators, explorers, and inventors base their future actions on their own past but necessarily incomplete experience is insufficiently recognized. Had all past experience been considered, many of these occupations would become deserted.

The constant advice to the young and gullible to "take a chance" in choosing a career and so secure an opportunity to win a handsome financial reward is not seen to be at one with advice to gamble on the stock market--or on a slot machine. And, of course, the prospect of a material reward (or fear of material failure) is assumed to motivate all choices in our society.

That statisticians are professionals at extracting the lessons of others' experience is recognized in a few circumscribed fields like research, development, and testing, or quality control, or acceptance sampling, but not, it seems, in opinion polling--or in safeguarding the [illegible] of a president.

My final figure enumerates these ten quantitative illusions of administrative practice. It is my belief that if administrators generally, almost universally, acquired an instinctive capacity to recognize instances of any of them on sight, despite their infinite capacity for disguise or fragmentary manifestation, then rational administration just might some day be possible.

TEQUILAP

1. Peas in a Pod
2. Profit and Loss
3. So Much the Better
4. That's Enuf, Cusp It
5. A One Dimensional View
6. Tail of the Dog
7. A Rose Is a Rose is a Rose
8. Procrustes' Bed
9. Sound and Substance
10. Whole and Part

FIGURE 15

COMBAT VEHICLE FLEET MANAGEMENT

C. J. Christianson and G. E. Cooper
Research Analysis Corporation, McLean, Virginia

Over the last six years the Operations Research Office and the Research Analysis Corporation have studied a variety of US Army vehicles from a maintenance-oriented point of view. All the studies have held the fundamental premise that to a greater extent many Army materiel decisions and programs should be based on certain mechanical properties of materiel in the real troop-machine environment. In the past the rarity of pertinent, real data too often resulted in the mechanical properties' being ignored or unrealistically appraised. It was apparent that many performance objectives were far from adequate indicators of actual achievement. ORO then and RAC now have sought to introduce realistic measures of mechanical effects into managerial decision processes. Of necessity this management research mission had to be confined to limited numbers and types of equipment. In order to determine meaningful equipment policies, it is necessary to consider troop-performance data in combination with a variety of monetary costs and with relative obsolescence. Neither dollars, nor obsolescence, nor mechanical performance alone can be expected to give infallible guidance to the necessary consideration of equipment management. Occasionally technological breakthroughs do occur, and sometimes a particular equipment model does develop a rash of breakdowns. However, in the long run the Army has to program much of its inventory between the relatively gradual changes in the designs for new production and the mechanical aging of previously manufactured equipment.

The most current study of Army vehicles has been selected as a general example of the ways in which RAC has contributed both greater qualitative understanding and improved numerical assessment techniques to the solution of important materiel management problems. Certain of the results have been disguised for relatively open presentation; however, wherever possible, numerical examples have been kept along the scales of their true values. Indeed, there is often a need to force experts to recognize the order of magnitude of the effects with which they profess intimate qualitative acquaintance.

1—The Research Analysis Corporation recently completed a comprehensive, multi-stage analysis of three different types of tracked vehicles operated and maintained in US Army combat units in Europe. A main objective of such study was to determine in-use lives of tanks, armored personnel carriers, and recovery vehicles. In addition to fulfilling its principal objective, the study program provided a variety of by-products essential to successful management of combat vehicle fleets. Measurement of fleet wearout, establishment of repair capacity requirements, projection of budget needs, assessment of combat readiness, and determination of fleet replacement factors (procurement requirements) are all management responsibilities that cannot be successfully fulfilled without basic information and analytical results of the type provided by RAC.

PROBLEM

COMBAT VEHICLE FLEET MANAGEMENT

Utilization
Component Replacement Forecasting
Readiness

Study conducted for the U. S. Army
by the Research Analysis Corporation

Figure 1

Inasmuch as the same analytical procedures were applied by the study to all three vehicle types, little generality is sacrificed if most further discussion is limited to just one vehicle--the tank. The determination of in-use lives is treated here as only a part of the general problem of tank fleet measurement. Within a manageable framework equipment life is interrelated with factors of utilization, component replacement forecasting and costing, and materiel readiness, to name only a few. The factors cannot be divided into one list of independent and another of dependent variables. They can be functionally related, but any adopted conventions of dependence and independence are likely to be unrealistic statements of causes and effects.

In all RAC equipment studies the starting point of meaningful concepts has always been a body of carefully collected empirical data. The hypotheses preceding observation have almost always had to be so modified under the light of experience that it now proves preferable to avoid prejudicing early theories. The logical elegance of classical scientific method has to be modified in favor of diffusely related series of observations resulting in near-saturation with both pertinent and extraneous data. Data processing and analysis then have to be not so much testing of theory as they must be sorting of the relevant from the irrelevant. Of necessity the rules of relevancy must be continually modified as more and more data are digested. And discouragingly it often becomes essential to completely reprocess earlier work.

Fig. 2—The body of empirical data used for the most current study accrued over the complete history sample is now shown. Maintenance events that occurred against all these vehicles were recorded by vehicle number, age, and mileage. The column on the right gives ample evidence of the fact that large numbers of combat vehicles are used extensively in peacetime. Peacetime is unquestionably a time of appreciable materiel consumption. That the severity of peacetime demands is still not universally recognized in all related civilian and military agencies continues to provide one of the major obstacles to completely successful combat vehicle management.

SAMPLE

		Average		
	No. of vehicles	Months observed	Miles observed	Usage miles per mo.
Tank	640	21	2739	130
Armored Personnel Carrier	708	18	2655	148
Tank Recovery Vehicle	83	19	1485	78

Figure 2

Fig. 3—One of the fundamental operations performed on the body of empirical data is the construction of mileage-dependent parts costs. The average direct costs for tank repair parts are shown for each successive 500-mile increment to 4500 miles. Track and engine replacements account for most of the dollar consumption. One of the most important features of parts consumption is the strong increase with mileage beyond 1500 miles.

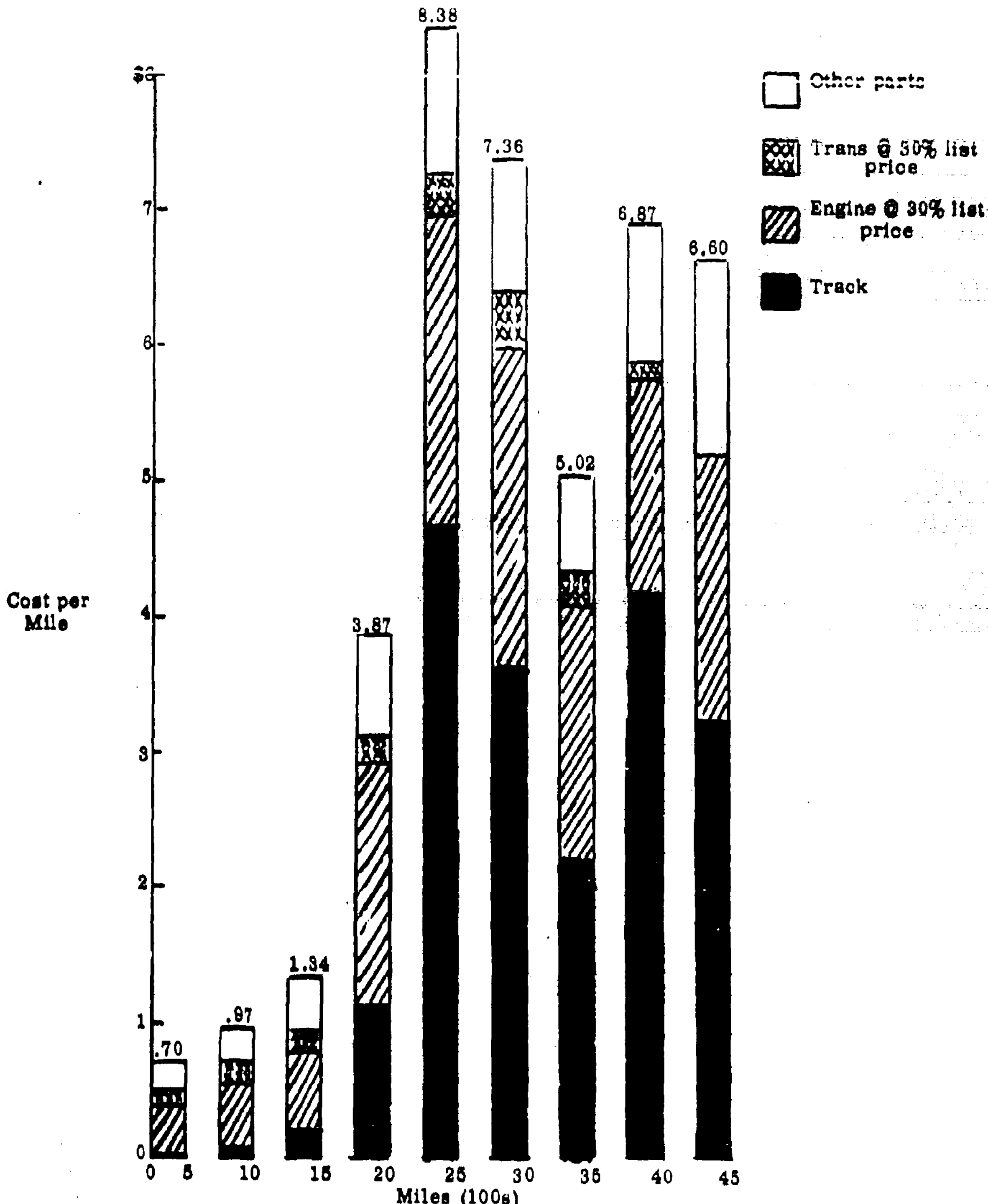


Fig. 3– Average Cost Per Tank by 500-Mile Increments
Total Miles Observed 1.75 Million

Through 1500 miles parts expense averages very close to $1 per mile. Beyond 1500 miles both track and engines enter pronounced replacement phases. Modal track replacement occurs at about 2200 miles. The dollar effect of this mode is to drive parts costs above $8 per mile in the 2000- to 2500-mile interval. Beyond 2500 miles there is a definite decline in expense, but a second generation of track replacement beyond 3500 miles forces the costs up again. From a separate examination of the same basic data it was estimated that, if all subsequent installations of parts and assemblies provided lives similar to those of the originally installed parts, an equilibrium cost would just exceed $6 per mile for a list of the principal mobility-affecting parts. Inclusion of weapon and fire control costs would add to the $6 per mile figure. The changes in parts consumption with mileage obviously have tremendous impact of the provision and budgeting of parts and maintenance support. If tanks are operated at 1500 miles per year, support during the first year of tank life be only one-sixth that required during an equilibrium year. Too much support planning is still dependent on an assumption that each new year is going to be like the last one. The importance of making predictions based on analyses of trends is revealed by data of the cost type.

Fig. 4—For the same tank sample, the purely historical cost average to nearly two years was $3.75 per mile. The truth of such history does not make the past a guaranteed base for prediction of identical futures. Vehicle support must be programmed in calendar time. In order to provide accurate support it is necessary to know both the basic consumption rates per mile and the mileages to be covered during given calendar intervals. Again in the case of the tank sample, adjustment of the average mileage cost to the average use rate yields monthly costs of $489 per tank per month. At the same rate of travel, but using tanks with much higher mileages, corresponding expenses would exceed $780 per tank per month.

Figure 4
PARTS COSTS FOR TANKS
(Usage: 130 miles per month)

	Cost per mile	Cost per month
Engines*	$1.19	$155
Transmissions*	.19	25
Other Parts	.32	42
Track--amortized at $2.05 per mile	2.05	267
Total	$3.75	$489

*30 percent of list price.

Fig. 5—The occurrences of replacement events are more fundamental than their associated costs. One of the advantages of the format employed for collection of the basic history data is that mortality statistics for individual parts types can be determined by straightforward actuarial calculation. For example, the comulative engine replacement experience is shown for the total tank sample. The accumulation of replacements increases non-linearly to just beyond 2000 miles, but then the engine activity becomes almost uniform.

Fig. 6—The replacement of original engines was interpreted as having been governed by an underlying distribution like that labelled first. That distribution is represented by two distinct phases. The first, extending to just beyond 1000 miles, averages only about one-half percent replacement per 100 miles. This first phase corresponds closely to the typical early "debugging" period so common to many kinds of equipment. However, unlike the frequently discussed "bath-tub" effect, tank engines do not then experience a phase of reduced replacement activity. Rather, the tank engines immediately being a second phase with sharp increase in the replacement rate. The second phase corresponds to what one would expect with entry into a wear-like mode of engine mortality. The presented distribution of mileages to replacement has a mean of 3700 miles.

If it is assumed that all succeeding engine installations will result in lives like those of the original engines, second and third replacements will occur as shown by the distributions labelled 2d and 3d in the figure. Such higher order replacements have been determined by taking repeated convolutions of the first distribution. The sum of replacements of all orders is often described as the renewal density. It represents the instantaneous replacement rate disregarding the order of a replacement. A pure coincidence of the chosen first replacement distribution is that its corresponding renewal density approaches equilibrium quickly but smoothly. In fact, by 3000 miles the renewal density is almost equal to its equilibrium value. The absence of strong oscillations in replacement activity is usually of great advantage in the prediction of maintenance support requirements. Somewhat later comments will be made by way of explanation of real effects that tend to drive engine replacement activity above the so-called equilibrium presented here.

Fig. 7—Conversion of a renewal density of corresponding budget costs is a relatively easy matter. Multiplication of the renewal density by the unit cost of a replacement yields the desired measure of expense. As an example of the type of results derived by such a processing operation,

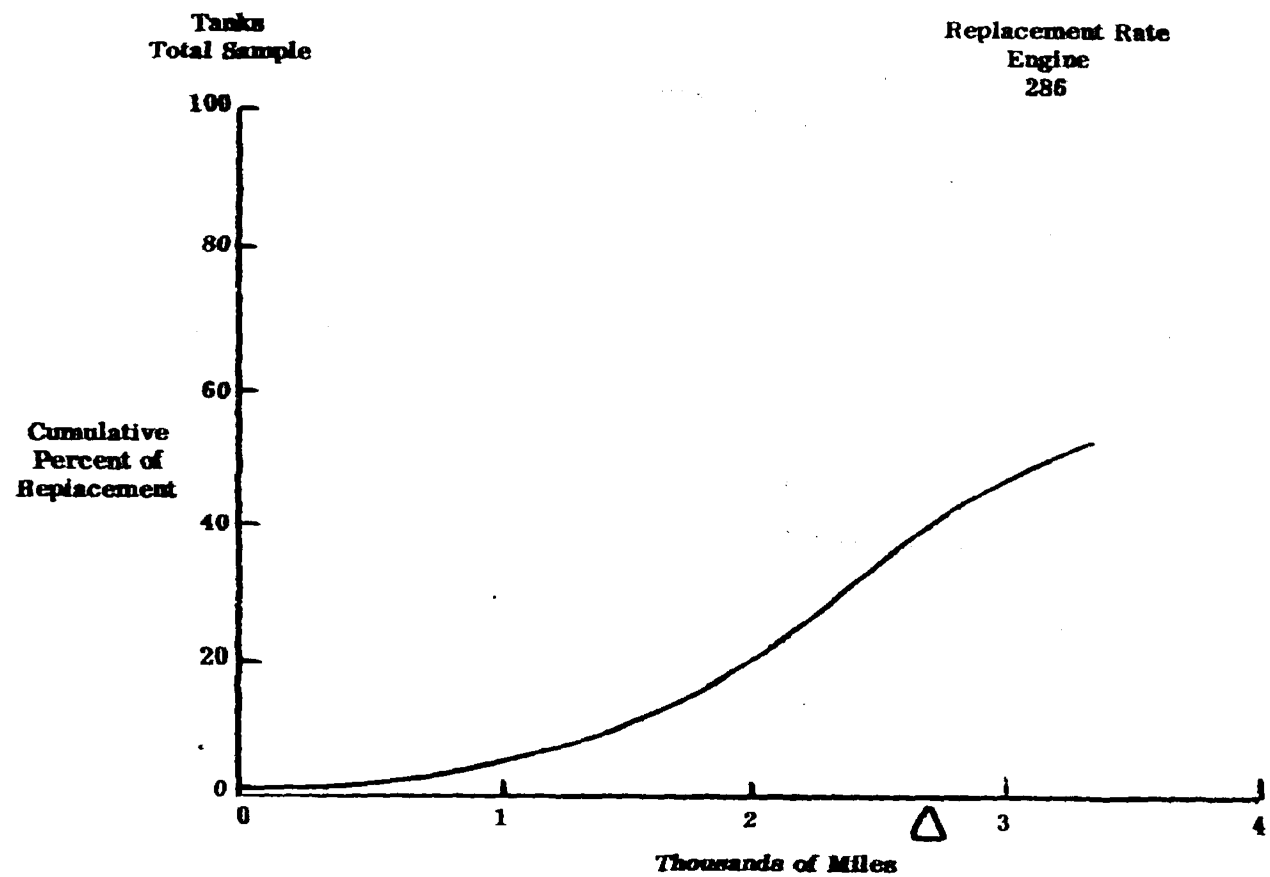
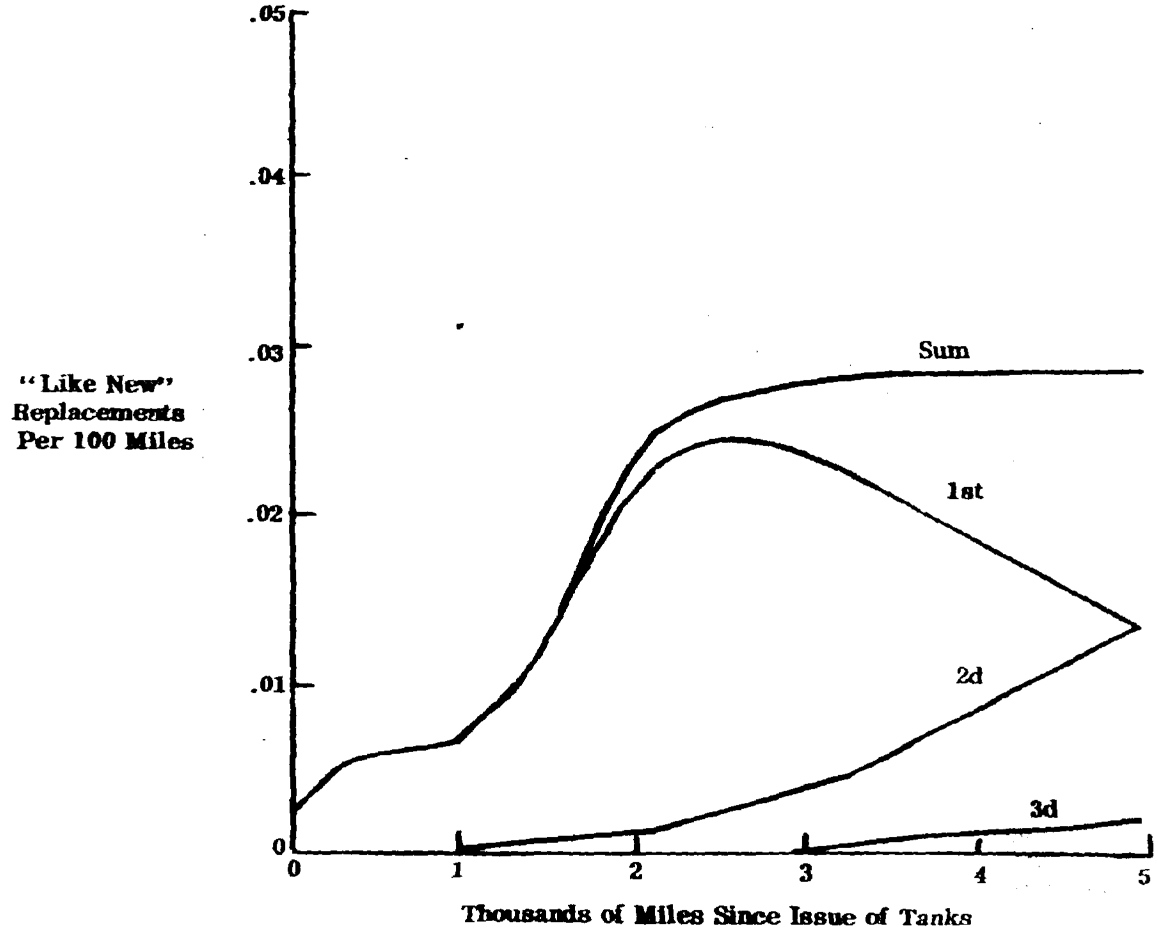

Tanks
Total Sample
Replacement Rate
Engine
286
Cumulative
Percent of
Replacement
100
80
60
40
20
0
0
1
2
3
4
Thousands of Miles

Fig. 5

Fig. 6—Tank Engines at 130 Miles/Month



a cost rate dependent on mileage is shown here for engine and transmission replacement activity. Such information represents interpretation and extrapolation of the originally collected empirical data. Transmissions have lower unit prices and longer lives than do tank engines. This double advantage to the transmissions makes their costs per mile considerably lower than that of the engines. Note that the total expense for these two assemblies accounts for nearly one-half the previously mentioned equilibrium cost rate of roughly $6 per mile.

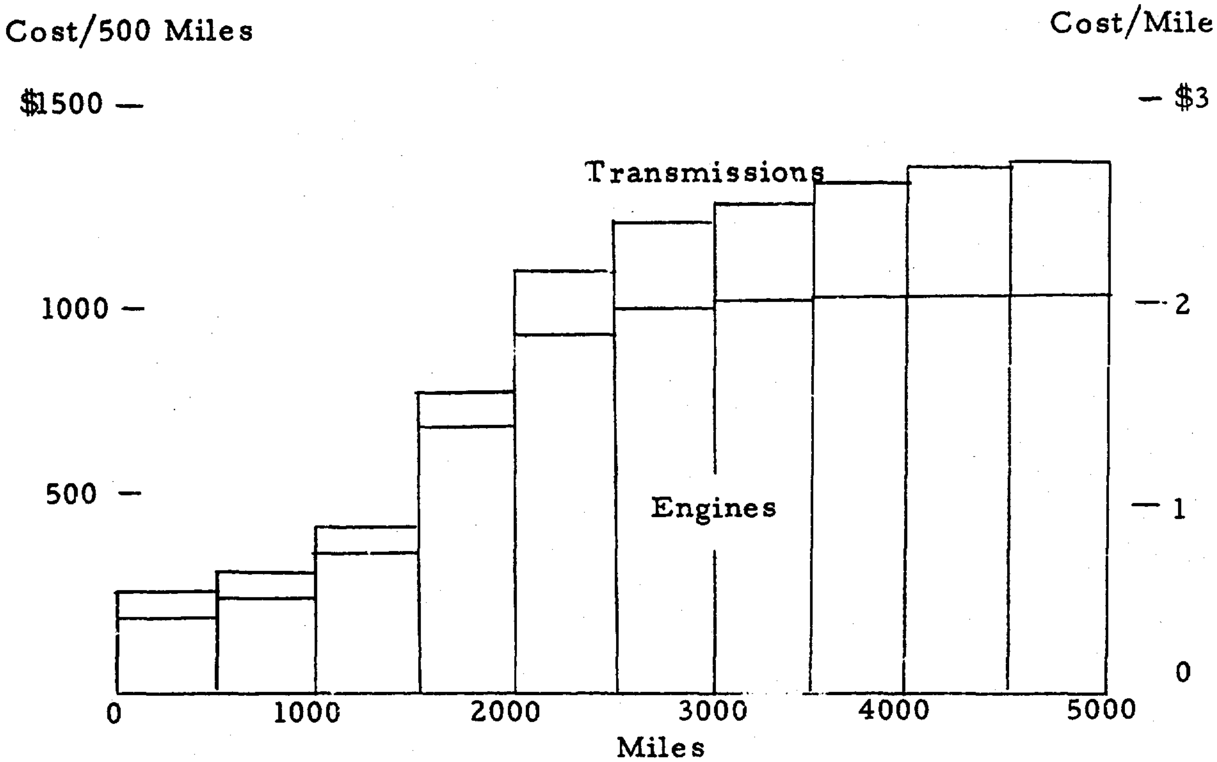


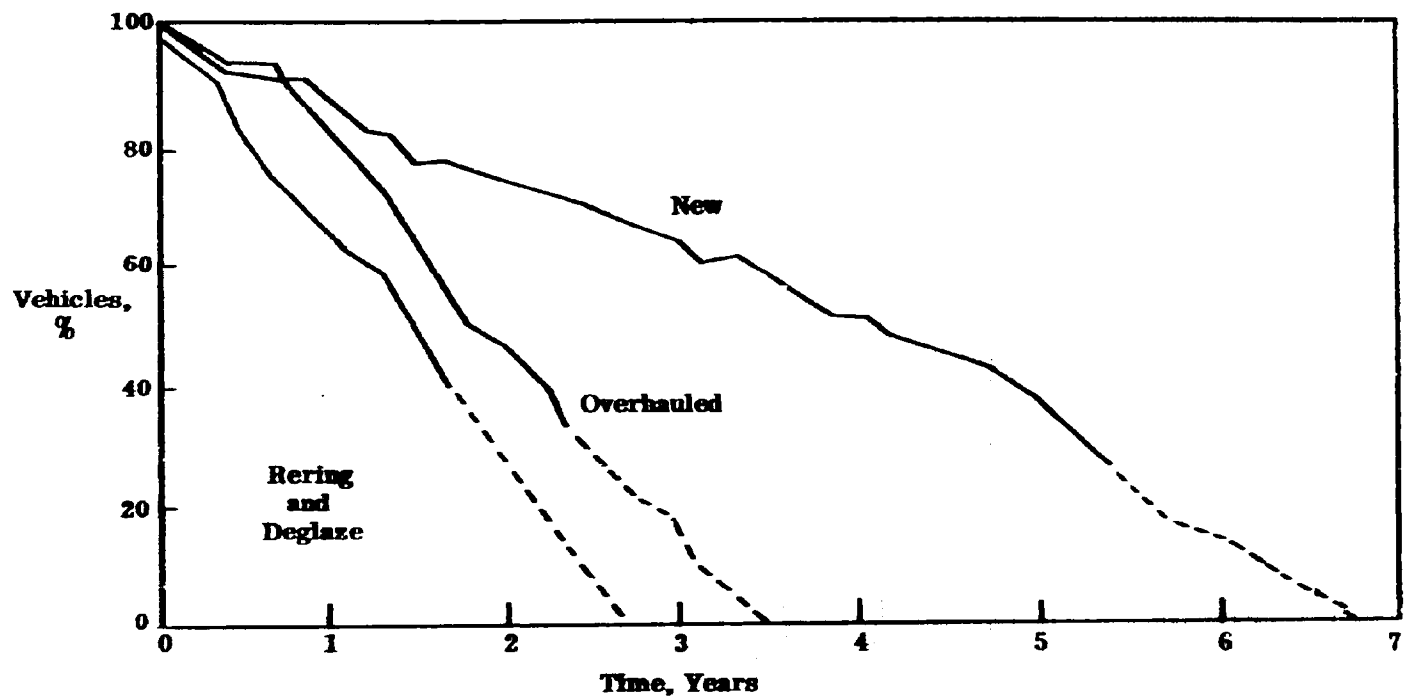
Figure 7--Tank Major Assembly Replacement Cost "Like New," 30% of List at 130 Miles/Month

Reality usually has a way of muddying the clear waters of analysis. The preceding engine and transmission predictions are all founded on an assumption of performance from replacement assemblies like that of the originals. Unfortunately very little data have accrued about the performance of replacement engines in the model tank most recently studied.

Fig. 8—However, the experiences of a preceding model tank and its engines are not ignorable. There survivor curves are shown in the projected figure. All these were determined for the other tank. That tank's original engines lasted as shown by the line labelled "new." Engines that were overhauled by a depot maintenance facility survived only as long as shown by the curve labelled "overhauled." Other engines were repaired by a mobile team (fourth echelon in the US Army). The lives of engines repaired by that team were even shorter and are represented by the line labelled "re-ring and de-glaze." After several years virtually all the engines that are installed in used tanks are repaired ones. The possibility must be recognized that equilibrium engine replacement activity may run twice as high with repaired engines as that determined from the originally performing assemblies.

Fig. 9—For the current model tanks, second engine replacements have been running somewhat higher than expected. Only about 10 percent of the second engines are known to have been repaired ones. The rest are presumed to have been issued from storage in unused condition. If the second engines were to do as well as the originals did only 4 percent of the tanks would experience a second replacement by 3000 miles. The predicted "like-new" experience is represented by the lower solid line in the accompanying figure. Empirical experience is shown by the lower dotted line and runs just over 8 percent replacement to 3000 miles. The real replacement mechanism seems to be one that is only partly renewal-like. Tank age regardless of assembly age appears to provide an important component of the probability of replacement. By now listeners are probably wondering how the pure renewal model can be of any use of management if so many of reality's disturbing influences tend to raise activity much higher. The renewal model provides a valuable reference point. First of all, the renewal estimates have already revealed that many support programs are set even lower than these predictions. The pure renewals usually specify the best that can be expected. If support is too low and is not geared to satisfy the demands of the best possible, it should be raised immediately to at least the level consistent with a

Fig. 8– Survival Rate, Tank Engines



renewal prediction. And second, the renewal estimates provide the basis of comparison of performance of repaired, stored, or modified assemblies with original quality. Too often statements are made that some component is good or bad without expressing goodness or badness relative to a real mechanical standard. The usual comparisons with paper standards have caused more confusion than they have eliminated. Even original engines do not compare favorably with their paper standards by which they were designed, built, and procured for Army use. Although performance of materiel in the hands of troops cannot be the ideal, absolute standard of reference, it remains the best, most meaningful basis of comparison at the present time.

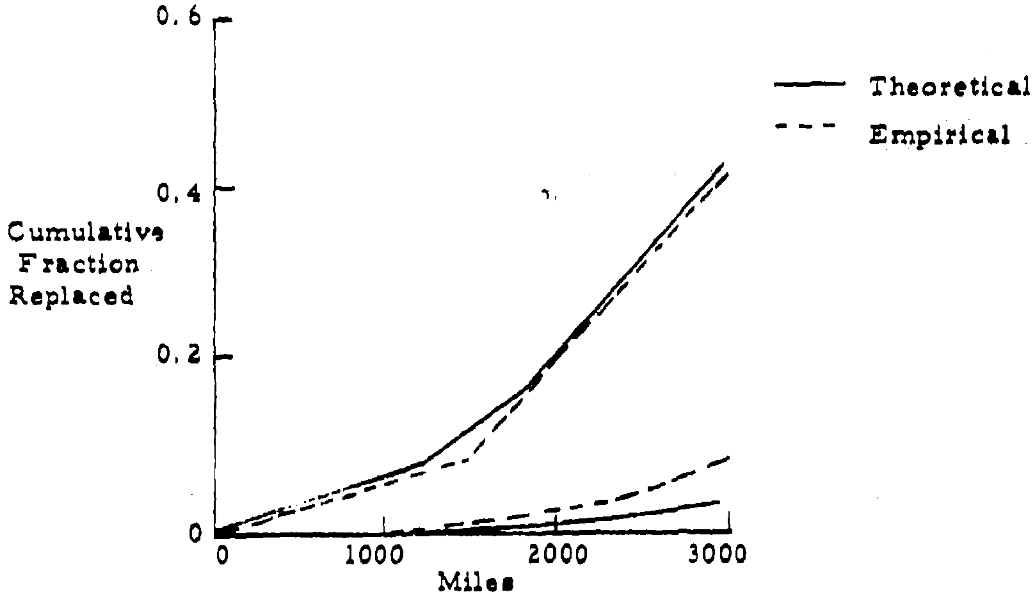


Figure 9--Comparison of History with Expected Like-New Renewals of Engines

Fig. 10—The replacement distributions of most parts and assemblies are widely distributed about their means. A particularly important example of a part that does not have this property is tank track. The lighter line in the projected figure shows the cumulative replacement experience for sets of tank track. The S-shaped curve represents the most compact distribution of replacements discovered in the whole series of studies of ground vehicles. A complete set of track is expensive. At the same time it has the shortest life among all the high-unit-priced parts of a tank. Conversion of life and unit price to costs per mile yields an estimate of $2.05 per mile for tank track. Although engines and transmissions cost more than track, their lives are sufficiently long to make their mileage costs lower than that of track. Track expense accounts for roughly one-third the predicted equilibrium parts costs for the mobility-affecting systems of tanks. A cumulative curve for second replacements is also shown. RAC pointed out that the early climb of the second curve was in part attributable to the use of much shorter-lived rebuilt track.

Fig. 11—Much effort has been directed toward interpretation of the empirical history relative to notions of materiel readiness. No one definition has proved satisfactory as a full description of readiness, but several less general notions have proved particularly useful. One concept that has been extremely helpful is that of "equipment availability potential." An item of equipment is considered to be assignable to one of several states of serviceability. For example, consider the four states shown. An item can be serviceable or it can be in one of three (or more) states of unserviceability. That item can undergo transitions from one state to another with probabilities associable with each particular type of transition. The "k's" can be associated with breakdown, and the "λ's" can be related to correction or repair times. The transition probabilities may depend on ages, mileages, generation numbers, or other factors. Too often it is not possible to detail all the inter-relations. In fact much of the time the greatest utility arises from considering a two-state model; that is, one state of serviceability and only one state of unserviceability. The instantaneous "availability potential" is defined as the probability of being in the serviceable state if transition probabilities were to permanently retain their current values. When the transition probabilities change sufficiently slowly, the availability potential gives a suitable measure of actual current serviceability or availability. Constant levels of availability potential may be represented by hyperbolas on a "k-λ" plane.

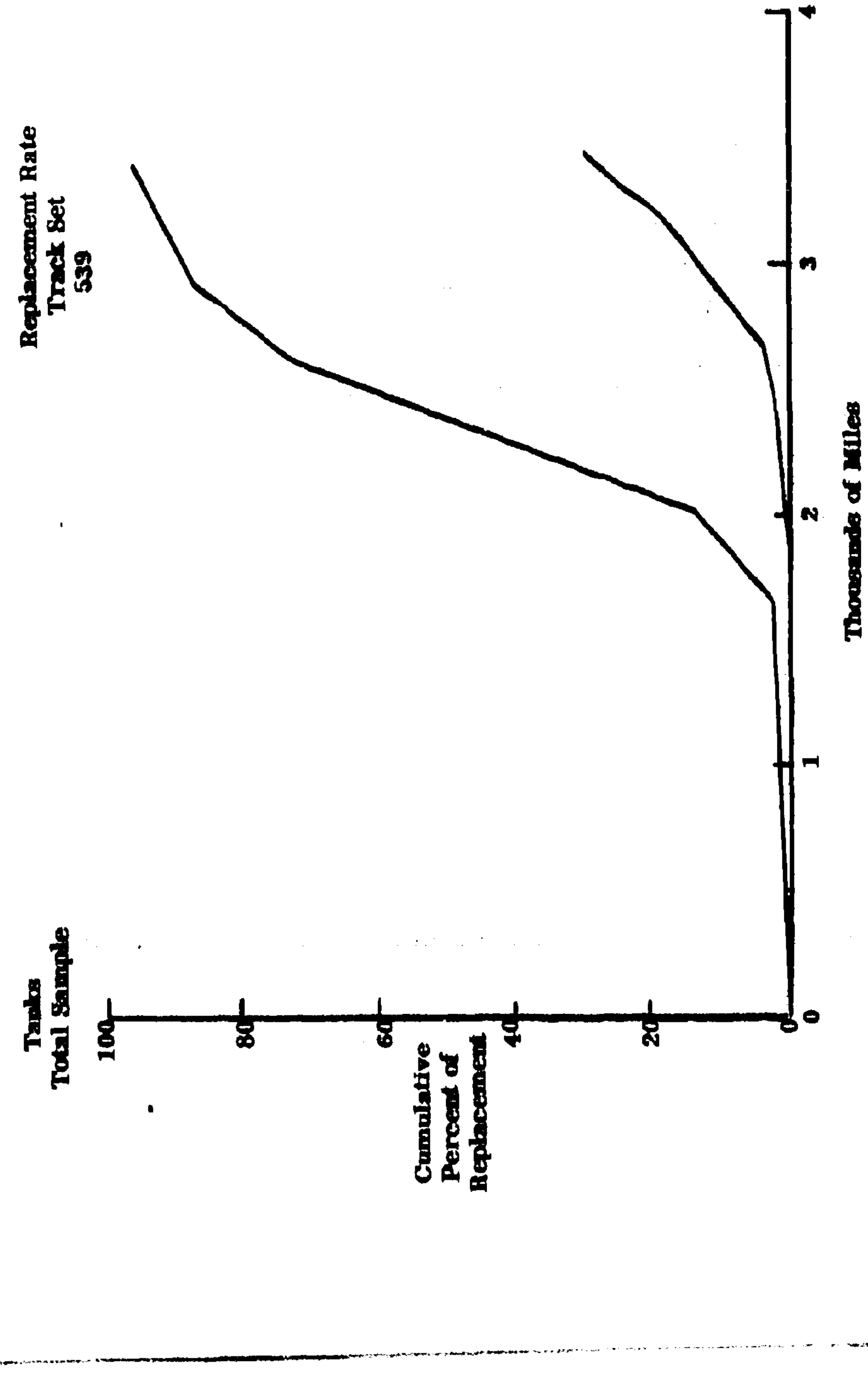

Replacement Rate
Track Set
539
Tanks
Total Sample
Cumulative
Percent of
Replacement
100
80
60
40
20
0
0
1
2
3
4
Thousands of Miles

Fig. 10

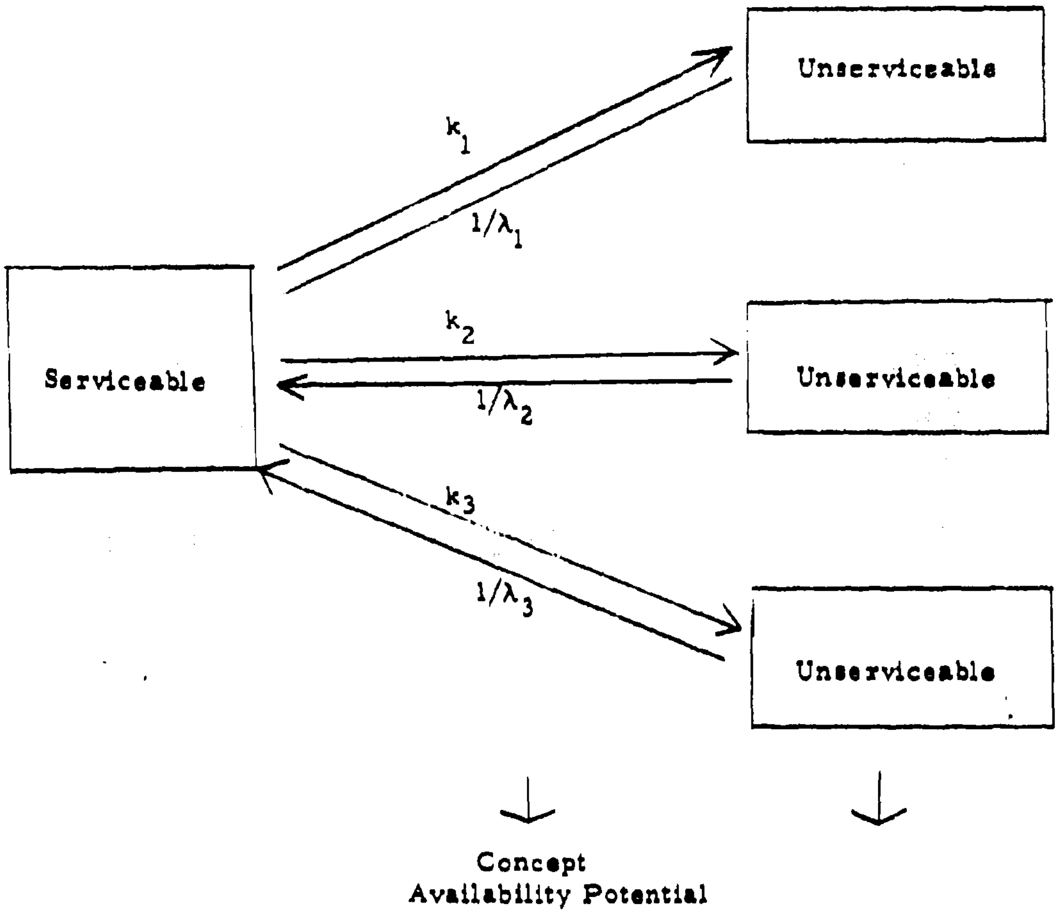
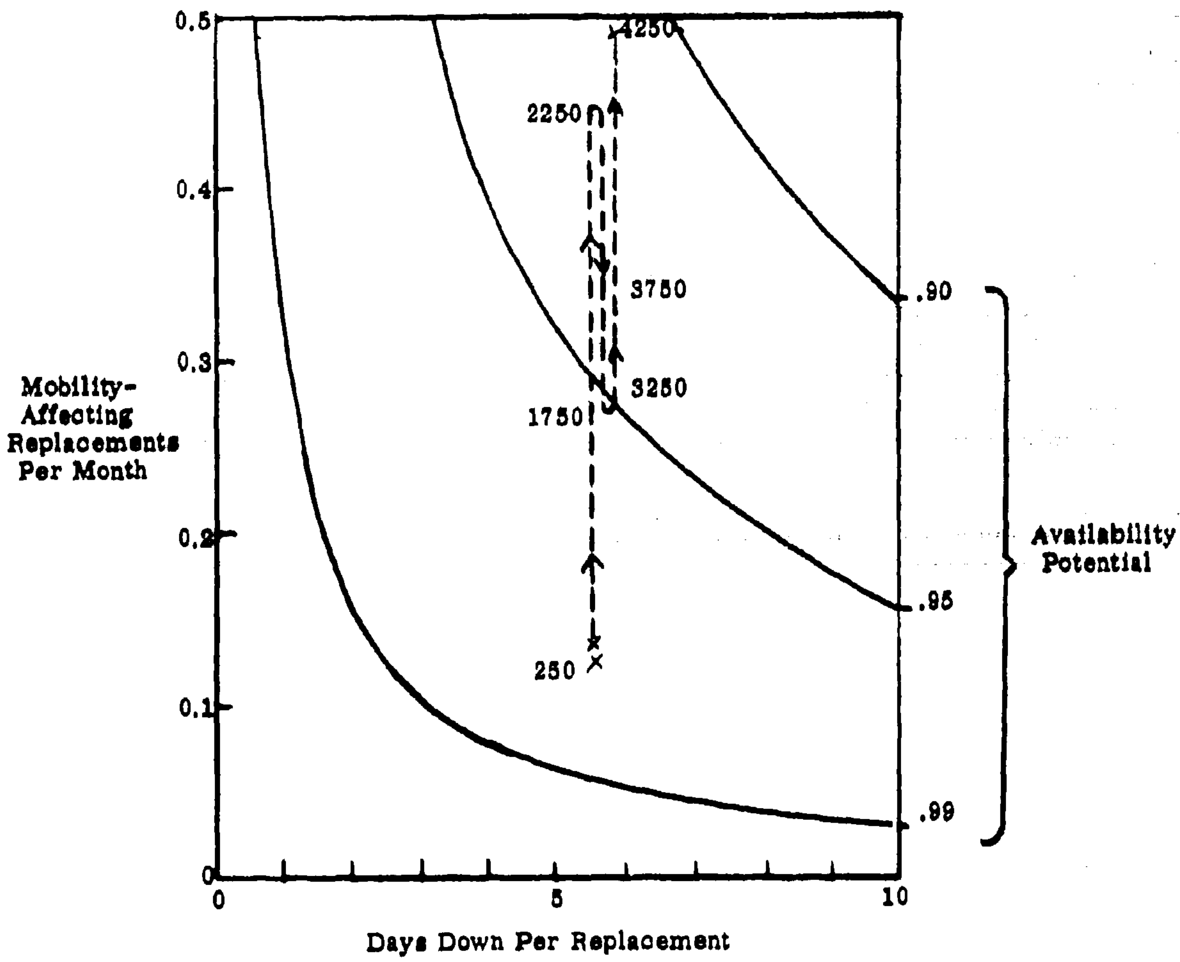


Figure 11

Fig. 12—Analysis of the mobility-affecting parts replacement data from the same tank history sample led to the mileage dependent availability path shown. During the entire history period the average response time stayed very close to 5.7 days per replacement job. However, because the rate of replacements per mile was changing, the availability at 125 miles per tank per month had to change as shown. During their early lives the studied tanks were operated with close to 0.98 availability potential. As they accumulated additional mileage, the tanks lost

Fig. 12--Tank Availability Potential
At 125 Miles per Month



availability. During the period of most track replacements, the availability dropped to about 0.95. Then followed a period of some improvement; the availability climbed to 0.95. However, beyond 3250 miles availability again dropped and reached about 0.91 in the interval 4000 to 4500 miles.

Fleet availability need never be considered an unmanageable aspect of operation and maintenance. Rather it is only a result of several directly manageable factors. Product quality, support response time, and rate of equipment use all affect equipment readiness.

Fig. 13—A great deal of attention is usually given to the supposed or predicted differences in product quality among different models of equipment. The normal approach of salesmanship is to promise that some new model will provide mechanical advantages far beyond the cabilities of its predecessor. Too often the demonstrated comparisons examine an unused, new model and an over-used, old model. The RAC studies have discovered that many models of different generations appear so different only if they are examined at different stages in their lives. At the same ages different models of tanks possess greater similarity with respect to parts replacement rates than do tanks of a single model at different ages. Judicious utilization of a particular tank model can increase overall mechanical capability more than can a transition of models amid a less carefully designed program of vehicle use.

As an example far less pronounced than reality, consider this figure. Suppose that at a uniform rate of use some tank model has the renewal density shown with respect to time. That density increases smoothly. A change of utilization can have a three-fold effect with respect to time. An increase in use, in effect, squeezes time by having the higher accumulated mileages occur that much sooner in time. Thus, in a given month the replacements per mile are higher. Then because more miles are traveled during that month, the replacements during a given month are given a second boost. The third effect may be to somewhat alter the mileage lives depending on the rate of use. It is not illogical to consider the possibility of increasing or decreasing mileage lives depending on the type of component involved. Actually it is usually sufficient to suppose that the occurrence of events per mile depends on the accumulated mileage but not on the rate of use. This assumption is nearly correct within the mileage ranges normally encountered and corresponds to having(u) equal to 1.0.

Fig. 13—Three-Fold Utilization Effect

$$h(t,u) = \left(\frac{u}{u_o}\right)\phi(u)\, h\left[\left(\frac{u}{u_o}\right)\phi(u)\, t,\, u_o\right]$$

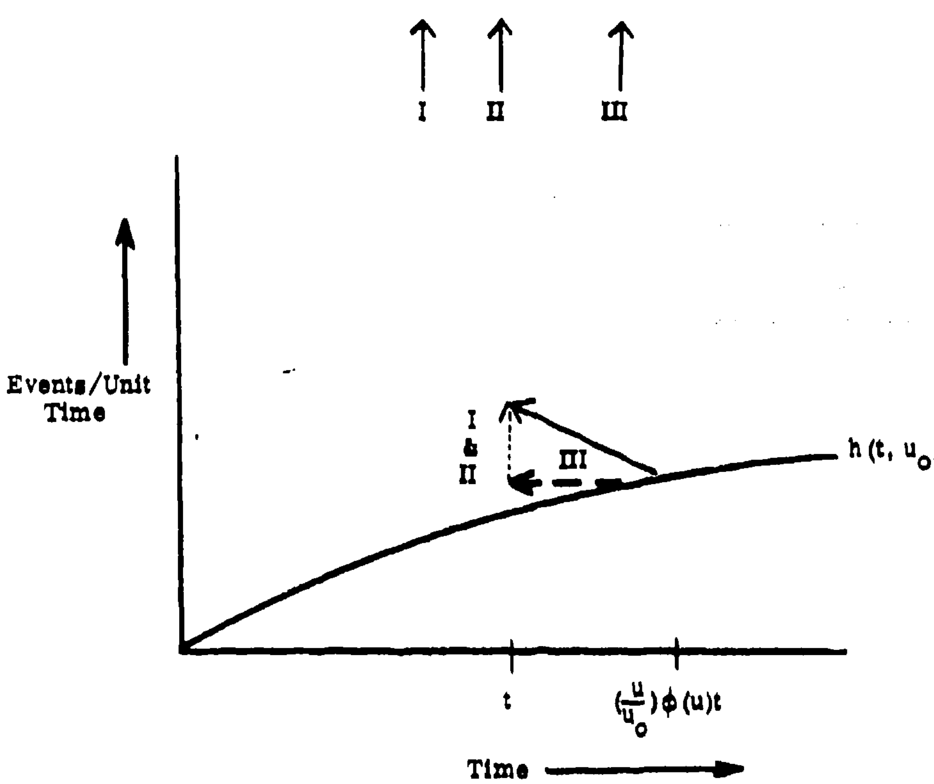


Fig. 14—Now consider an example from the real life of the studied tanks. In the accompanying illustration the calculated renewal density for engine replacements is shown along a time base of tank use at 130 miles per tank per month. At that rate of use engine replacements at 20 months amount to about 4.6 per 100 tanks per month. Now consider what would happen if use were increased 50 percent to 195 miles per tank per month. At 20 months the faster tanks would have the same total mileage as do the slower tanks at 30 months. At that mileage the engine replacement rate per mile is higher than at the lower mileage. At the same time the tanks are going 1.5 times as many miles per month. The net result is that at 20 months tanks going only 1.5 times as fast may be expected to experience over 1.8 times as many replacements.

Next consider the prospect of having operated those same tanks at only one-half of 130 or 65 miles per tank per month. At 20 months the slower tanks have accumulated only as many miles as had the 130-mile-per-month tanks at 10 months. Thus at 20 months the slower tanks experience engine replacement at a much lower rate per mile, and because the slower tanks cover only half as many miles per month, their engine activity is even lower. In fact, at 20 months the reduction of use by one-half results in only 0.17 times as many engine replacements.

The magnitudes of change with respect to time that can be effected by utilization control are obviously greater than many of the differences asserted to exist among different models. It should also be obvious that the utilization impact extends to materiel readiness, assembly repair, assembly floats, maintenance allocations, and so on throughout much of fleet management.

Fig. 15—A nomogram was constructed as an illustration of the utilization effect on major assembly maintenance activity. The nomogram expresses the interdependence among mileage replacement rates, equipment utilization, time replacement rates, durations of repair pipelines, and assembly float requirements.

For example, consider an assembly that is being replaced at a rate of one per 3700 miles (reference the point along the lower left scale). A vertical trace to roughly 130 miles per month reveals that 1000 tanks experience about 38 replacements of that assembly per month. If three months are required between the removal of the assembly to the time that it is repaired and available for re-use, it is necessary to keep just

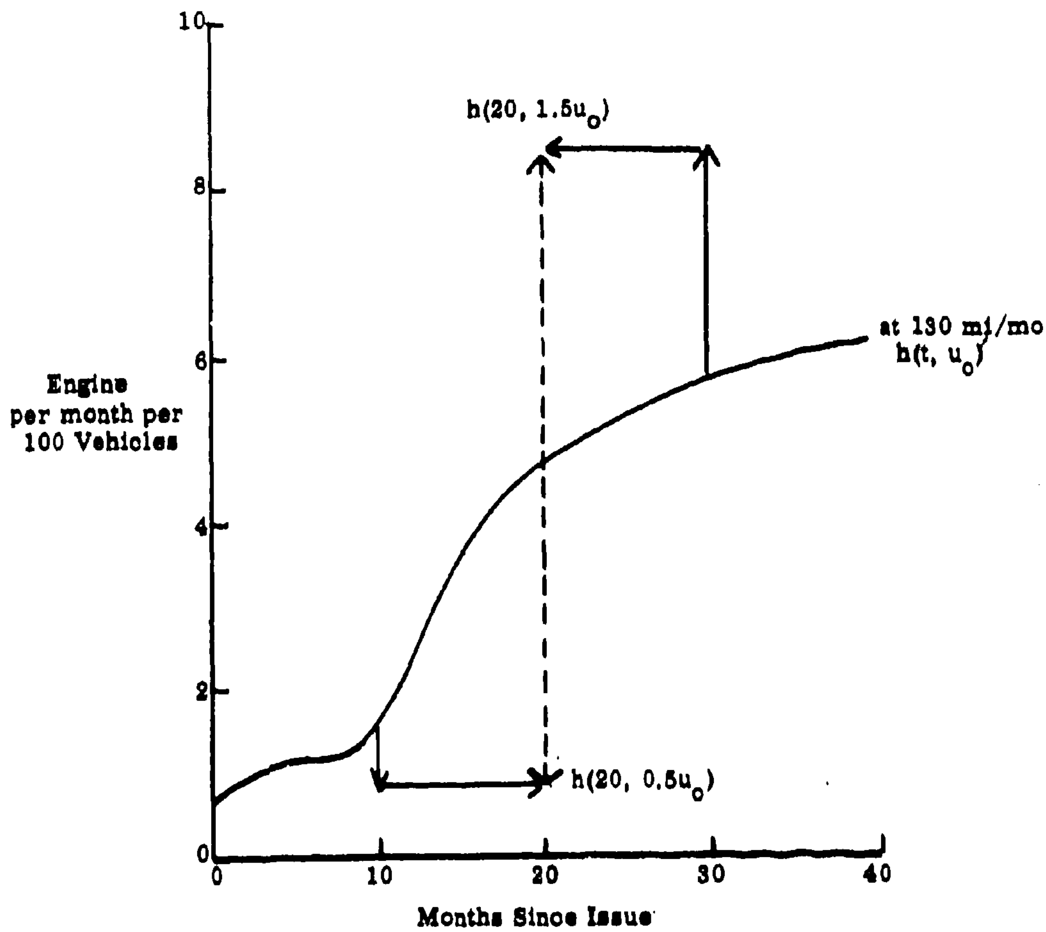


Fig. 14—Direct Effect on Assembly Replacements of 50 Percent Changes in Vehicle Use

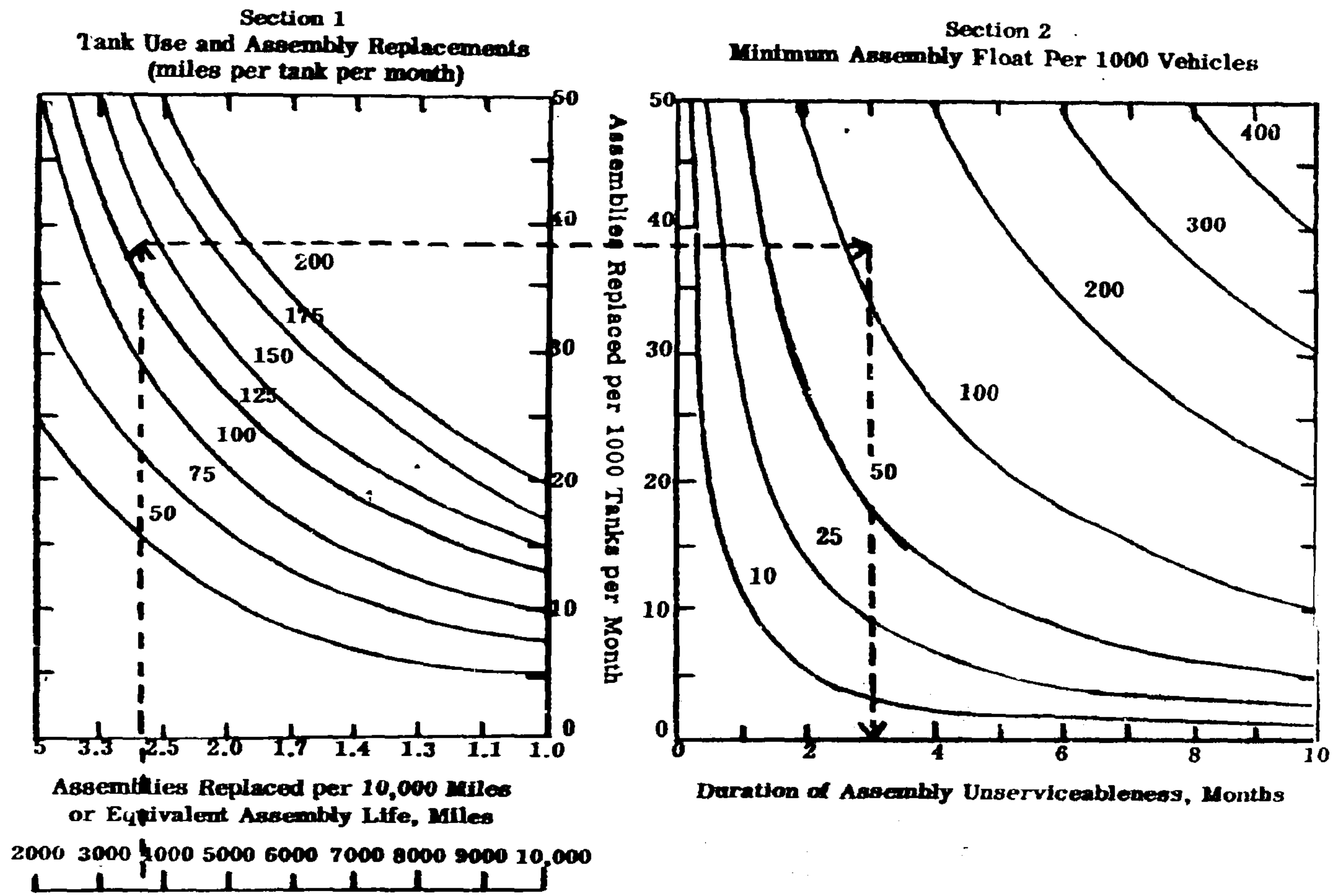


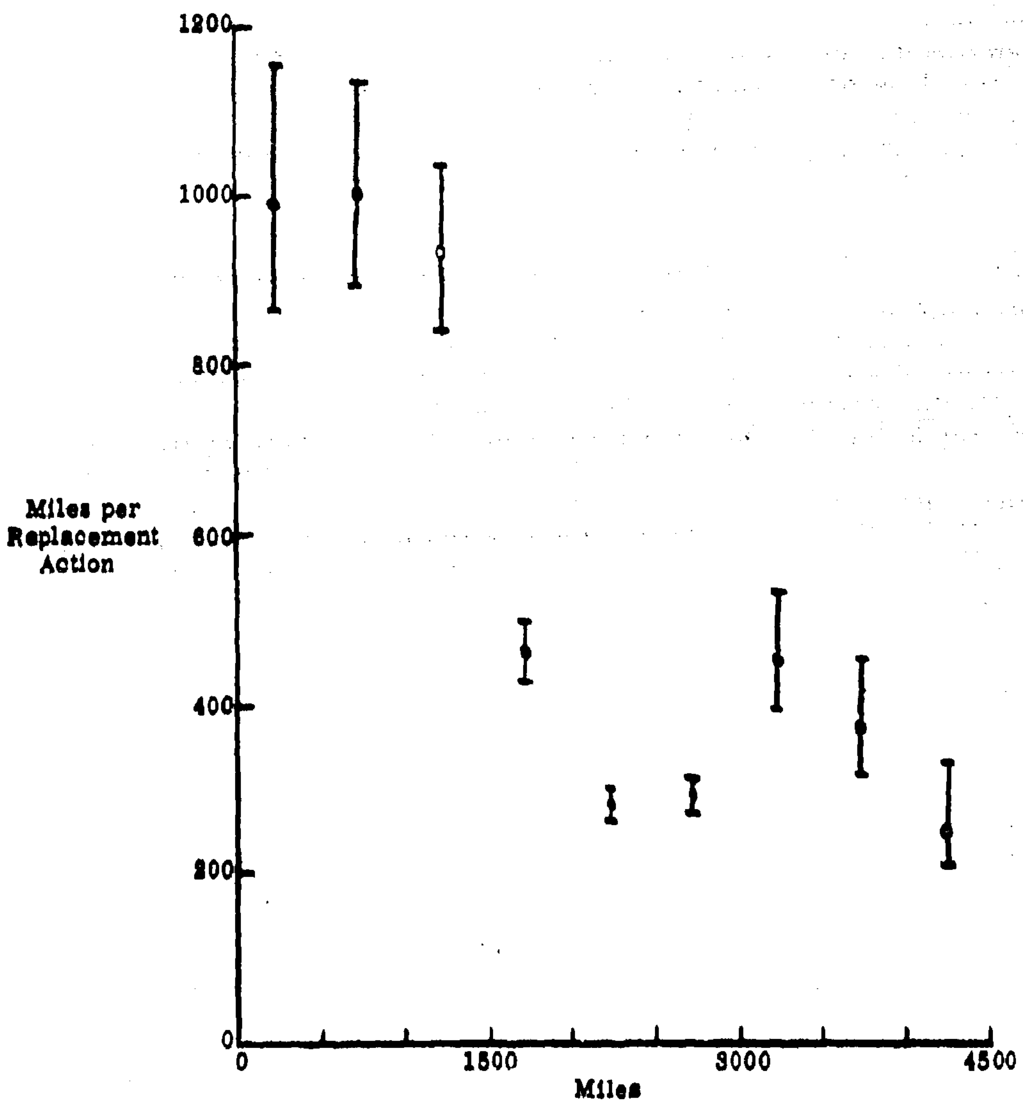
Fig. 15—Relation Among Assembly Life, Rate of Use, Duration of Unserviceableness and Minimum Assembly Float

over 100 of those assemblies in the assembly float or pipeline. Efficient exploitation of combat vehicle resources demands that close, continuous attention be given to all the quantities described in the nomogram. At a given time the fleet generates demands for assembly replacements in a way that depends on both the assembly quality and the fleet use. The duration of the pipeline depends on heavy maintenance programs, stocks of parts, and the geographic location of facilities. A trans-Atlantic pipeline by surface transport can easily run to many months and result in gigantic increases in the required float size.

Fig. 16—The aggregate of mobility-affecting parts replacements provides a basic indicator of what to expect in the way of vehicle performance. Such data were already used to determine the mileage dependent availability potentials at a given rate of use. Perhaps a more fundamental way of viewing the replacement activity is to consider the average miles per parts replacement action over a range of mileages. Such information is provided in the projected figure. To 1500 miles the studied tanks performed with about one replacement action per 1000 miles. Beyond 1500 miles the performance dropped rapidly with the incidence of a great deal of track replacement activity. Improvement occurred in the 3000- to 3500-mile range, but then a decline again appeared. From the detailed basic data it was estimated that the trend would eventually lead to an equilibrium activity close to 165 miles per replacement job for the mobility-affection parts. This level is based on the assumption that all installed parts and components last as well as did the originals.

Fig. 17—The figure now presented provides an example of the effect of using repaired assemblies as replacements in older tanks. Tank A represents the model studied most currently. The tanks down for engine or transmission replacement are shown for Models A and B when all replacement assemblies perform as well as did their originals. Model B was actually studied several years ago, and considerable data were collected for it during periods when it did receive repaired assemblies. Model B was operated at a much lower rate of travel, but translation of its major assembly experiences to the same tank use rate as that of Model A led to the much steeper line. In other words, had the older model tanks been operated at 130 miles per tank per month, at 40 months they would have most likely experienced about a 5-percent deadline rate for engine or transmission replacements assuming enough assemblies would have been available from the repair facilities. Model A was not observed much beyond 20 months

Fig. 16–Tank Miles per Mobility-Affecting Replacement Action



and hence had not had an opportunity to experience use with mostly repaired engines. The similarity between the predicted behavior of both Models A and B with all-new assemblies is cause to suspect that Model A might then do as poorly as Model B when mostly repaired assemblies are provided it. A very likely consequence of the observed trend of performance is that fleet users will probably, quietly reduce their level of tank use to one allowing more relaxed maintenance support.

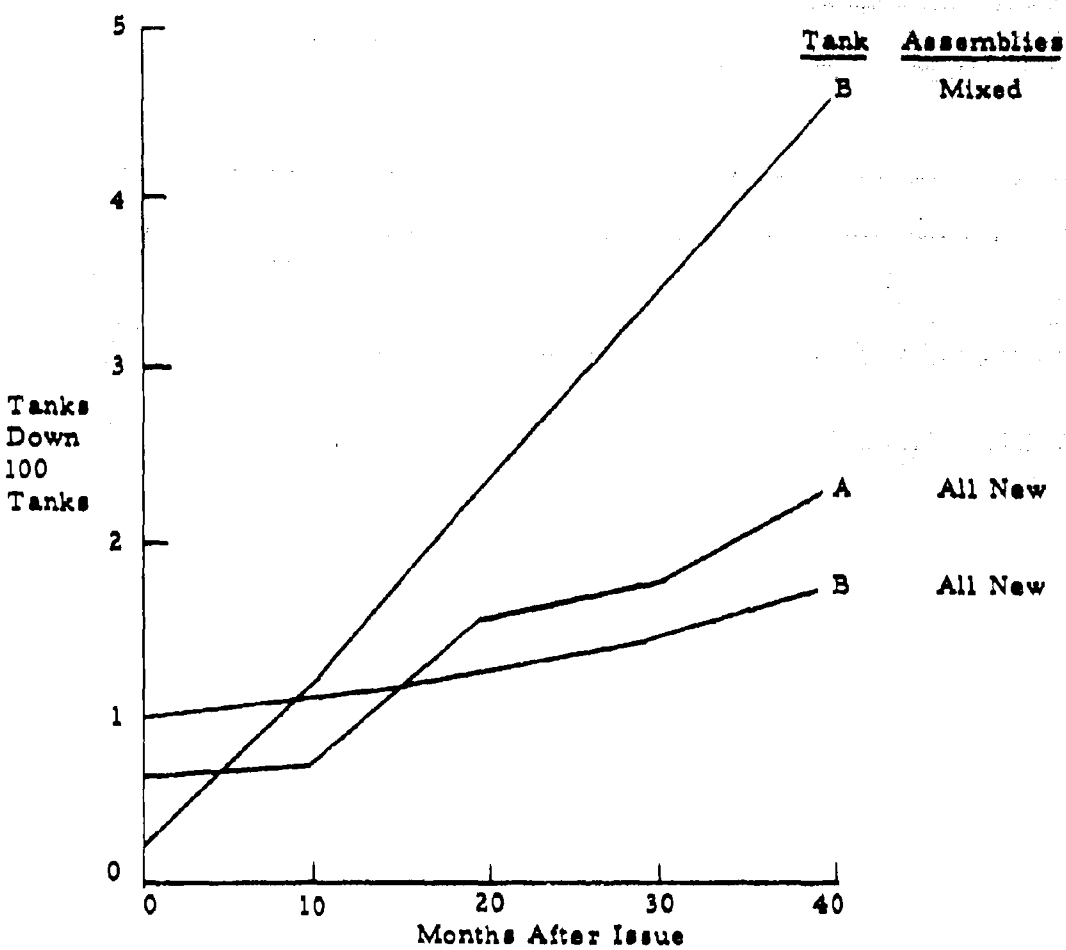


Figure 17--Tanks Down for Engine and/or Transmission Replacement at 130 Miles per Month

Fig. 18—Equipment availability is only a partial measure of materiel readiness for combat. In general availability can always be increased by decreasing the rate of use. Such a phenomenon seems to occur very often. A matter of equal importance to readiness is the performance to be expected from any equipment that is available. Vast segments of the US Army's inventory face a very severe dual requirement. Such equipment is used in extensive peacetime training programs. The equipment has to be available not only for training but also for any emergency deployment to combat. In order to survive if combat should arise, it is necessary that the equipment continues to possess an adequate of residual or combat life. The preceding examples from tank life give sufficient evidence of a probable loss of residual life as the equipment is used and ages. The data of mobility-affecting parts replacement activity were further translated into a measure of what would be expected in the way of tank endurance in the event of an unexpected 50-mile march over hard-surfaced roads. In order to score a success, the tanks subjected to this hypothetical test must be available to start such a march and then complete it without a mobility-affecting deficiency. The march was made short for several reasons. The principal reason was that over several years, the observation of several scheduled marches revealed that tanks experienced most of any deficiences to about 100 miles during the first 50 of those miles. Hence a success to 50 miles is good assurance of success to somewhat more than 100 miles. Too many so-called readiness tests are not previously unannounced. On announced exercises units very often are able to deploy all their vehicles initially. Were a deployment requirement to be issued at some other time the results would be likely to be far different.

For the hypothetical march measure, the mileage-incidence of mobility-affecting replacement actions was assumed to follow the trend developed from the main body of tank history data. In the particular chart shown the training requirement was taken to be about 125 miles per month. Over a period of many years tanks would continue to accumulate mileages, lose availability, and do less well against the unexpected march. The 10,000-mile line is just one rough approximation of what an equilibrium tank might achieve. The march model has been given considerable elaboration with study devoted to the implications of various functional effects for different utilization dependencies. The requirement for uncomplicated visual presentation has led to the reduction of all results to formats very similar to that shown.

Fig. 18—Effects of Age on Hypothetical Performance Tanks on a 50-Mile March

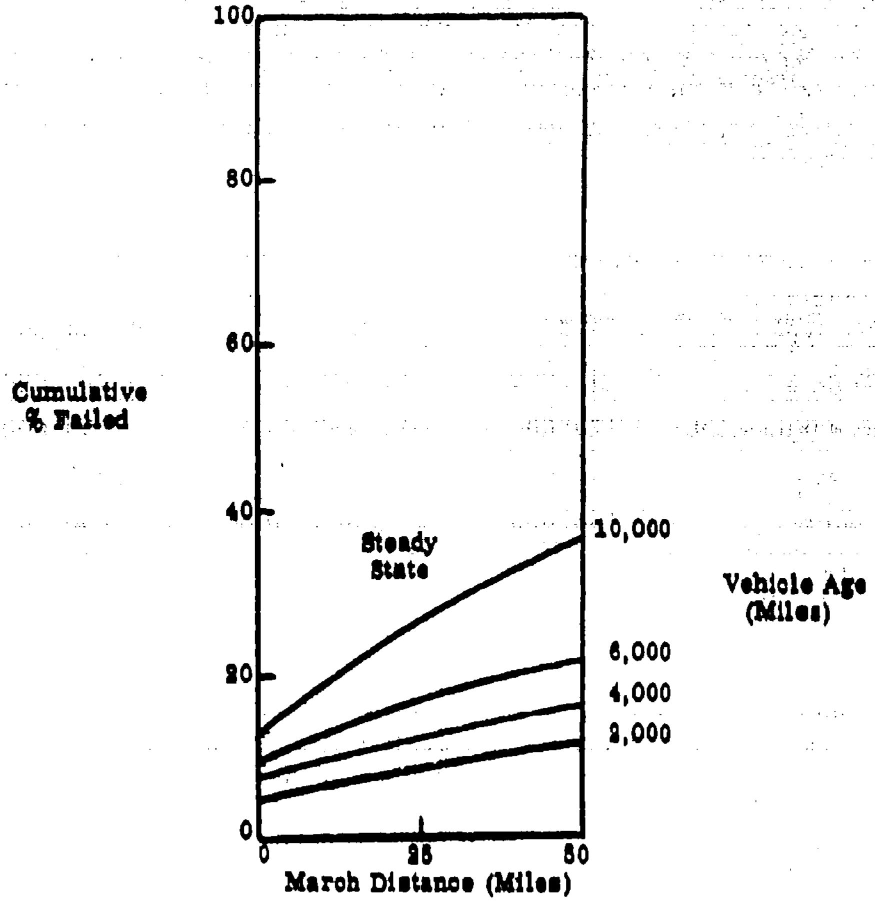


Fig. 19— As part of the conclusion of this general survey of RAC's management assisting activities in the area of combat vehicles, it is appropriate to reintroduce consideration of parts support costs. At 125 miles per tank per month the yearly support of a fleet of 1000 tanks of the model studied requires the funds shown in the accompanying table. During the first year the fleet consumes about $1.5 million worth of parts and assemblies. By the third year over $10 million worth are being consumed. Track replacement by far comprises the biggest single slice of the total bill. In the third year the "other parts" account for only about 16 percent of the total cost, but they involve a great variety of different kinds and numbers of repair parts.

ANNUAL EXPENSE--1000 TANKS

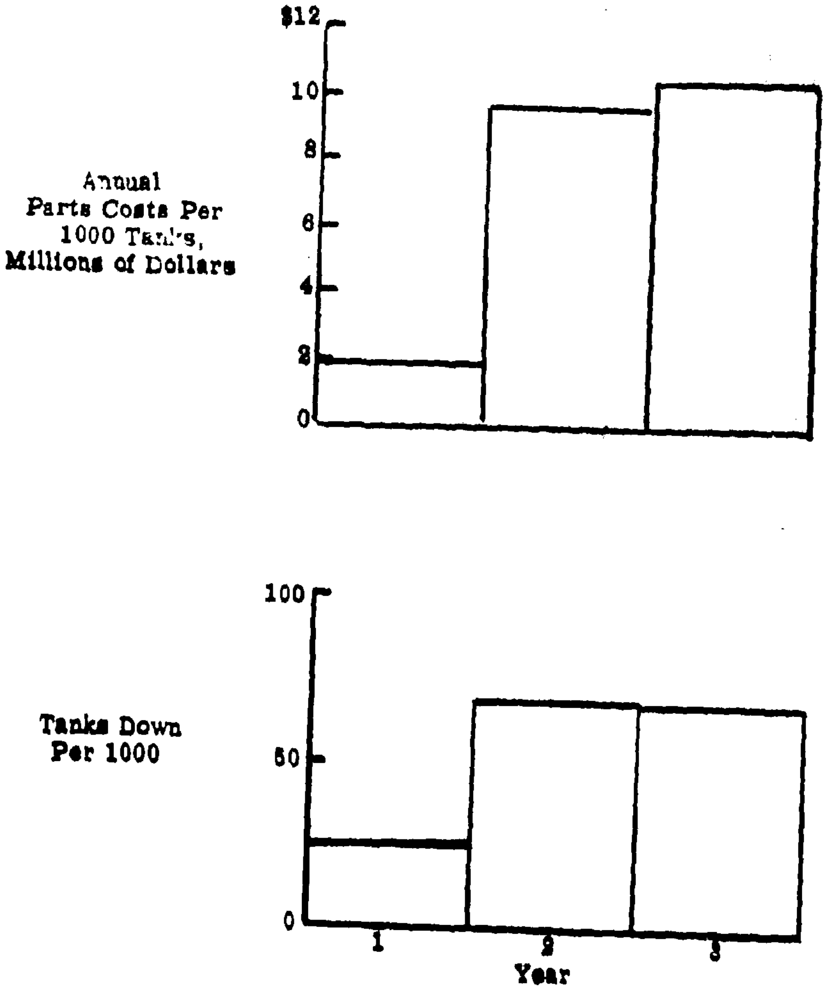
Year Mileage	1 0-1500	2 1500-3000	3 3000-4500
Engine	$ 740,000	$2,540,000	$ 3,000,000
Transmission	170,000	430,000	830,000
Track	500,000	5,130,000	4,900,000
Other	460,000	1,400,000	1,600,000
Sum	$1,870,000	$9,500,000	$10,330,000

Figure 19

Fig. 20—Money is not the only penalty of vehicular old age. Even though parts are fed into the tank fleet, the net availability of tanks drop. In reality tank users have to pay more and get less as their vehicles accumulate mileage. In this chart both the parts costs and unavailability of tanks like those studied are shown. Out of a force of 1000 tanks, the equivalent of more than an entire battalion are on the average unserviceable and unavailable during the second and third years. Compared with the first year, parts costs have increased about five-fold and unavailability about three-fold be the second and third years.

Fig. 21—Practically all the preceding results may be categorized as by-products of the general analysis leading to the determination of target ages for the effective in-use lives of vehicles. Through suitable weighted

Fig. 20—THE DOUBLE PENALTY—
MONEY and DOWNTIME



The effective in-use lives of tanks and tank mobility systems at different rates of uniform use

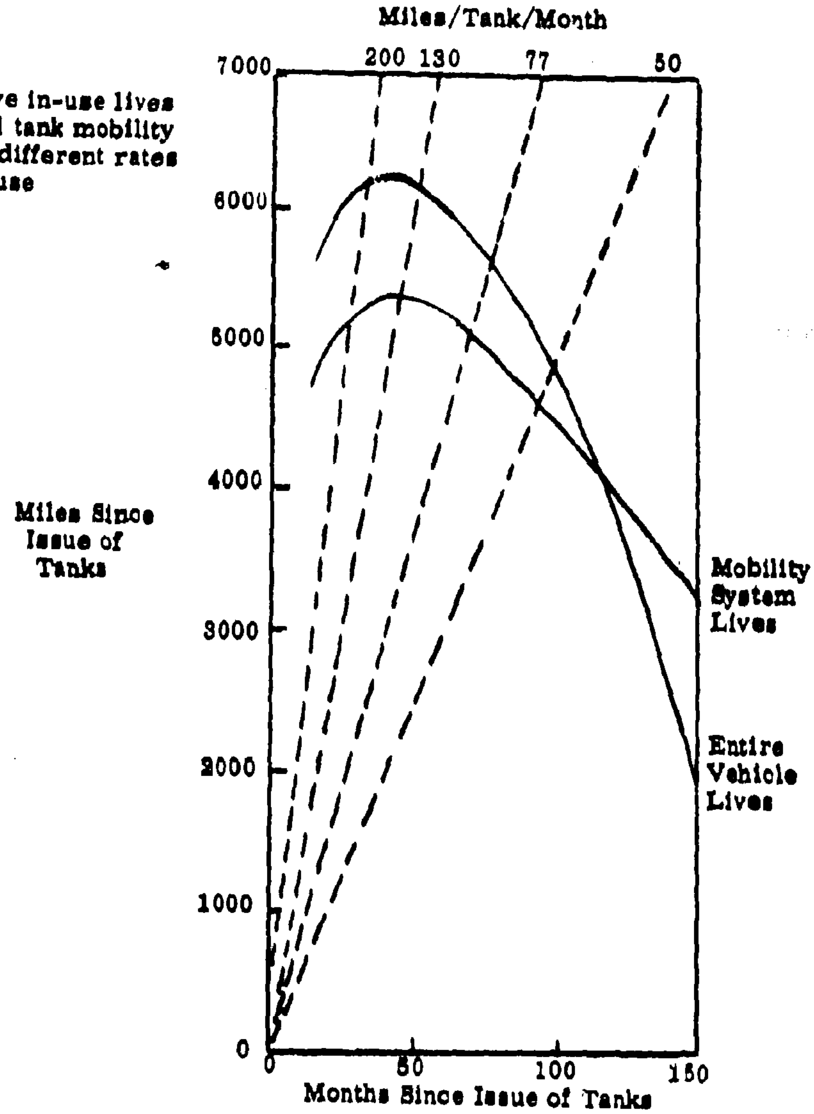


Fig. 21

combination of the foregoing kinds of information schedules of lives were determined for the tank model studied. These lives are best represented by a curve on a mileage-calendar age plane. In general effective in-use life depends on rate of utilization. In fact the presented curve applies only to an operation spectrum of uniform utilization. The general accumulation of mileage may be represented by a path belonging to a parametered family of usages. Each use family will result in a different life curve. As long as the mileage paths of a given use family do not intersect one another, the time mileage plane consists of points associable with at most a single utilization curve, and a single effective life curve can be constructed uniquely. Existing history data do not provide an adequate basis for the construction of a realistic life model employing intersecting mileage paths of single families and leading to life differences at the points of intersection. Such questions are interesting from the programming and function-theoretic point of view, but they remain well beyond current capacities for empirical, experimental resolution.

Two separate effective in-use life curves are shown in the figure. The one applies to entire tanks and the other to separately defined tank mobility systems. In general higher utilization rates may be expected to increase mileage lives except at very high use. However, the higher mileages are achieved in much shorter times. Much of the management problem arises because the training program results in large numbers of old model tanks with relatively low mileages and in equally large numbers of newer model tanks with much higher mileages. It becomes necessary to establish tank life paths through the inventory in such a way that in the long run the less obsolescent tanks are also the ones with the lower mileages. To have obsolete, low-mileage tanks and modern, over-used ones at the same time achieves nothing more than a use-, readiness-, budget-paradox.

RAC has presented a rapid survey of many of the factors considered in a well-integrated program of assistance to the US Army in the management of its combat vehicles. Time and space do not permit treatment of all factors, nor do they suffice for adequate explanation of the integration to final result.

RAC's activities in this area represent a continuing sequence of alternating empirical and theoretical efforts. Over the years the guiding doctrine has been to provide a steady output of information of general and specific utility to the Army in the management of its combat vehicles. The information has had to be consistent with the materiel experience of US Army troops in combat units undergoing actual training and yet remaining constantly responsible for preserving a combat ready posture. Applicability to high population fleets and not elegance in an artifically reduced inventory has been a test to be satisfied for all resulting conclusions and recommendations.

APPLICATION OF STATISTICS TO EVALUATE SWIVEL HOOK-TYPE CROSS CHAIN FASTENERS FOR MILITARY APPLICATIONS OF TIRE CHAINS

Otto H. Pfeiffer
U. S. Army Tank-Automotive Center
Warren, Michigan

ABSTRACT. The test was conducted according to a developed experimental design to determine the utility of swivel-hook cross-chain fasteners for military applications of tire chains.

Dual tire chain assemblies (mounted on M35A1 test vehicle) consisting of standard Military-type side and cross chains and three types (standard and two swivels) of cross chain fasteners were subjected to an accelerated wear test of 425 miles on dry concrete road surfaces. The experimental results were expressed in terms of: (1) Miles to failure for an individual cross chain, (2) Weight losses of selected cross chains and (3) Replacement times for each of three types of fasteners. The principal response, (1) miles to failure, was considered to be exponentially distributed; therefore, logarithms of miles to failure were analyzed in accordance with the structure of the experimental design.

Cross chains connected with swivel-type fasteners remained functionable about twice as long as the cross chains connected with the standard-type fasteners. Both swivel-type fasteners permitted significantly faster cross-chain replacement than the standard type, although one swivel type was also significantly faster to manipulate than the other swivel type.

The statistical analysis of the experimental results indicated the swivel hook-type cross chain fasteners used in this test resulted in a significant increase of cross chain life as well as simplification of replacement.

INTRODUCTION. It became necessary for the government to make a decision whether to consider swivel hook-type cross chain fasteners in future procurement of tire chains and components. Some information on swivel hooks was available, but it was considered inadequate for the basis of a decision.

A test was proposed to utilize two military trucks equipped with various types of tire chains, to be conducted on both concrete and gravel road surfaces (in a two-to-one proportion) for a total distance of approximately 300 miles.

Products of two manufacturers of tire chains and chain components were available for the test. Each had a swivel hook of a particular design which they were interested in selling to the Government.

Since a large proportion of the Military's wheeled cargo vehicles fall within the $2\frac{1}{2}$ ton payload class (see Figure 1) having 9.00 x 20 size tires, it was desirable to test tire chains of this dimension.

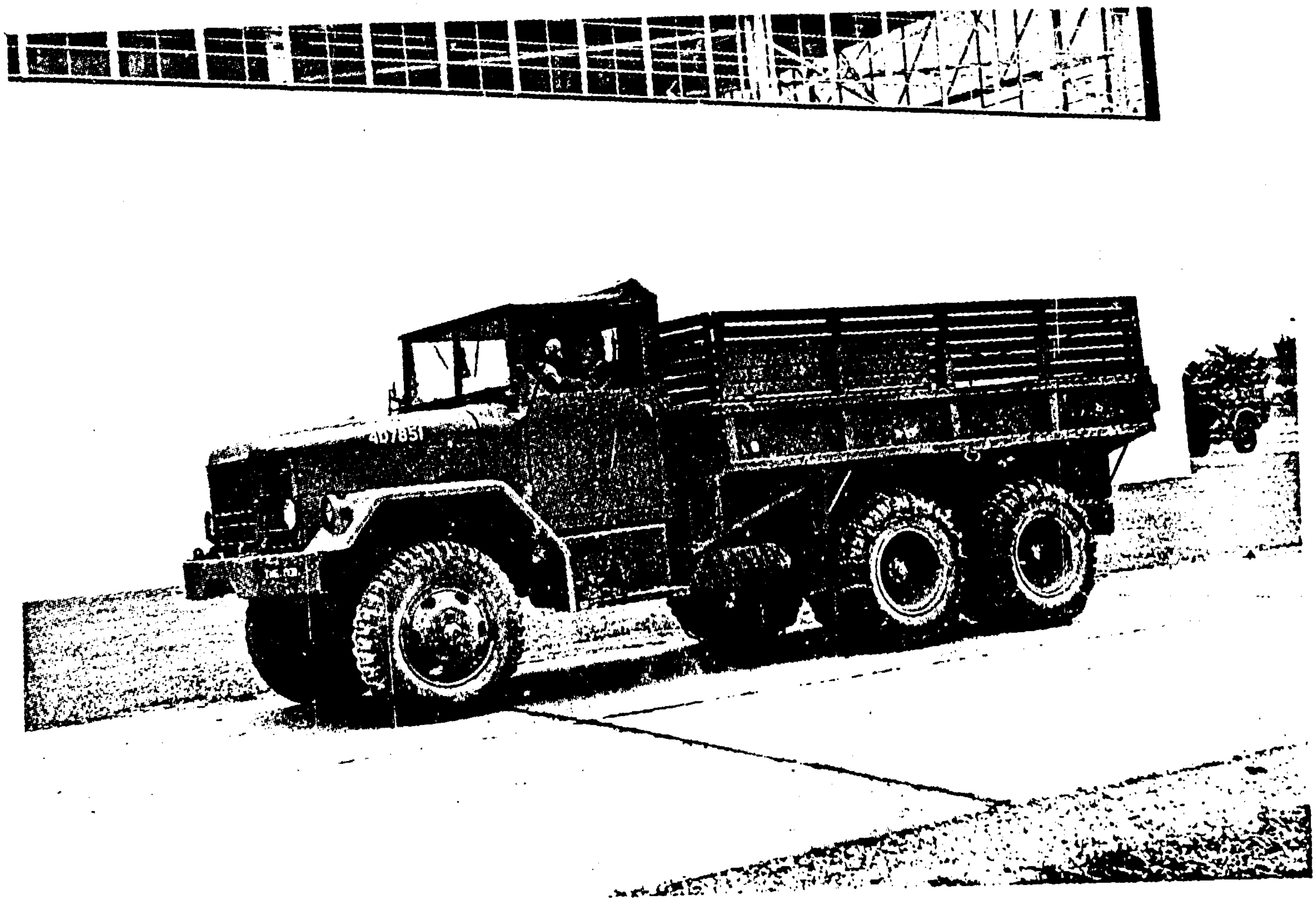
It was requested that representative cross chain samples of all test chain assemblies from the two manufacturers hereafter referred to as "Code B" and "Code C" and a standard Military item referred to as "Code A" be subjected to a metallurgical examination to determine: mechanical properties, macro-etch quality, case-hardened depth, and cross-sectional hardness.

Two complete tire chain assemblies of Code B were to be placed on two rear dual wheels of one vehicle, diagonally opposite to two Code A chain assemblies. The same arrangement utilizing Code C and Code A was to be adhered to on a second test vehicle.

During the course of testing, it was proposed that two brake panic-stops per mile be made after the vehicle had attained maximum speed. Wheel spining on take-off was also requested.

A review disclosed a test [3] had previously been conducted on tire chains by the Government. This test compared the Code B and Code A chain only. A similar chain arrangement to this proposed test was used, but the vehicle was driven 500 miles over various terrains and on surfaces ranging from marsh and swamp to concrete. Although this test indicated some advantages of the swivel hooks in comparison with the standard Military chain, the test did not adequately establish the quantitative nature of the advantages.

Only letters of recommendation from commercial sources were available regarding the Code C chain.



U. S. ARMY TANK-AUTOMOTIVE CENTER NEG. NO. 72275 DATE 11 June 63
Tire Chain Test. M35A1, 2 1/2 Ton Truck, Cargo, 6x6. Left Front View of Truck with Chains Installed.

FIGURE 1

The similarity between the newly proposed test and the test previously conducted brought up the question: What could be learned with this test that had not been found out before? The only answer that appeared obvious was -- Nothing! Further study of the situation anticipated many problems if the proposed test arrangement and procedure were to be followed:

1. Differences in the vehicles, not only in weight but manufacturing variances as tire sizes, tracking characteristics, and braking characteristics, to mention a few.

2. Differences in the tire chains, both length of cross chains as well as the cross-chain wire diameter and of major concern -- the difference in metallurgy of the cross chain steel.

3. The difference in road contact surfaces when chains are compared on different wheels, or worse -- on different vehicles.

4. Difficulty of data analysis or supportable conclusions if the test were to be conducted in a haphazard manner or without accounting for known variables.

We decided a specially-designed plan must be developed which would produce usable data. The services of Dr. Emil H. Jebe, a Research Mathematician from the University of Michigan's Institute of Science and Technology, were obtained. He became an inseparable part of the project until final conclusions were reached.

<u>CHOICE OF RESPONSE AND EXPERIMENTAL UNIT</u>. It was apparent that several responses should be studied. Of primary interest were:

1. Miles to failure of each individual cross chain,

2. The weight loss of the cross chains associated with each type of cross chain fastener,

3. The time needed for replacement of worn-out or broken cross chains.

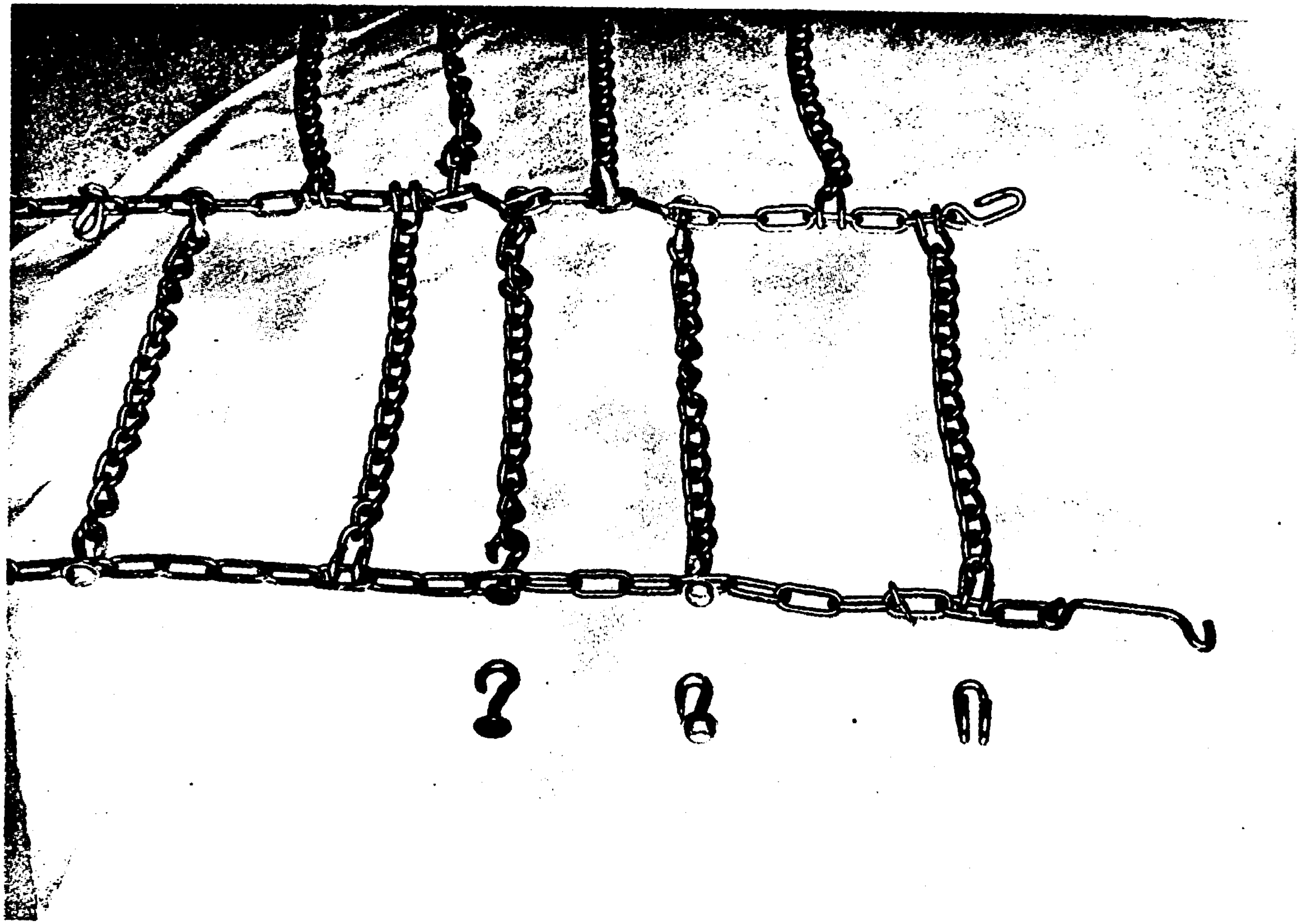
The first major problem in development of a suitable experimental design of plan was to determine the unit for measuring responses. The four rear dual wheels and the inside and outside tire of each dual wheel were the first kinds of units considered. However, since a single tire chain covers the complete dual wheel (resulting in only four units being available at the same time), no satisfactory design could be based on a wheel as the unit unless 1200 to 2000 miles could be driven. Even using a single tire as the unit would have given only eight units and differences between outside and inside tires would have to be eliminated. This kind of unit would also require an extended period of driving.

Further discussions disclosed that the government was considering at this time only the utilization of swivel hooks as repair items for the ample supply of standard Code A chains presently in stock.

An interesting thought occurred -- why not use standard Code A tire chains with Code B and Code C swivel-hook fasteners inserted as if they were repair items? This would eliminate several of the anticipated problems. The cross chains of Code A or standard military tire chains were considered to be for all practical purposes of uniform size and metallurgical composition.

It suddenly became evident that with this concept, a large number of experimental units would become available if we considered each individual cross chain as the unit of measurement. A dual tire chain consists of 26 cross chains or 13 on each half. A total of 104 units became available which could be used for one run of, say -- 300 to 500 miles.

Some minor problems were encountered in the acceptance of this experimental unit. The swivel hook-type fastener could not be inserted into the same link of the side chain as the standard crimp hook. It required an adjacent link 90° out of phase. This problem was solved by spacing the cross chains unevenly around the tire as shown in Figure 2. Since three types of cross chain fasteners were to be tested, four cross chains were fastened with each type. The remaining cross chain was left fastened with the standard type and was not used for test data. Using this arrangement, there were 32 cross chains for each fastener-type, evenly distributed over the eight tires of the four dual wheels.



U.S. ARMY TANK-AUTOMOTIVE COMMAND NEG. NO. 71825 DATE 1 Apr 63
Modified standard military dual tire chain utilizing Code A, Code B, and Code C cross chain fasteners

FIGURE 2

A completely randomized arrangement of the cross chain fasteners on each wheel seemed undesirable. A randomized starting order followed by a systematic order was suggested. The cluster of four cross chains connected with the same type fastener on each tire became the experimental unit with eight replications over the set of wheels.

Four chain assemblies were fabricated according to prescribed procedure and mounted on the test vehicle. To accelerate the rate of wear and thereby reducing driving distance, a test course (see Figure 3) consisting entirely of concrete was selected. Provisions were made to equalize right and left turns necessary during test driving.

STRUCTURE OF THE DESIGN. With the experimental unit now clearly defined, it was possible to develop the complete overall plan for data analysis. There were certain obvious sources of environmental variation present which could not only be removed but estimated for magnitude. These sources are listed as:

1. Difference between front and rear dual wheels.

2. Difference between right and left side dual wheels.

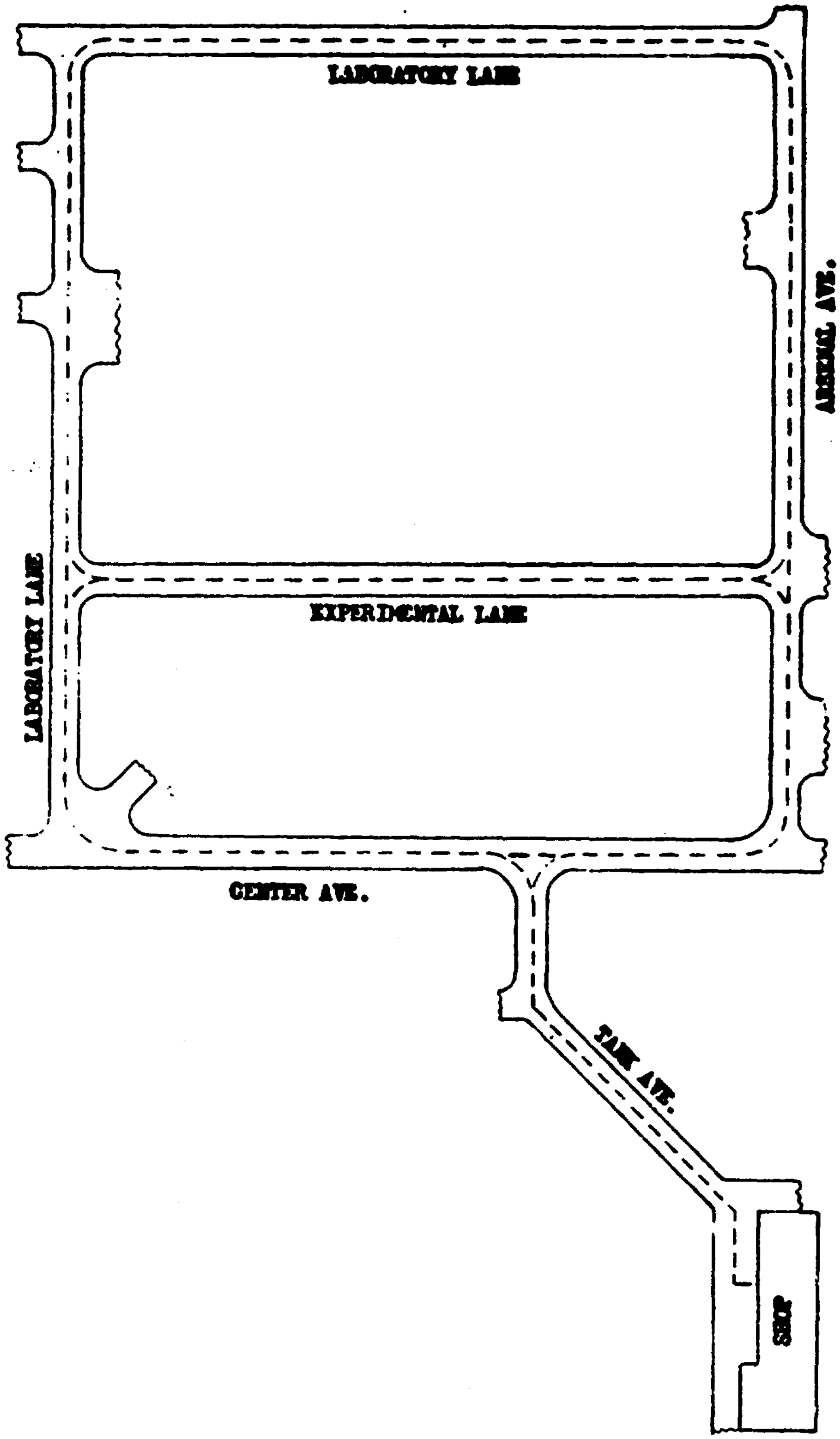
3. Outside versus inside tires.

4. Interactions of these effects with each other.

5. Possible interactions of the treatments (types of cross chain fasteners) with these positional differences.

Since there were three clusters - three experimental units - of four chains (each cluster with different types of fasteners) on each tire the plan may be described as a Randomized Complete Block Design in eight replicates considering each tire as a block. The variation among blocks was also to be subdivided in the manner just outlined.

A formal structure for the design could then be established. The usual textbook model for a Randomized Complete Design would be satisfactory for preparing an analysis of variance of these observed results, providing the usual assumptions could be made. One of these



TIRE CHAIN WEAR TEST COURSE

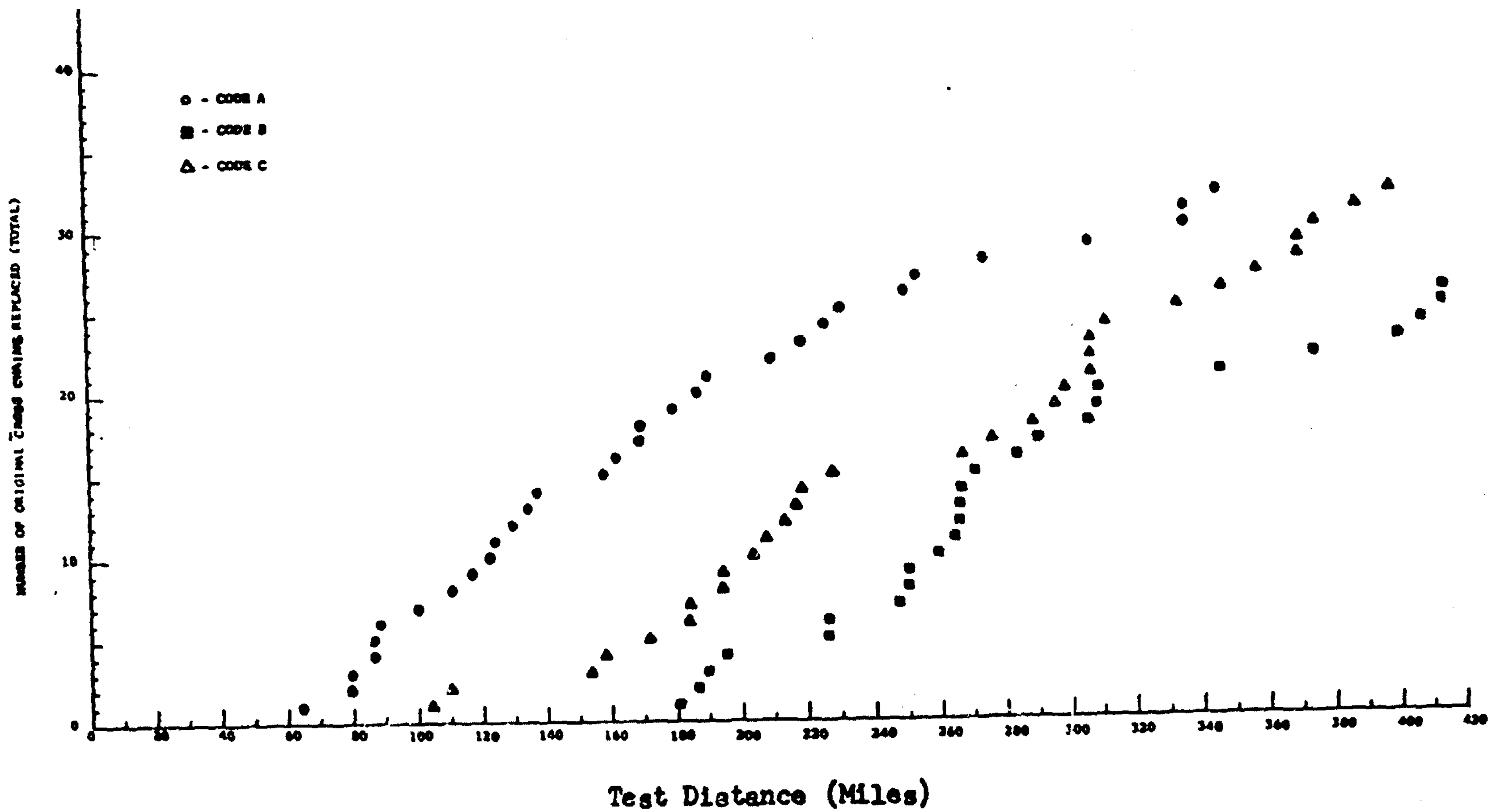
FIGURE 3

assumptions is that the observations are independently and normally distributed [1] [10]. The response of primary concern here is miles to failure of an individual cross chain. Therefore, the test plan may well be considered a "life test" or a "wear test". Considering that our observations were "miles to failure", we realized they would not likely be well-described by the normal probability distribution. This point is well demonstrated in Figure 4. It appears therefore that the exponential distribution may be regarded as an acceptable probability model for the observed miles to failure. With this in mind, our analyses of these failure data have been generally guided by the considerations set forth in a series of papers by B. Epstein, B. Epstein and M. Sobel, and M. Zelen which appeared in several Mathematical and Statistic Journals (See reference list).

The exponential distribution in its probability density form usually expresses the random variable in some quantity directly related to time. In our case, d = miles-to-failure may be used as the random variable. In this form, a constant uniform failure rate is assumed. As was indicated previously, we do not have a uniform situation in this tire chain test since there are a number of sources of variation present. The parameter θ appearing in the probability density form represents the mean time before failures of the cross chains. Based on Zelen's work [19], a more complex model (Figure 5) was written for this θ in the exponential distribution.

Another view may be taken for observations following exponential distribution. The procedure established here suggests taking logarithms of the observations, in this case miles-to-failure for each cross chain, and then carrying out an analysis of variance considering necessary assumptions. The functional relation between mean and variance for exponential distribution, that is, the standard deviation equals the mean [2] suggests the use of the logarithmic transformation (see Figure 6). This approach is also discussed by Zelen in his paper [19]. He finds the technique acceptable against possible departures from the strict exponential distribution form.

The analysis of variance of the logarithms was prepared on the basis of the model and will be discussed later.



MILEAGE AT FAILURE OF ORIGINAL CROSS CHAINS

FIGURE 4

EXPONENTIAL DISTRIBUTION MODEL

The model $\theta_{ijkr} = \theta\alpha_i \gamma_j \tau_k \rho_r \eta_{ij}$ was developed from works by M. Zelen. This model states the Mean Time Before Failure (MTBF) for the four crossbars secured with the same type of fastener on the same tire depends upon

- θ — a general mean time to failure
- α_i — an effect associated with the front or rear set of dual wheels, i=1, for Front, and i=2, for Rear
- γ_j — an effect associated with left or right side of truck, i.e., j=1 for Left, j=2 for Right
- τ_k — an effect associated with the tire position, k=1 for Outside, and k=2 for Inside
- ρ_r — an effect associated with the type of cross-chain fastener, say 0_A, 0_B and 0_C, or the subscript r takes three values
- η_{ij} — an interaction effect associated with the dual wheel position

FIGURE 5

LOGARITHMIC TRANSFORMATION MODEL FROM EXPONENTIAL DISTRIBUTION FORM

Let $Y = \log d$ where d is miles to failure.

The transform model is written as:

$$Y_{ijkrs} = m + a_i + c_j + (ac)_{ij} + t_{ijk} + f_r + e_{ijkrs}$$

where

m	is an overall mean logarithm
a_i	is the front or rear effect
c_j	is the left or right side effect
$(ac)_{ij}$	is an interaction or component associated with i, jth wheel position
t_{ijk}	is the tire position effect within the i, jth wheel
e_{ijkrs}	is a random deviation for the sth individual cross-bar within the rth type and its associated tire and wheel position

FIGURE 6

The test was originally planned to run for 300 to 500 miles. At 300 miles, the test was temporarily stopped to assess the situation up to that time. It was already clear that there were large differences in miles to failure for types B and C versus type A even though approximately 1/3 of the original cross chains had not yet failed. This large number of unfailed chains would have created considerable difficulty when analyzing the results. Driving was continued to about 425 miles before being terminated for other reasons. At that time, six of the original cross chains remained. These were all equipped with type B fasteners and were positioned as follows: One on each of two wheels, and four on one wheel, as shown in Table I. It was necessary to estimate data values for these six in order to maintain a balanced and simple, straight-forward analysis of the results. In general, procedures derived from work by B. Epstein [8] were used for estimating the missing data. Using Epstein's formula (see Figure 7) we obtained estimates for the missing values on the two single tires. In the first case, $n = 4$, $r = 3$, $d_r = 309.8$ and $d_1 = 226.6$. Solving for $\hat{\theta}$ we obtained the value of 123.25. This formula assumes that failures occur randomly at any time starting from zero miles of travel. Since failures did not occur for some distance of travel, we estimated the minimum "guarantee" distance A by the formula $\hat{A} = d_1 - \hat{\theta}/_n$. In this case, $\hat{A} = 195.79$. Combining $\hat{A}$ and $\hat{\theta}$, we obtained 319.04 as the proper estimated MTBF* for the cell. Using the estimated cell mean, we estimated the cell total as n (estimated MTBF) = 4(319.04) = 1276.16. We already knew the actual miles to failure of three of the cross chains in the cell; therefore, subtracting this value from the estimated total left 433.1 as the estimated missing value.

The same method was used to estimate the second missing value at 504.3.

No failures of B type swivel hook fasteners were evidenced on the outside tire of the left front wheel. This situation posed a real problem. The Epstein formula used for estimating the previous two missing values requires at least two failures in a cell if it were to be applied directly. A variety of methods for solving this problem were considered, including schemes based on using the weights of the unfailed cross chains at termination of the test driving. All these schemes were rejected as unsuitable.

*Mean Time Before Failure

	LEFT			RIGHT		
	A	B	C	A	B	C
	MILES	MILES	MILES	MILES	MILES	MILES
Front Outside	336.4	-	289.2	336.4	414.6	312.4
	275.0	-	153.9	180.1	400.8	347.7
	250.6	-	371.3	86.8	414.8	194.7
	346.8	-	334.3	64.7	346.8	208.3
Front Inside	254.1	309.8	158.3	219.3	266.1	269.7
	169.9	-	184.1	122.4	271.9	307.7
	158.3	306.7	358.4	129.8	247.1	399.6
	88.5	226.6	307.7	110.6	186.5	171.8
Rear Outside	231.5	250.6	183.9	307.7	266.8	110.1
	86.8	195.6	213.9	162.6	308.6	376.1
	134.5	264.8	388.8	116.9	-	204.1
	187.5	309.2	217.2	79.6	375.1	228.6
Rear Inside	211.5	259.8	371.0	190.8	250.6	296.7
	169.7	180.6	104.3	79.6	267.0	194.4
	100.0	189.9	276.9	124.0	291.0	307.7
	137.7	284.6	299.6	226.3	226.5	219.0

MILES TO FAILURE OF ORIGINAL CROSS CHAINS

TABLE I

EPSTEIN'S FORMULA

$$\hat{\theta} = \frac{\sum d_i + (u-r)d_r - nd_1}{r-1}$$

where (in this particular case)

$\hat{\theta}$ = estimated MTBF

d_i = observed miles to failure for ith failed item

n = number of items under test

r = number of failed items

d_r = the largest value or last failure

d_1 = the smallest value or first failure

FIGURE 7

In order to obtain a useable solution, the Epstein formula was reapplied to the whole of Type B fastener data. Utilizing the two previously estimated missing values, the parameters now became n = 32 and r = 28 in this instance. The mathematics will not be described here. But stating briefly, $\hat{\theta}$ and $\hat{A}$ were estimated, then $32(\hat{A} + \hat{\theta}) - \sum_{i=1}^{28} d_i$ yielded an estimated total for the entire missing cell. This estimated total /4 provided the estimated MTBF or $\hat{\theta}$ for the cell. Individual values were then determined by proportionality of each cross chain weight to mean weight of the cell. The estimated missing values for this cell ranged between 563 and 572 miles. These values appeared to be too high, giving the impression we were favoring type B. Applying the standard analysis of variance "missing plot procedure" for the Randomized Complete Block Design to the logarithms of the miles to failure data [14] [18] provided a mean value for the entire cell although it was not a useable value. The estimated value based on averaging the data available was 276 miles, but it was known these cross chains had already traveled over 400 miles. The distance traveled at termination of test, 424.9 miles, could also have been assigned to each unfailed cross chain but this would have been unfavorable to type B. Therefore, values estimated as already described were used and they appeared to yield an acceptable solution.

There are several approaches which may be followed in considering the estimation of the effects of interest in this experiment. For completeness, three methods were considered and the differences among the methods were small for this test program. The methods considered were:

1. Calculating the appropriate simple averages of the miles to failure data

2. Estimating the parameters in Zelen's model as described above

3. Estimating in terms of averages of the logarithms of miles to failure.

The latter of the three methods was used in the analysis of variance and will be discussed further.

Logarithms of the original data, miles to failure, and the anti-logs of the mean logarithms are shown in Table II.

Averages of Logarithms of Miles to Failure
and the Anti-logs of these Averages

Effect	Tire Position: Inside Log Miles	Inside Miles	Outside Log Miles	Outside Miles	Combined Log Miles	Average Miles
Fastener Type:						
A	2.1682	147.3	2.2381	173.0	2.2032	159.7
B	2.4088	256.3	2.5744	375.3	2.4916	310.2
C	2.3974	249.7	2.3871	243.8	2.3923	246.8
Wheel Position:						
Left Front	2.3549	226.4	2.5554	359.3	2.4551	285.1
Left Rear	2.3004	199.7	2.3191	208.5	2.3098	204.1
Right Front	2.3220	209.9	2.3795	239.6	2.3508	224.3
Right Rear	2.3220	209.9	2.3455	221.6	2.3337	215.6
Sides:						
Left	2.3276	212.6	2.4372	273.7	2.3824	241.2
Right	2.3220	209.9	2.3625	230.4	2.3423	219.9
Wheels:						
Front	2.3385	218.0	2.4673	293.3	2.4029	252.9
Rear	2.3112	204.7	2.3323	214.9	2.3218	209.8
Overall Mean:	2.3248	211.2	2.3999	251.1	2.3623	230.3

TABLE II

Comparing the average for type A fastener (inside tires only, expressed in anti-log form as 147.3 miles with values calculated by the other two methods -- 155.8 miles and 155.9 miles, respectively) we find it slightly less. This somewhat lower figure is the result of the non-linearity of the logarithmic transformation. This anti-log value is really an estimator of a median value rather than a mean.

Estimating the differences among the types of cross chain fasteners, or the ratios as, say -- B/A, C/A and B/C was the next concern. It not a simple problem and considerable study was devoted to finding a reasonable solution. A method discussed by Zelen [19] [20] for estimating such ratios and finding confidence limits for the ratio gave extremely wide limits for these ratios from our experimental data. From the averages of "log miles" for "Insides Tires", "Outside Tires" and "Combined Averages" presented in a previous table, the fastener-type differences expressed in logarithms were calculated. The variances of these differences were estimated directly by taking 2(Experimental Error Mean Square) /r, where r is the number of values averaged to form a mean for a single fastener type. This was based on the theorem that the variance of a difference is the sum of the variances of the quantities used to form the difference. The standard deviation was obtained from the variance result just stated. Confidence intervals were then formed by taking the observed mean differences: $(B-A) \pm t_{\alpha k}$ (standard error) where t in this case was $t_{(0.95, 86)} = 1.987$. The k = 86 degrees of freedom comes from the Pooled Error Mean Square determined from the analysis of variance presented later. The confidence intervals obtained are for differences of averages expressed in logarithms.

There was a considerable difference in the average miles to failure (about 50 miles) considering all the fasteners combined between the Inside Tires and the Outside Tires. Considering this difference, it was decided to present separate results for Inside and Outside Tires and then combined averages, as seen in Tables II and III. Returning results to original scale of miles to failure was desirable. It was observed that the anti-log of the difference (B - A on Inside Tires) was nearly the ratio of miles to failure for B/A as given by the data in Table II. Anti-logs taken for the lower and upper confidence limits for the differences likewise became approximate confidence limits for the ratios. These results are also listed in Table III.

Estimated Type Differences in Logarithms and Estimated Ratios of Miles to Failure by Types of Crossbar Fasteners and the Associated Confidence Intervals

Description Difference	Ratio	Estimate	Lower Limit	Upper Limit
Inside Tires				
B-A		0.2406	0.1289	0.3523
	B/A	1.7400	1.3460	2.2510
C-A		0.2292	0.1175	0.3409
	C/A	1.6950	1.3110	2.1923
B-C		0.0114	-0.1003	0.1231
	B/C	1.0260	0.7938	1.3280
Outside Tires				
B-A		0.3363	0.2246	0.4480
	B/A	2.1690	1.6770	2.8050
C-A		0.1490	0.0373	0.2607
	C/A	1.4090	1.0900	1.8230
B-C		0.1873	0.0756	0.2990
	B/C	1.5390	1.1900	1.9910
Combined				
B-A		0.2884	0.2095	0.3673
	B/A	1.9430	1.6200	2.3300
C-A		0.1891	0.1102	0.2680
	C/A	1.5460	1.2890	1.8540
B-C		0.0993	0.0204	0.1782
	B/C	1.2560	1.0480	1.5070

TABLE III

It is to be noted that these confidence limits, when expressed in the original scale in miles, are really confidence limits for a median mileage ratio figure.

ANALYSIS OF TOTAL VARIATION. The analysis of variance in terms of logarithms of miles to failure is presented in Table IV.

The selected chain arrangement as previously described was a cluster of four cross chains for each type of fastener systematically arranged around each tire. When comparing error, it was found the error mean square (based on the clusters) was about equal to the mean square for the cross chains within the clusters (except on the inside tires where there was some difference, but in the wrong direction). It was decided to calculate the pooled error.

The results in Table IV bear out the large differences between the types of fasteners already displayed in Tables II and III. Other points to be noted from Table IV are:

1. Left side versus right side effect is small (about equal to error).

2. The front wheels versus rear wheels effect is large in relation to error although this effect is mostly associated with the outside tires.

3. Individual wheels differ considerably from what might be expected if predictions were based only on the left versus right and front versus rear effects. This effect is shown by the interaction line which is again largest for the outside tires.

4. The difference in miles to failure for outside tires versus inside tires, about 40 miles, does not appear to be a chance effect. Most of this difference was associated with the left front wheel, however.

5. During the detailed examination there was some question regarding the uniformity or behavior of the types of fasteners on the outside tires versus inside tires. Although not shown in Table IV, the effect of an interaction of types by outside versus inside was calculated (removed

Analysis of Variance of Logarithms of Miles to Failure of Cross Chains in Test of Cross Chain Fasteners for Truck Tire Chains

Source of Variation	Degrees of Freedom	Inside Tires	Outside Tires		Combined Analysis
		Sums of Squares			
Total	48	260.8730	278.9757	(96)*	539.8537
Mean	1	259.4277	276.4531		535.7455
		Mean Squares			
Outside vs Inside				(1)	0.1353
Front vs Rear	1	0.00897	0.21902		0.1583
Left vs Right	1	0.00373	0.06701		0.0388
Interaction	1	0.00564	0.12268		0.0987
Wheels				(3)	(0.0986)
Tires				(7)	(0.0803)
Fastener Types	2	0.29490	0.45442		0.6872
Error	6	0.00701	0.03021	(14)	0.0248
Crossbars within Clusters	36	0.02222	0.02844	(72)	0.0253
Standard Error of a Difference	=	.0562	.0562		.0397

$$\text{Pooled Error} = \frac{14(.0248) + 72(.0253)}{86} = .025249$$

Standard Deviation = 0.1589

*Quantities in ( ) are degrees of freedom for the combined analysis.

TABLE IV

from error with 14 degrees of freedom). An F ratio of about 2.46 was obtained when comparing the mean square obtained with pooled error mean square. The probability of such a value of more extreme occurring by chance was about .09.

OTHER TESTS. During the temporary termination of the test at 300 miles, a Median Test described in works by Mood [13] was applied to the data. From this test at the 300 mile level, it was reasonable to conclude that B and C type fasteners were much superior to type A; however, it seemed desirable to perform additional test driving to further quantify the experimental results.

OTHER RESULTS OF INTEREST.

1. Weight loss:

Weight loss measurements for individual cross chains were made at 20-mile intervals during the test program while original chains remained intact. To remove correlation between successive weighings, different cross chains were selected for weighing at the end of each 20-mile interval. At each weighing, one cross chain for each type of fastener was removed from each wheel and then replaced in its original position. With the limited number of cross chains available, it was necessary to repeat weighting at the end of 180 miles. The resulting weight loss data was plotted against distance driven (see Figure 8). These weight loss data show that all cross chains (regardless of fastener type) tended to lose weight at approximately the same rate. Considering this result, the longer life of cross chains associated with the B and C type fasteners must be due to a spreading of the wear over the entire surface of the cross chain produced by the rotational motion permitted by the swivel-type fasteners. Many of the type C swivel fasteners had a forge flashing on the hook shank (shown in Figure 9) which restricted the rotational motion. Depending on whether there were two, one, or none of these hooks with "flash", a particular cross chain might rotate freely, only partially (wind-up), or not at all. Such results could account for the observed difference in life of the B and C types. Figures 10, 11, and 12 show specific wear patterns associated with each type of fastener. The curved wear pattern established in the C type is assumed to be the result of one non-rotating hook causing chain wind-up.

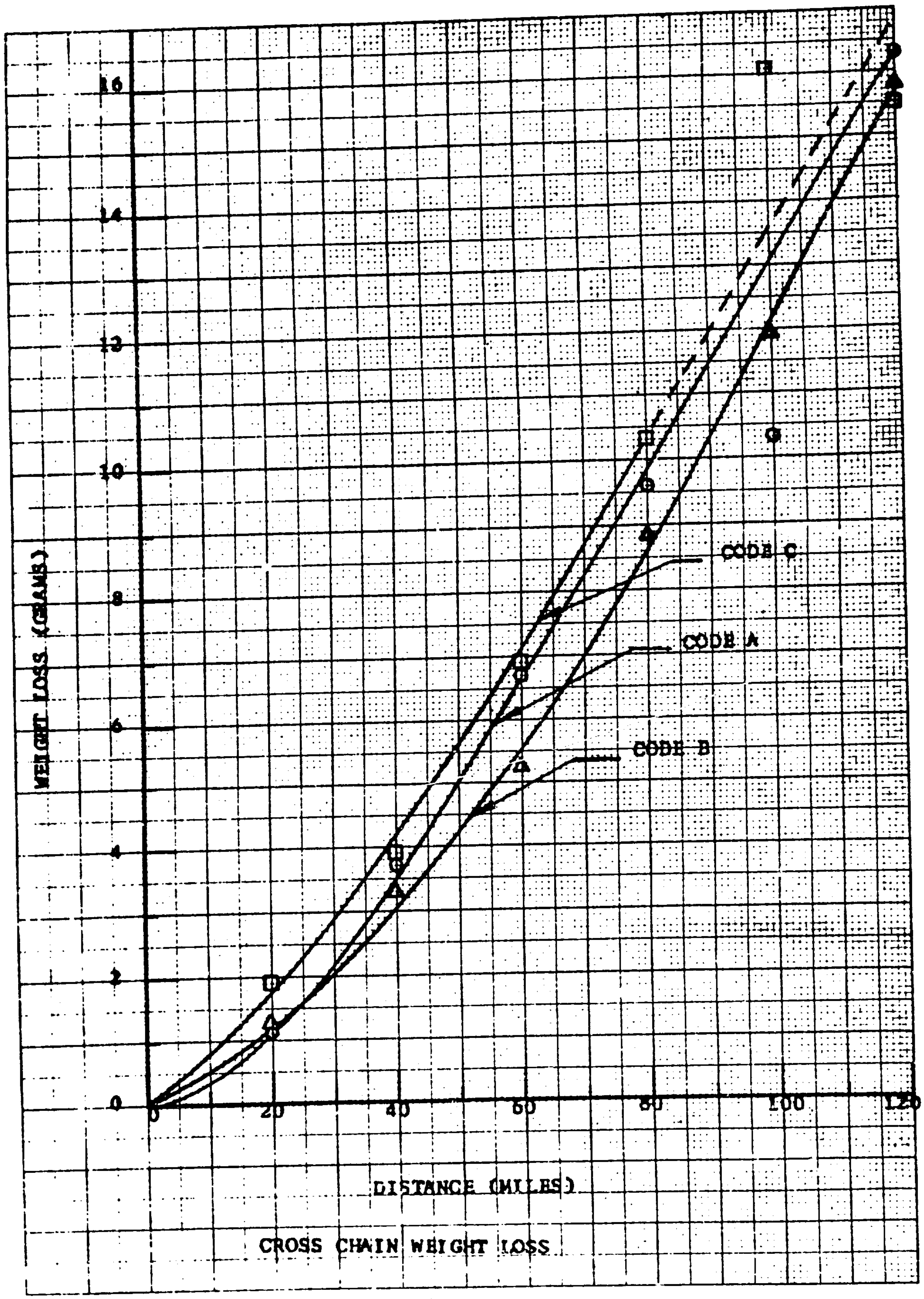
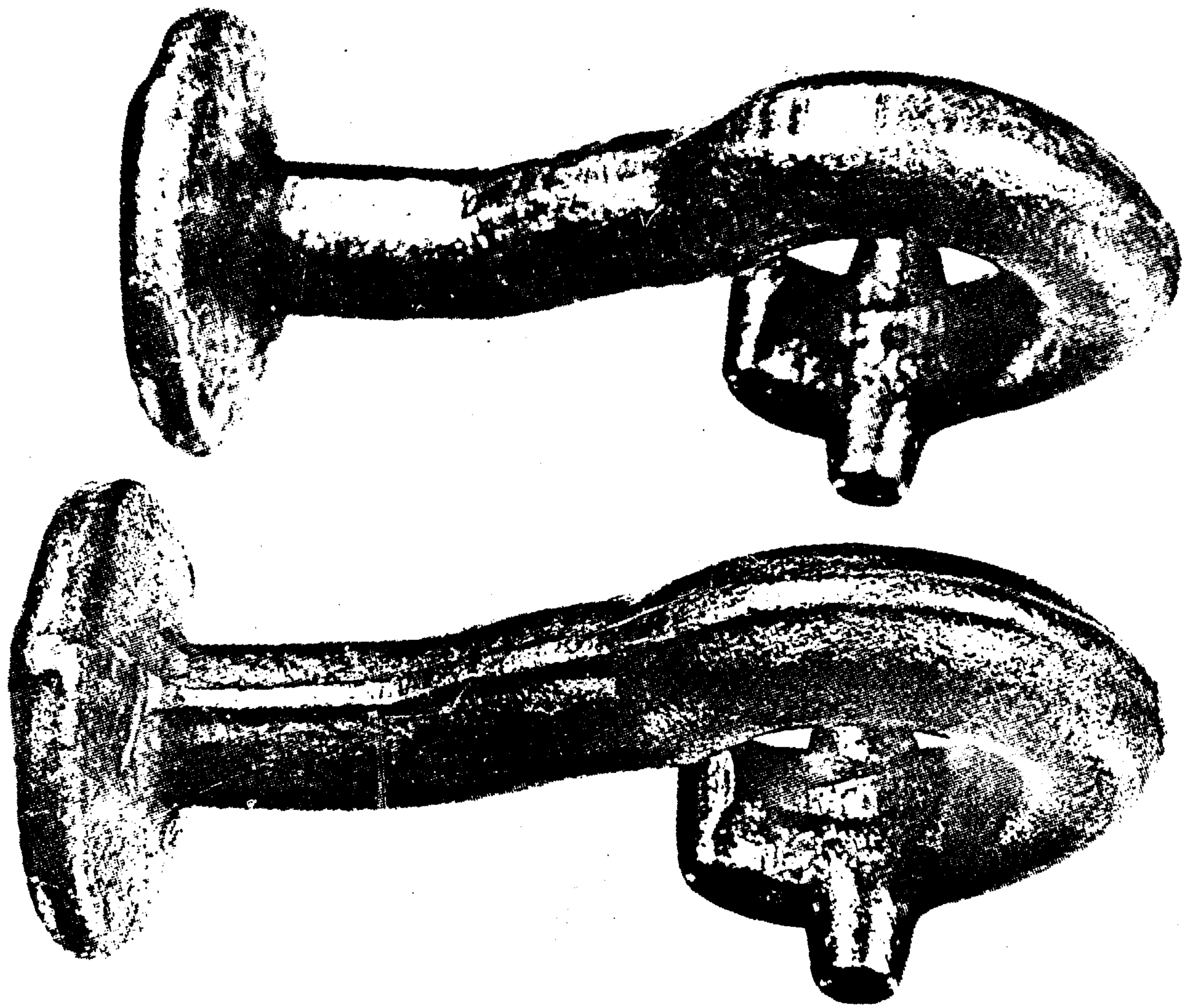
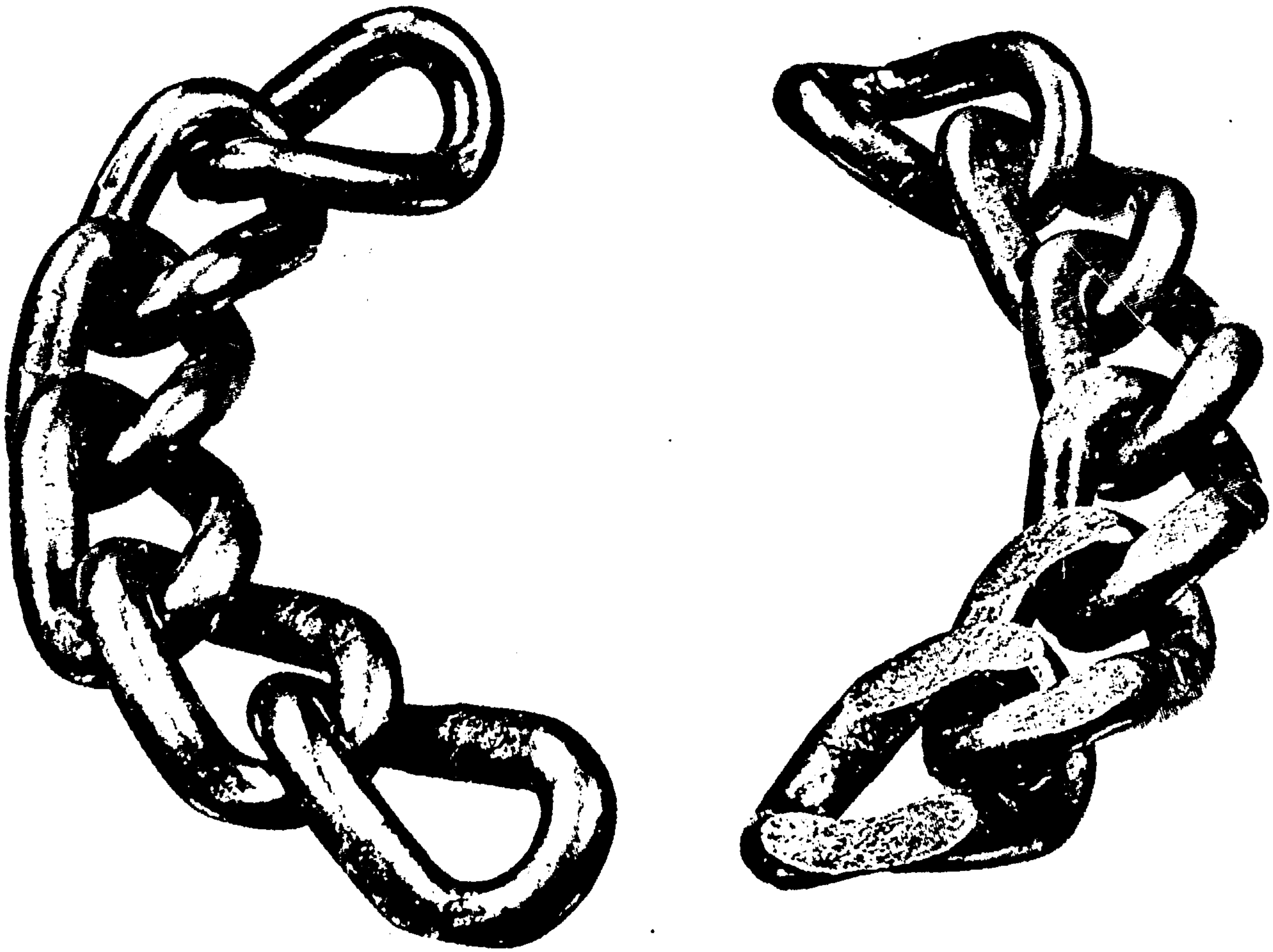

WEIGHT LOSS (GRAMS)
DISTANCE (MILES)
CODE C
CODE A
CODE B
0
2
4
6
8
10
12
14
16
20
40
60
80
100
120
CROSS CHAIN WEIGHT LOSS

FIGURE 8



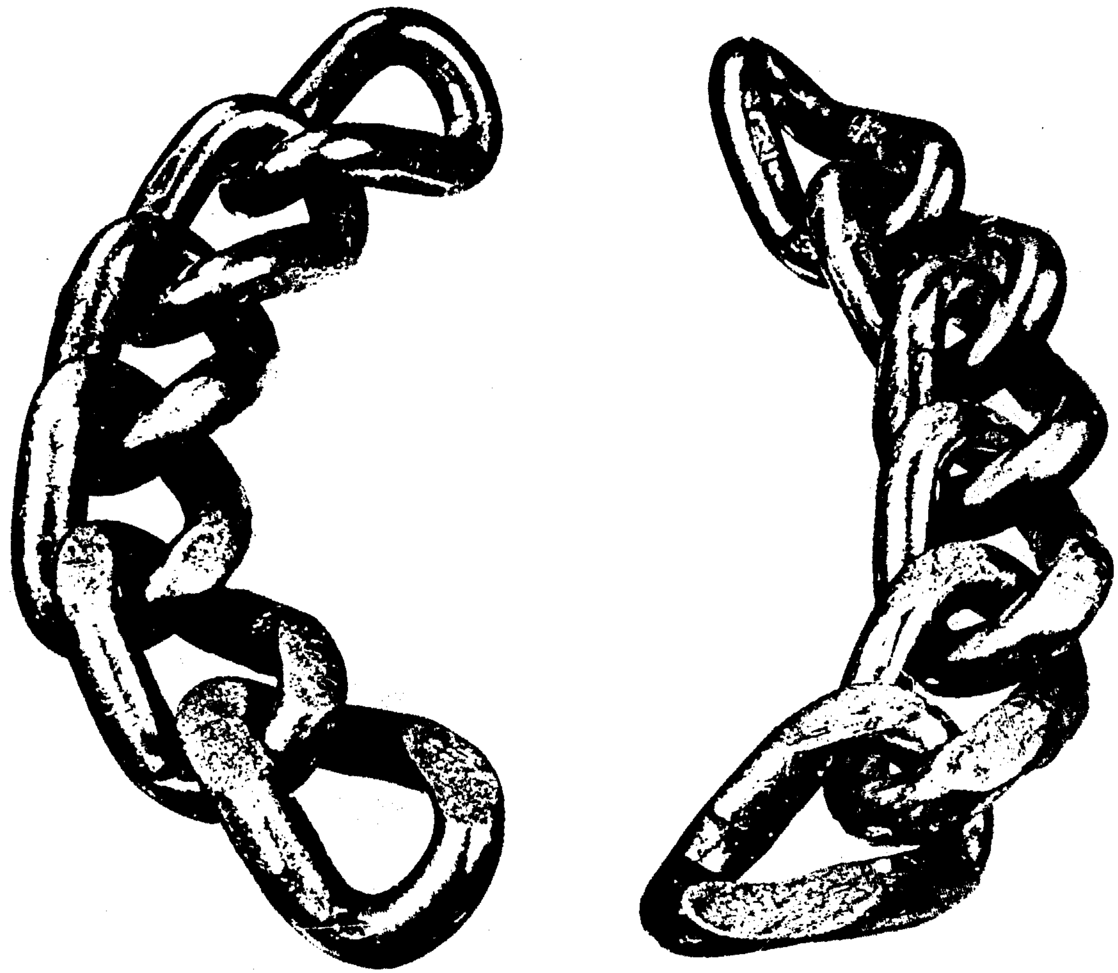
U. S. ARMY TANK-AUTOMOTIVE CENTER NEG. NO. 72213 DATE 20 May 63
Tire Chain Wear Tests. Non-uniformity of Code C connectors. "Flash" present on hook.

FIGURE 9

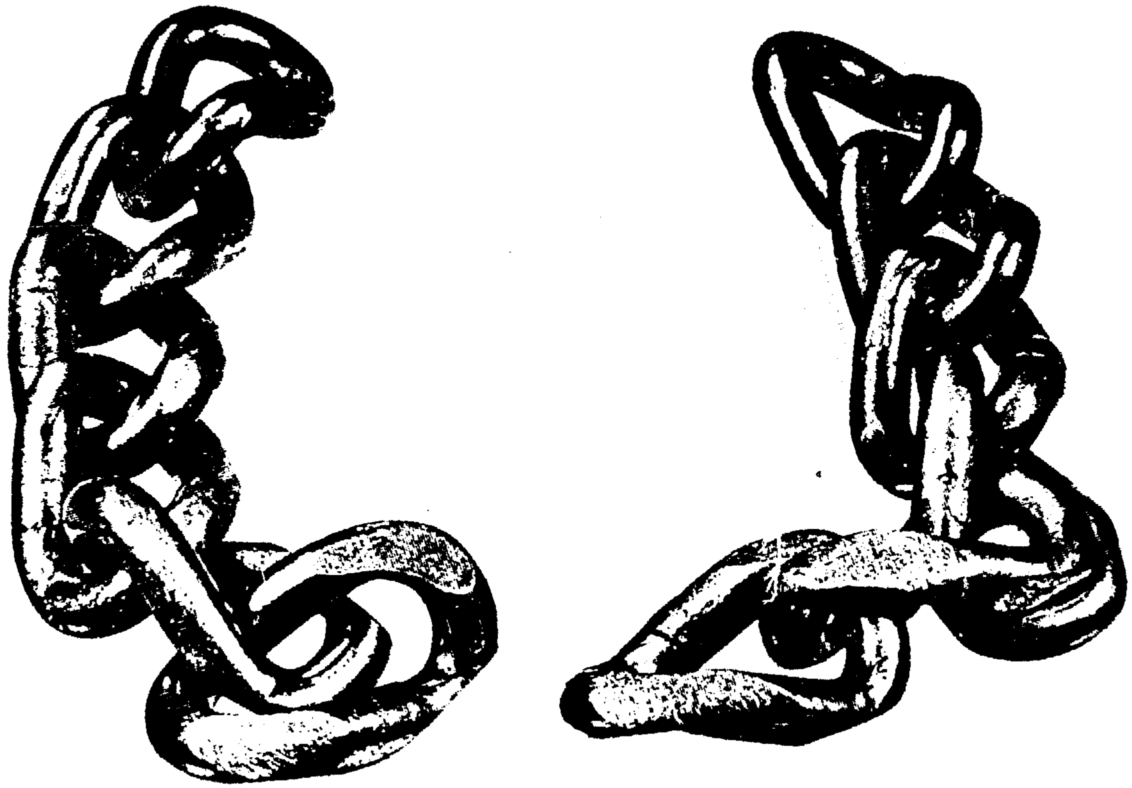


U.S. ARMY TANK-AUTOMOTIVE CEN[illegible] NEG. NO. 72215 DATE 20 May 63
Tire Chain Wear Tests. Typical wear at failure of cross chain connected with Code A type fastener. One-half of chain shown upside down.

FIGURE 10



U. S. ARMY TANK-AUTOMOTIVE CENTER **NEG. NO.** 72216 **DATE** 20 May 63
Tire Chain Wear Tests. Typical wear at failure of cross chain connected with Code B type fastener. One-half of chain shown upside down.
FIGURE 11



U. S. ARMY TANK-AUTOMOTIVE CENTER NEG. NO. 72218 DATE 20 May 63
Tire Chain Wear Tests. Typical wear at failure of cross chain connected with Code C type fastener. One-half of chain shown upside down.

FIGURE 12

2. Replacement Time:

At each replacement of a failed cross chain and also during the weighing procedure in the shop, the time required for removal and reinstallation of the cross chain was recorded. A separate record was kept for shop work and field work. There was some difference associated with the work site; however, the most pronounced difference was associated with type of fastener. An analysis of variance of these differences could have been calculated, but the mean differences in observed time between the three types were so large that a detailed analysis did not seem to be needed. The data were analyzed using the Wilcoxon-Mann-Whitney Statistic (ranking method) [12] [15]. There was found to be a significant difference between the averages in all cases. The average replacement time and re-installation time and ratios are tabulated in Table V.

All removals and installations of the A-type fastener were done with a special tool (see Figure 13) provided for this test. An unsuccessful attempt was made to remove an A-type fastener with the tools provided in the standard tool kit of the vehicle.

CONCLUSIONS. It is clear from the results obtained that the cross chain fasteners type B and C (swivel hook) are superior to the standard type fastener. This superiority is primarily described by comparing average miles to failure or by the ratios of average miles to failure. From these ratios for inside tires only, we observe the swivel-hook fasteners to be about 70 percent better on the average. On the outside tires, we find type B about 117 percent better than type A, and type C 55 percent better than type A. Confidence limits for the true ratios of superiority show a minimum of at least 30 percent improvement on inside tires, and possibly as much as 120 percent for the swivel hooks when expressed in terms of two-sided 95 percent confidence intervals. These results were far less uniform on the outside tires.

The next question raised is, "Are the type B swivel hooks better than the type C?" The observed difference on the inside tires is small. On the outside tires the data indicate a significant difference between type B and C. A large part of the superiority of B and C is found on the left front outside tire. Type B is also better than C on the other three outside tires, but to a varying degree.

REPLACEMENT TIME AVERAGES AND RATIOS

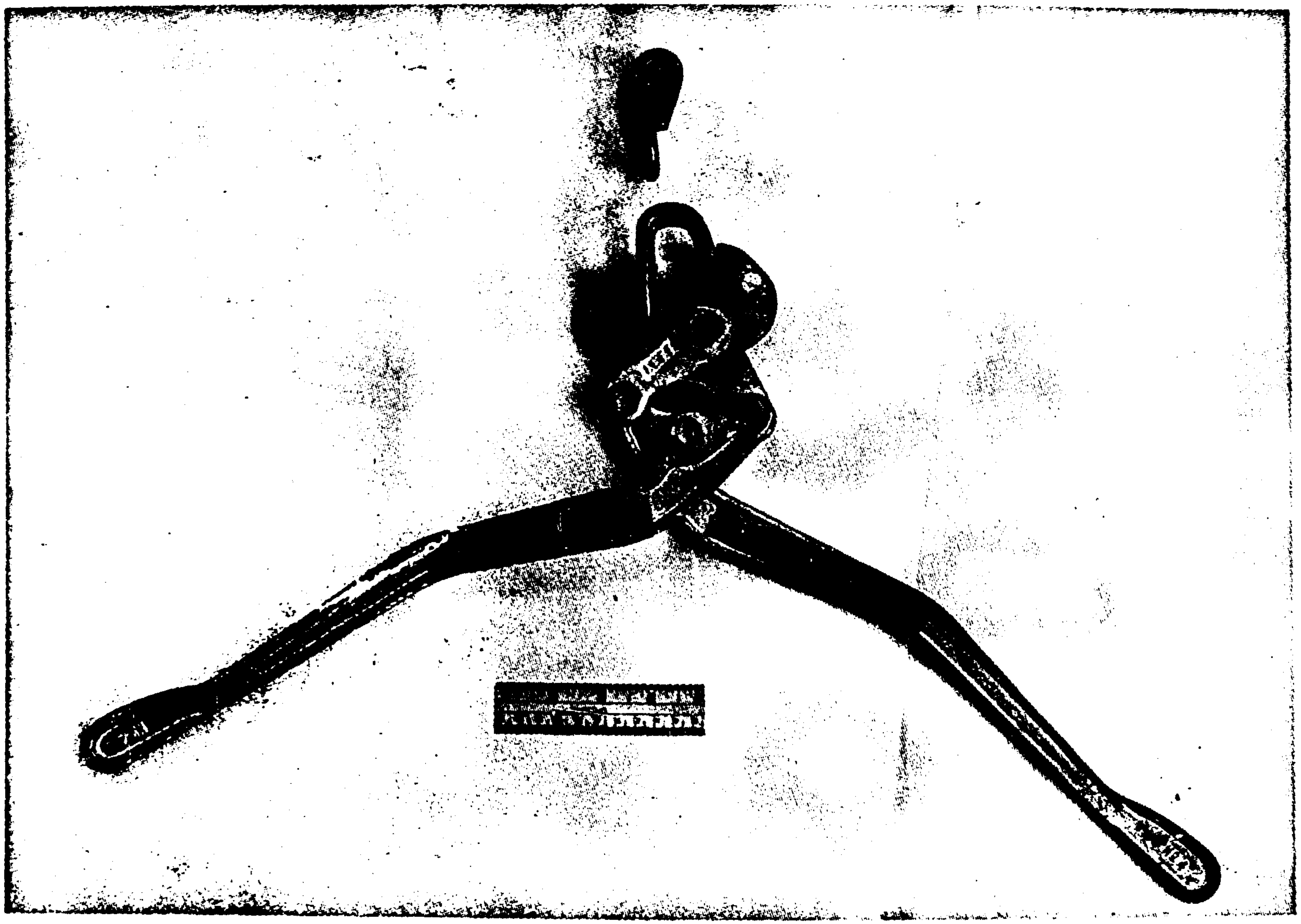
REPLACEMENT OF FAILED CHAINS IN THE FIELD

	CODE A	CODE B	CODE C
Average Time in Seconds	261.4	83.1	55.2
Ratios	A/B= 3.15	A/C= 4.74	B/C= 1.51

REMOVAL AND RE-INSTALLATION OF CROSS CHAINS IN SHOP

	CODE A	CODE B	CODE C
Average Time in Seconds	225.5	98.0	40.5
Ratios	A/B= 2.30	A/C=5.57	B/C= 2.42

TABLE V



U. S. ARMY TANK-AUTOMOTIVE CENTER NEG. NO.P/R 243-64-3 DATE 2 August 63
Special Chain Tool with Code A Fastener in "Ready to Close" position.
FIGURE 13

As described before, it was noted the type B fastener allowed better cross-chain rotation than the type C so that part of the difference between B and C may be ascribed to this characteristic, although the sizeable difference in performance for outside and inside tires is baffling.

The replacement times are highly favorable to the swivel hooks although type C was found to be somewhat better than B. Thus, it seems that minor modifications might make the two swivel hooks about equal in performance and replacement time.

A field trial conducted using chains completely assembled with swivel hooks would be worthwhile to determine the extrapolation factor for normal field conditions from the accelerated test conditions.

When considering the use of the swivel-hook type fasteners as replacements in military tire chains, it appears from the data obtained that the present experiment has been adequate.

REFERENCES

AMS = Annals of Mathematical Statistics
ASTM = American Society for Testing Materials
JASA = Journal of the American Statistical Association

1. Anderson, R. L., and Bancroft, T. A., "Statistical Theory in Research," McGraw-Hill, New York, 1952.

2. Bartlett, M. S., "The Use of Transformations," Biometrics, 3, 39 (1947).

3. Cuevas, R. N., Jr., "Test of Removable Swivel Hooks for Tire Chains," Report No. DPS/TT1-649/53, August 1957, Dev. and Proof Services, APG, Md.

4. Davies, O. L., "Design and Analysis of Industrial Experiments," Oliver and Boyd, Edinburgh, 1954.

5. Epstein, B., "Truncated Life Tests in the Exponential Case," AMS, 25, 555 (1954).

6. Epstein, B., and Sobel, M., "Life Testing I," JASA, 48, 486 (1953).

7. Epstein, B. and Sobel, M., "Some Theorems Relevant to Life Testing from an Exponential Distribution," AMS, 25 373 (1954).

8. Epstein, B., "Estimation of the Parameters of Two Parameter Exponential Distributions from Censored Samples," Technometrics, 2, 403 (1960).

9. Epstein, B., "Estimation from Life Test Data," Technometrics, 2, 447 (1960).

10. Kempthorne, O., "The Design and Analysis of Experiments," J. Wiley and Sons, New York, 1952.

11. Laurent, A. G., "The Lognormal Distribution and the Translation Method ...," JASA, 58, 231 (1963).

12. Mann, H. B. and Whitney, D. R., "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," AMS, 18, 50 (1947).

13. Mood, A. M., "Introduction to the Theory of Statistics," McGraw-Hill, New York, 1950.

14. Neyman, J. and Scott, E., "Correction for Bias Introduced by a Transformation of Variables," AMS 31, 643 (1960).

15. ORDNANCE Corps Pamphlet 20-113, Ordnance Engineering Design Handbook, Experimental Statistics, Section 4, Special Topics, Ordnance Corps, June 1962.

16. ORDNANCE Corps Pamphlet 20-114, Ordnance Engineering Design Handbook, Experimental Statistics, Section 5, Tables, Ordnance Corps, June 1962.

17. Parker, W. H., "The Wearing Qualities of Tire Chains," ASTM Procedures 28 (II), 332 (1928).

18. Snedecor, G. W., "Statistical Methods," 5th Edition, Iowa State College Press, Ames (1956).

19. Zelen, M., "Factorial Experiments in Life Testing," Technometrics, 1, 269 (1959).

20. Zelen, M., "Analysis of Two-Factor Classifications with Respect to Life Tests," paper No. 42 in "Contributions to Probability and Statistics, Essays in Honor of H. Hotelling," edited by I. Olkin, Stanford University Press, 1960.

ERROR ANALYSIS PROBLEMS IN THE ESTIMATION OF SPECTRA

Virginia B. Tipton
White Sands Missile Range, New Mexico

ABSTRACT. Power spectral density functions are estimated digitally by evaluating the Fourier cosine transform of the autocorrelation function. In order to obtain reliable averages with which to describe the autocorrelation function it is necessary to limit the resolution with which it, and its transform, can be described. Is it possible to evaluate, or to express analytically, the accuracy with which the computed spectrum represents the true spectral density function?

INTRODUCTION. The use of power spectral density functions to describe the frequency content of a time function has been a common engineering practice for some time, developing originally from the communications engineers' concern with separating signals from noise in transmission systems. At the same time the statisticians' approach to the study of random fluctuations in time series data led to the development of autocorrelation functions as a descriptive tool. The bridge between these two approaches to the study of noise, which is simply high frequency random variations superimposed on the desired data, was the discovery of the now well-known Wiener-Khinchin relationship. This relationship simply states that, except for a constant factor, the power spectral density function and the autocorrelation function of a stationary random process are a Fourier transform pair. Since the autocorrelation function is an even function of its time lag τ, the complex Fourier transformation process simplifies to a real cosine transformation which can easily be carried out by a digital computer.

The digital computation of power spectral density functions is becoming an increasingly more important part of data reduction work. It is now being applied experimentally to the study of random errors in trajectory measuring instrumentation systems, as well as to the more traditional applications in vibration data analysis and telemetry problems.

However, in order that the spectral estimates computed may be of value to the data user, we must be able to describe in some way the reliability with which the computer spectrum approximates the true spectral density function; that is, we must be able with some degree of confidence to place limits upon the errors of our estimation.

THE WSMR DATA REDUCTION SPECTRUM ANALYSIS PROGRAM. The derivation of the computer programming equations used at WSMR Data Reduction Directorate was first given in a report written in 1957, "The Digital Computation of Power Spectra," by L. M. Spetner of Johns Hopkins University. This digital process is based on the Wiener-Khinchin relationship; that is, it first computes the autocorrelation function of the random data and then determines its Fourier transform, which is the power spectral density function. In order to separate the noise data from any constant (zero frequency) component, the input data are first averaged and then this data mean is subtracted from the original data. This process insures that the average of the residuals will be zero, a condition which must be met if the Fourier transform is to exist. In order to eliminate any linear trend, or a quadratic, a least squares 2nd degree curve is then fit and removed. We are now ready to compute the autocorrelation function of the residuals, which we assume then to be both random and stationary.

At this point it is well to say a few words about random processes in general. A random process is a collection, or ensemble, of time functions such that the ensemble can be characterized by its statistical properties. In studying noise problems we are usually not overly concerned with that individual time function which we happened to observe, since any of the member functions of the ensemble could have occurred with equal probability. Rather we are interested in determining from the observed function the statistical properties which characterize the entire ensemble. For a special class of random processes (that is, for those which are both stationary and ergodic) this can be done because it has been shown (elsewhere) that in such cases the process averages across the ensemble are equal to the time averages along a single representative function from the ensemble (See Figure 1).

The autocorrelation function for a random process is defined as the ensemble average of the product of each function times itself shifted by a time delay τ.

$$R(\tau) = \widetilde{f(t) \cdot f(t+\tau)} , \qquad (1)$$

where the wavy bar indicates averaging across the ensemble.

If we are dealing with a single random function from an ergodic ensemble, equation (1) for the autocorrelation function becomes a time average over the function

$$(2) \qquad R(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(t)\, f(t+\tau)\, dt$$

In the digital case where the integral is replaced by a summation over the range of data points N and time delay m, this becomes

$$(3) \qquad R(m) = \frac{1}{N-m} \sum_{i=1}^{N-m} f(i)\, f(i+m) .$$

This autocorrelation function has several interesting properties:

(1) It is an even function, i.e., $R(-\tau) = R(\tau)$ (a property which is useful in determining its Fourier transform.)

(2) The value of $R(\tau)$ for $\tau = 0$ equals the average power of f(t), or in statistician's language, the variance of the function.

(3) The value of $R(\tau)$ is bounded by its value at $\tau = 0$, so that the computed autocorrelation coefficients can easily be normalized to give unity autocorrelation for zero time delay.

If the function is truly random then its autocorrelation function will rapidly approach zero, since the values of $f(t+\tau)$, as τ increases, are not dependent upon the value of f(t). Thus a typical normalized autocorrelation curve of a random noise record will have the shape indicated in Figure 2.

However, it should also be pointed out that the converse is not so - it cannot be shown that because the autocorrelation function approaches zero as τ increases that the given function is necessarily random.

Once the autocorrelation function has been found, the power spectral density function is computed from it by taking its Fourier transform

$$(4) \qquad \phi(\omega) = \int_{\tau = -\infty}^{\infty} R(\tau) e^{-i\omega\tau} \, d\tau \, .$$

Using the property that $R(-\tau) = R(\tau)$, this becomes

$$(5) \qquad \phi(\omega) = 2 \int_{0}^{\infty} R(\tau) \cos \omega\tau \, d\tau \, .$$

The spectrum is estimated for discrete values of $\omega = \frac{\pi K}{m}$, where K is an index ranging from 0 to m, and m the number of autocorrelation coefficients computed.

The resulting estimates are smoothed using a 3-point symmetric filter, with weights (0.23, 0.54, 0.23) and plotted as function of frequency.

(For reference, the digital computing formulas used in the program will be listed as an appendix to the paper.)

THE PROBLEM OF ERROR ANALYSIS. The problem confronting us now is chiefly this: How can we express the errors involved in estimating spectra by this digital process? Or in other words, with what confidence can we say that the spectrum we have computed represents the true spectrum of the process we are studying? Can we put limits on our error, perhaps in the form of a statement such as, "our estimate is within $\pm$5% of true spectrum" and have perhaps 90 or 95% confidence that we are right?

The problem appears to be in balancing the frequency resolution we can achieve, that is, the number of points used to estimate the spectral curve, against the reliability with which they are computed. The maximum frequency resolution (Δf) in our digital process is determined by the highest frequency we can distinguish in the data (f_{max}) and the number of time delay averages (M) for which we computed the autocorrelation function.

$$\Delta f = \frac{2 \cdot f_{max}}{M}$$

But the highest distinguishable frequency is limited by the rate at which the original data samples were digitized. The sampling rate as given in Spetner's equations must be at least twice the highest frequency present.

By the time the data arrive in digital form at the Data Reduction Computer facility we no longer have any control over the digitizing rate $(1/\Delta t)$ or the length of the data sample $[T(\text{seconds}) = (N \text{ points}) \cdot (\Delta t \text{ seconds})]$. We must assume that the data users chose a sampling rate high enough to minimize aliasing errors, that is, the folding back of frequencies higher than f_{max} so that they appear as some sub-multiple of themselves in the frequency range we can observe.

In addition, the number of time delay averages used to describe the autocorrelation function is limited by the length of the data sample. In practice, we generally limit M to approximately one-tenth of the number of data points N. $(M \doteq N/10.)$ We could increase the number of time delay averages computed, but only at the cost of reliability of them. As M increases the number of data points available to average decreases. Thus this could not solve our problem, and at present, no other solution has been found.

DIGITAL COMPUTING FORMULAS

1. Autocorrelation function is estimated at M points

$$R(m) = \frac{1}{N-m} \sum_{i=1}^{N-m} f(i) \cdot f(i+m), \text{ for } 0 \leq m \leq M.$$

2. Cosine transforms estimated for each value of M

$$L(0) = R(0) + 2 \sum_{P=1}^{M-1} R(P) + R(M)$$

$$L(h) = 2R(0) + 4 \sum_{P=1}^{M-1} R(P) \cos \frac{hP\pi}{M} + 2 R(M) \cos (h\pi),$$

$$\text{for } 0 < h < M$$

$$L(M) = R(0) + 2 \sum_{P=1}^{M-1} (-1)^P R(P) + (-1)^M R(M).$$

3. Smoothed spectral estimates

$$u(0) = 0.54 \, L(0) + 0.46 \, L(1)$$

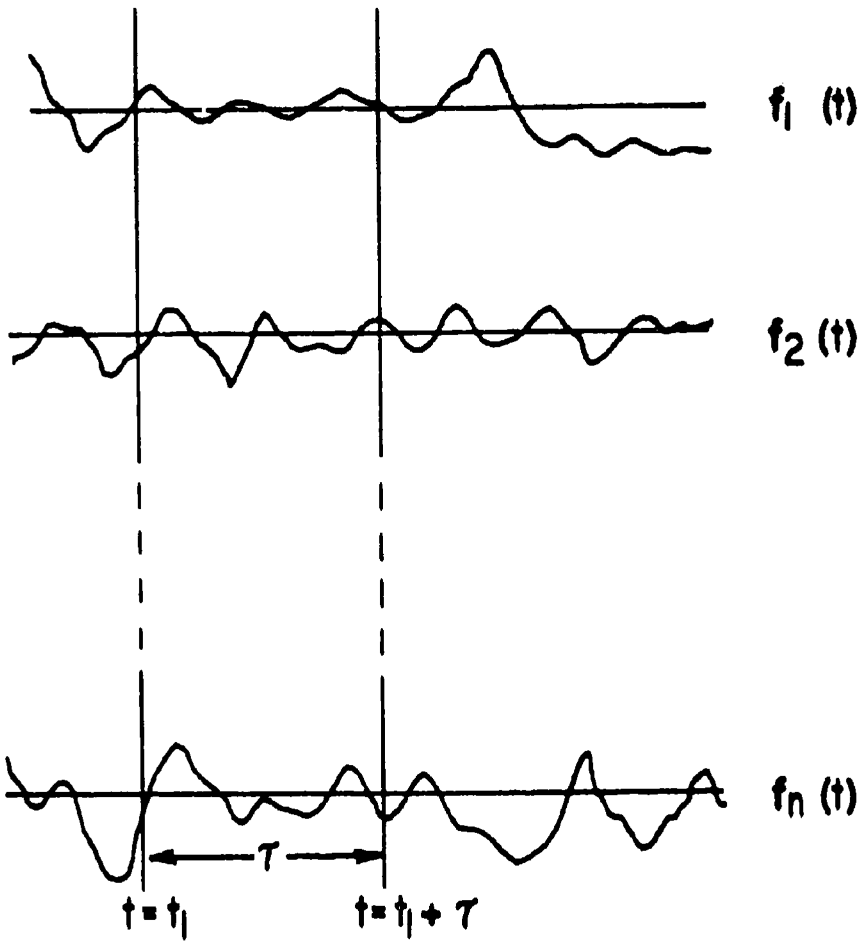
$$u(h) = 0.54 \, L(h) + 0.23 \, [L(h-1) + L(h+1)] \text{ for } 0 < h < M$$

$$u(M) = 0.54 \, L(M) + 0.46 \, L(M-1) .$$

4. Plot of $u(h)$, for $0 \leq h \leq M$, where Δt is the time interval in seconds between original data points, vs, frequency, $f(h)$, where

$$f(h) = \frac{h}{2m \, \Delta t} \text{ cycles per second.}$$

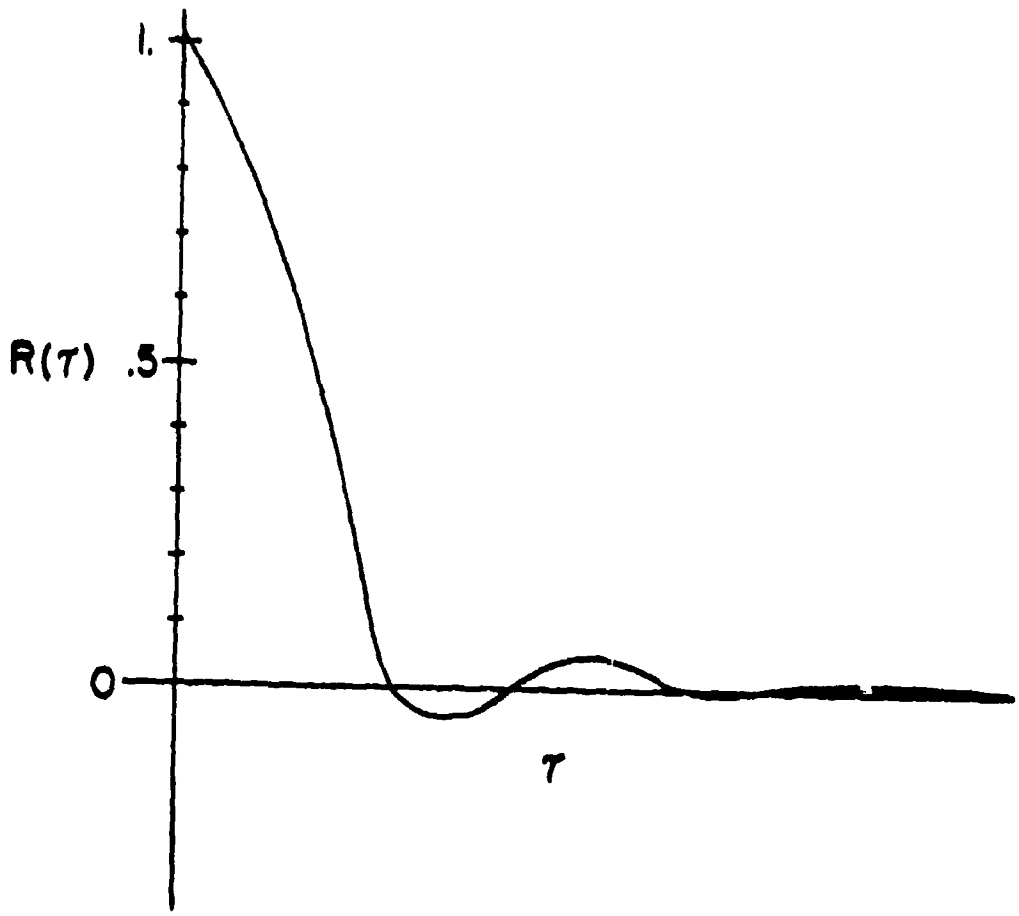
AN ENSEMBLE OF n TIME FUNCTIONS



ENSEMBLE AVERAGE $\widetilde{f(t = t_1)} = \frac{1}{N}\left[f_1(t_1) + f_2(t_1) + \cdots + f_n(t_1)\right]$

TIME AVERAGE $\overline{f_2(t)} = \frac{1}{N(\Delta t)}\left[f_2(t_1) + f_2(t_2) + \cdots + f_2(t_n)\right]$

FIGURE 1.



A TYPICAL NORMALIZED AUTOCORRELATION FUNCTION FOR RANDOM NOISE

FIGURE 2.

VALIDATION PROBLEMS OF AN INTERFERENCE PREDICTION MODEL

William B. McIntosh
U. S. Army Electronics Proving Ground

The Electromagnetic Environmental Test Facility (EMETF) of the US Army Electronic Proving Ground, Fort Huachuca, Arizona, is being developed to give solutions to a host of communications-electronic problems, most of which in one way or another arise from the fact that military demands imposed upon the electromagnetic spectrum require vastly more space in the spectrum than is available for this purpose. A consequence of the resulting crowding is interference between electromagnetic equipments. The EMETF is designed to provide experimental data bearing on the interference problem in its broadest sense. A subsidiary and included feature is provision of data on the ability of communications-electronic systems to perform intended missions in the absence of competing electromagnetic signals.

Although the ultimate EMETF will be a facility capable of providing these answers for communications equipment, for radar, for navigation devices, for data transmission links, and in fact, for all army electronic activities, this presentation will be confined to aspects of voice communication by radio.

Several years ago, the EMETF was conceived of as primarily a huge outdoor field test facility, spread over some 2400 square miles. In the District of Columbia area, this facility would have stretched from one side of Washington to the other side of Baltimore. An artist's concept of the field facility is shown in figure 1. Originally, 24 transmitter sites and two transmitter-receiver siteswere deployed; the latter two sites were the test sites. At each of the transmitter sites equipment was grouped around a control van. The master control center is located at one of the test sites. From the center, one or more transmitters at any one or more van sites may be controlled for test purpose.

The basic test unit is a cycle, figure 2, which requires 30 seconds; it consists of energizing the desired transmitter, and recording the test link performance, as will be described later. Then the entire environment is turned on, and the link performance again measured. If degradation has occurred, the next step - and a long one - consists of a search for the one or more transmitters responsible for the degradation. These operations produce the basic field data.

The concept of test, and with it the field facility, has evolved to the point where now the field facility has been reduced in size, and the major share of useful output is being derived from a computer simulation program known as the Interference Prediction Model (IPM). The IPM requires a considerable amount of input data of various sorts, which has lead to the creation of a third unit, the Instrumentation Workshop.

The ultimate form of the IPM, as it is currently envisioned, is shown in figure 3.

The model requires input data specific to the equipment or concept under test as well as a specific description of the problem. Internally, the model consists basically of five modules. Figure 3 also shows in block diagram form how each module is developed and validated.

The propagation module performs one basic function. It describes the attenuation which is expected as an electromagnetic wave front travels between a transmitter and a receiver.

The equipment module incorporates the necessary equations which describe what happens to electrical signals as they pass through the equipment. The scoring module translates the processed signal at the receiver output into a measure of how much of the original intelligence remains. The tactical deployments module incorporates several preselected deployments, each containing the physical location, in three dimensions, of every piece of emitting or receiving equipment, plus information on the topography in the form of an XYZ matrix, in meters, at 500 meter intervals in the XY plane. The radio frequency assignment also becomes a part of this module.

The statistical module is at present largely undeveloped. In time, however, it will be used to convert an essentially deterministic model into a stochastic one. The essential purpose of this paper is to describe certain problems which have been encountered in an attempt to provide an interim stochastic capability by different methods.

I will next outline a test problem in which the important outputs will be obtained from the IPM, but for which actual hardware of the test item is available for certain measurements.

The preparation consists of establishing the location of all equipments, both of the test type and other types which will share the environment, the assignment of radio frequencies, the inputs of the equipment characteristics, and the like. From the entire collection of communication nets in the chosen deployment, a sort of stratified random sample is chosen for test. Stratification is based upon the frequency of different type nets as well as on the relative importance of net type to mission success. Thus, if there is only one command net from Corps to Division, that net is included. From the many command nets between, say, infantry platoons and squads, several are chosen at random.

In the present use of the model, one basic question is asked. What, on some scale, is the overall system effectiveness? This question is often asked for a standard system and for a proposed replacement, whereupon the comparison will provide useful information to those who make procurement decisions. The systems effectiveness measure now in use depends upon a somewhat involved procedure which results in every test link being classified as providing or not providing acceptable performance. Then the effectiveness measure is merely an index, being the ratio of acceptable links to the total tested. Since the initial measure, intelligibility, is changed from a continuous variate to a binomial, many links which have been measured inaccurately will still be classified correctly. Further, to the extent that the model is imprecise but lacking in bias, errors of classification in one direction will tend to be balanced by other errors in the opposite direction. Thus, for the index, we really do not need to be concerned in great detail with the goodness of our answer for each link.

But people do ask such questions. And ultimately we would like to answer such queries as: How well can some specific platoon communicate with its company headquarters? We no longer will be satisfied with knowing how well on the average a platoon can communicate with a company, nor will we accept a simple yes - no answer.

Given this ultimate desire to answer questions about any communications link, of necessity we must accept a stochastic answer. Even if we have developed a perfect model, in the technical sense, this will be true. Communications equipment will continue to exhibit interunit variability; operators will not all have identical hearing ability or training; and most of all, propagation loss will continue to be an important variate. Even if we should come to know the form and moments of all pertinent distributions of equipment characteristics, we won't have any way of knowing the individual characteristics of the specific equipment at any given geographic point. Anyway, in reality, those equipments may be a few to a few hundred meters at least away from where we have them located in the problem. Even though we learn all there is to know about the effects of atmospheric conditions, terrain, and vegetation on path loss, we can't know what the precise atmospheric factors would be if the tactical operation we are simulating were to exist. Nor can we precisely describe the minutiae of the terrain over all direct paths and multipaths between all necessary pairs of points taken from a vast set.

The best we can hope for -- and this it seems is realistic -- is to say with some chosen confidence, that _if_ equipment is approximately where it is supposed to be, _if_ the various environmental conditions are approximately those used as model inputs, _if_ we have studied a sufficiently large sample of the equipments in question -- then a given communication link will exhibit a performance somewhere between points A and B on some scale.

It may have become apparent that the word validation is being used in the EMETF in two somewhat different contexts. One may be called validation for development. This consists of whatever tests or comparisons may be useful in checking out the development of a module, particularly the propagation, equipment and scoring modules, as implied in figure 3.

The IPM is designed to be a theoretical model rather than an empirical one. That is, it performs the calculations textbooks and research papers give in explanation of what happens. It is not supposed to store empirical data on what has been observed at various times and places, and regurgitate the solution to the stored problem which is most similar to the desired problem. While this theoretical approach is the cause of considerable grief during development, the advantages of a good theoretical model over an empirical one are evident.

However, as it happened, we could not wait the millenium without useful outputs from the facility. Our sponsors presently began to clamor for results, and results we had to produce, even though we knew that none of the modules performed to our standards. We were thus, in part, forced into interim empirical solutions. From this there arose another concept, validation for utilization, a statement about the goodness of our results. We shall henceforth be concerned here with this latter sort of validation.

A block diagram of the entire test problem, figure 4, will be a helpful introduction to the next section. This shows the test transmitter with its special test signal and the propagation path to the test receiver and scoring device described later, labelled VIAS. The figure also shows potential interfering transmitters (IG), each of which is supplied with normal modulation, i.e., voice, radioteletype, etc.

The two scales below the diagram are merely subjective guesses, and no precise quantitative interpretations should be given to them. For example, on the adequacy of representation scale, which applies to the IPM, we know that both the test signal and the test transmitter are represented more adequately than is the test receiver. This is because most of the various things which happen to a signal as it passes through the equipment occur in the receiver. We also note that the receiver in turn is more adequately represented than is propagation loss.

The bottom scale includes not only factors related to less than perfect representation in the model, but also includes variation in electrical characteristics among equipments, time-dependent variation over a propagation path, and the like.

However inaccurate these judgements may be, they did provide some guidance for separating the total problem into parts. One easy choice to make, and one which is also required by the operational scheme, consists of fragmenting the problem into the interfence versus the non-interference cases; that is, study the problem with and without the interfering transmitters activated. The balance of this paper will be concerned only with the non-interference case.

Another division point was taken at the input to the receiver. This was selected on a recent test for several reasons. One is, it appears from the diagram that the first portion of the chain, from test signal to receiver input, would be basically a measure of the ability of the model to predict propagation loss. Another reason is that better measurements can be made on the low level signals at the receiver than on the high level signals at the transmitter. A third is that in the workshop we studied in detail the receiver-VIAS subsystem, and here the input to the subsystem was of necessity the input to the receiver. Thus the non-interference case was divided into two segments.

From the field facility, data on the received signal can be obtained from a number of transmitters at one or more receiver sites, each transmitter-receiver combination defining a path. In the IPM, these paths may be simulated and the computer signal at the receiver obtained for each. Thus we generate a set of bivariate data as shown in figure 5.

These data were treated by the method of simple linear regression. Since the regression will be used to provide an interval estimate for the expected value of a hypothetical "field" signal given a model signal, the latter was used as the independent variate. The confidence band shown is that for the line as a whole. In other words, it is based on the tabular factor

$\sqrt{2F_{2, n-2}}$ rather than on t_{n-2}, which is valid for only one

prediction of the expected value for Y, given X. This confidence belt has roughly 25 percent greater width than the one of the same confidence level which is computed using the Student t. We used this method for showing how well, on the average, the IPM was predicting path loss.

The first specific question directed to the Panel arises here. Is there a method for providing an interval estimate for any number of individual predictions?

To lead into the second question, further details are helpful. The regression is based on the received signal measured in negative dbm, that is, in decibels below one milliwatt. This is a measure of the signal power induced across the input impedance of the receiver. The gain of the receiver antenna is thus included in the signal power measurement. In practice, however, it may be necessary to measure the field intensity, a voltage impinging upon the receiver antenna. Thus the dbm measure shown may contain a computed element. Most likely this would be computed once for each type of antenna-receiver combination, and would not include interantenna variability or variable ground plane effects. Thus the dependent variate may among other things contain a fixed, computed component rather than a measured, variable one. Clearly, we need in such cases to assess the effects on our predictions.

This and other examples not cited pertain to the general question of whether we do in fact satisfy the several assumptions inherent in the regression model, which now suggests the second question. If conventional regression, upon further examination, is not applicable, can the Panel suggest alternative approaches? Remember that for subject matter reasons we desire to obtain, at approximately this point in the chain of events, a measure of how well the IPM is performing its job.

I will close this section by pointing out that we fully realize the measure of our ability to predict path loss over one terrain type, in an area of sparse vegetation, is not necessarily indicative of how well the model will perform under other terrain-vegetation combinations. The Army and others are presently engaged in collecting propagation data in various areas of the world. At present, however, the Arizona data are all we have to work with. It is our hope that by the time suitable data from other areas are available, we will have established the techniques to use these.

Before proceeding to the second portion of the non-interference chain, it will be helpful to describe the scoring device. The Voice Interference Analysis Set (VIAS), is a commercial device designed to convert signal-to-noise type information from the terminal end of the receiver audio section into a measure which is monotonically related to intelligibility. The result is the Articulation Index (AI). The conversion is accomplished by subdividing the audio frequencies from 200 to 6100 cycles per second into 14 bands, each of which is supposed to contribute equally to speech intelligibility. In each band the signal-to-noise ratio is measured during 17 seconds of the 30-second test period. For signal-to-noise ratios of +18 db or higher, the signal-to-noise ratio is converted to unity; for ratios of -12 db or lower, the conversion results in zero. In between +18 and -12 db, the conversion is approximately a linear function of the signal-to-noise ratio. The final articulation index is simply the mean of the 14 increments. There are some additional manipulations involved, and a special test signal is required, but these details need not concern us here. This device is based upon studies by French and Steinberg, and by Beranek.

It should also be noted that if a voice communication system does not possess the full bandpass of 200 to 6100 cycles per second, the VIAS bounds the AI between zero and some value less than unity.

The second segment of the chain concerns the radio receiver and the scoring device. In the shop, a rather precise curve can be established which relates the signal power, developed across the receiver input impedance, to the AI output. This curve generally resembles that shown in figure 6. Three things should be noted. First, the figure shows hypothetical data and, if anything, the point scatter is excessive. An individual receiver produces data points which scarcely deviate from a smooth curve. Second, what information we have indicates that variation among receivers results primarily in a horizontal translation of the curve by no more than a very few db. This is apparently the result of variation in receiver sensitivity. Third, measurements which fall in the lowest fourth or fifth of the AI scale are difficult to make, and these exhibit a higher variability than those resulting from stronger signals.

It should also be stated that the effects of varying the modulation level at the transmitter are at present unknown in detail, but presumably are very important.

The mathematical nature of the AI/signal relationship is not known from theory, as far as we have been able to ascertain. An understanding of the manner in which the VIAS operates on the signal clearly explains the rounded corners. It also allows for a strictly linear portion in the descending leg of the curve, at least for some values of signal and noise. Finally, the bounds on the function are easily understood. Perhaps this is enough.

In practice, the probit transformation was applied to the AI axis and a reasonable linear trend was established. Although the potential applicability of the probit transformation is not immediately obvious, a study by our contractor indicated that it could be used. Confidence intervals were established by the methods appropriate to probit analyses, and then mapped through the inverse of the transformation to provide the confidence belt shown in figure 7.

The three curves to the right represent the fitted line and its confidence bands for a specific equipment type. The single line to the left shows only the fitted curve for another equipment type.

Note that the line on the left drops from maximum AI to zero over a spread of about 25 db, whereas the other curve takes a little over 50 db to drop to zero. When it is considered that the power output of the transmitters normally associated with these receiver types is in the vicinity of +40 to +45 dbm, we see that a receiver over most of the possible range of signals is either performing at its best, or else is not extracting any intelligence whatever from the desired signal.

The use of the probit transformation was an expedient. It is clear that in some cases it is not appropriate, clear because the best fitting probit line obviously does not fit well. In particular, if there is a considerable segment of the curve which is linear in the descending region, the probit transformation is not suited.

The third question for the Panel is this: Please comment on the problem of providing both point and interval estimates for the functional relationship between input power and output AI.

An earlier topic, the scoring device, dealt with a conversion from an electrical measure, the signal-to-noise ration, to a psychoacoustical measure, aural intelligibility. The question naturally arises: Is the AI scale a suitable measure of intelligibility?

Previous work, notably by Kryter, had shown that AI was not linearly related to the articulation score (AS), where the latter is defined as the proportion of words recorded correctly by a listener. Kryter showed, further, that different AS/AI relationships were obtained depending upon the size of the word list. He and others have shown or suggested that such other factors as the type of noise, i.e., white noise, voice babble, or meaningful single voice interference, also affect the AS/AI transformation. We have recently verified that the electronic circuitry of the communication equipment also affects this relationship.

While there are some theoretical results which predict the functional relationship between AS and AI, our position is that, at present, the relationship must be established empirically. Naturally, we anticipate the day when the appropriate theory, to include parameter values, has been established and can be used to convert, in the IPM, from the last electrical-type measure to intelligibility.

The articulation score, as we use it, is defined as the mean proportion of correct responses given by five listeners to a transmission involving one of several 50-word phonetically balanced lists.

The experimental procedure requires that the word list be transmitted and recorded on magnetic tape. The transmission also includes the special test signal required for the AI measurement. For various reasons, we now imbed the test words in carrier phrases, and this procedure necessitates a transmission time of 16 minutes. Each tape is scored in a special listening facility by five operators. The AI signal is scored separately by the VIAS. Thus, one transmission produces one AS/AI datum point.

Figure 8 presents some recent AS/AI data acquired from different equipments. The actual points are shown only for the middle curve. The scatter of points shown is roughly typical of each curve.

These curves were supplied by our contractor. The center and lower curves are based on the Gompertz curve while the upper one is hyperbolic. The Gompertz curve,

$$Y = ab^{c^X}$$

with a taken as unity, has been given some theoretical justification by previous psychoacoustic studies. It was fitted in linear form by means of a log log transformation in which Y is 1/AS and X is AI. The transformation enabled simple linear regression techniques, including confidence bands, to be applied. The confidence bands were mapped through the inverse of the transformation to provide an approximate confidence belt for the line as a whole.

The next question to the Panel is doubtless now evident. Please comment on the problem of converting AI to AS.

In summary, we began with a complex total problem, restricted it to voice communication, and further restricted it to the non-interference case. The non-interference case has been broken into three segments, each treated as a regression. The dependent variate for one becomes the independent variate for the next, with the articulation score as the ultimate dependent variate. At present, approximate and conservative confidence limits can be placed on the expected value of the AS. We are aiming for the ability to place exact confidence limits on the individual AS predictions.

My final question to the Panel asks for discussion of this problem of ultimate interest.

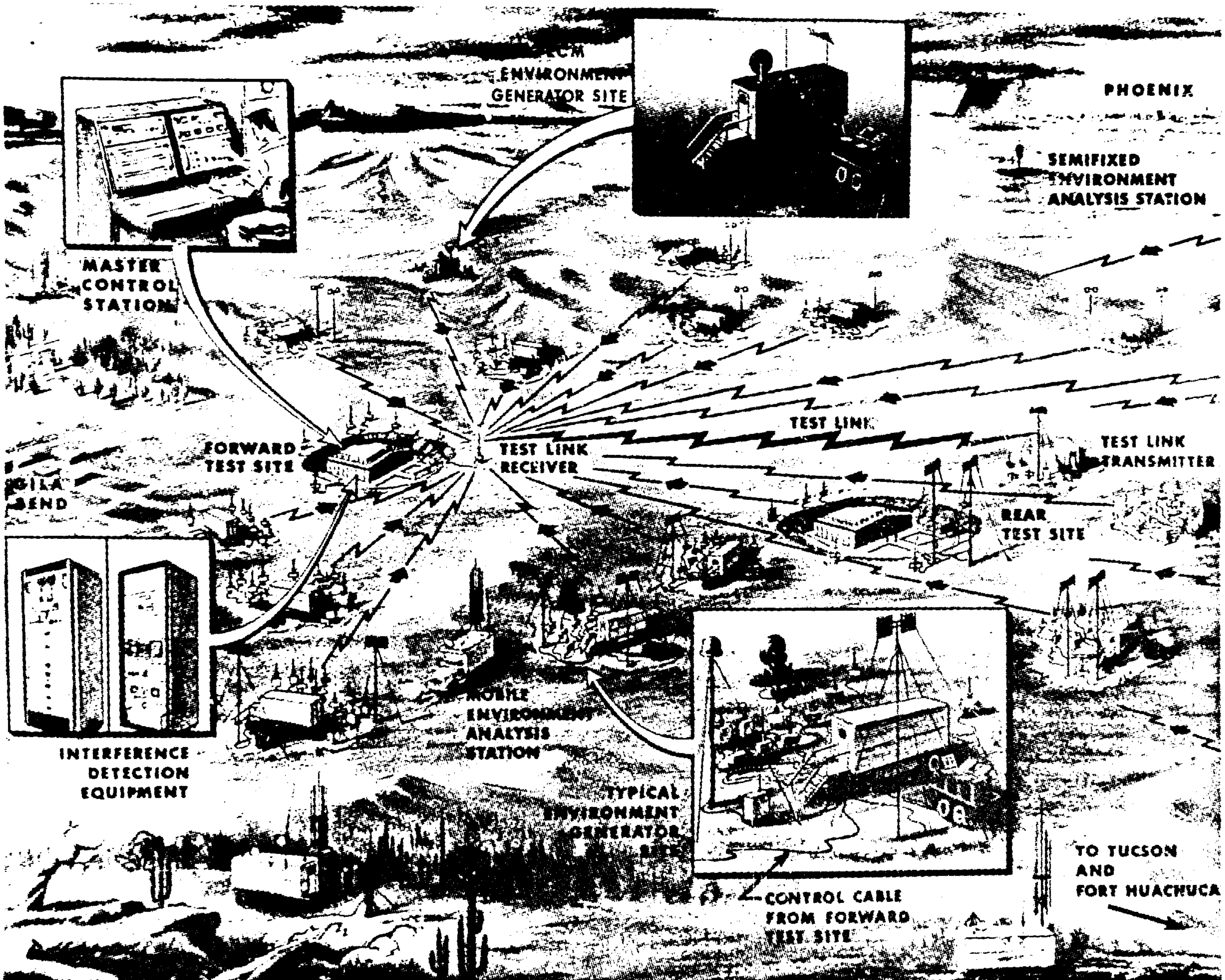


Figure 1. Artist's concept of the EMETF field facility.

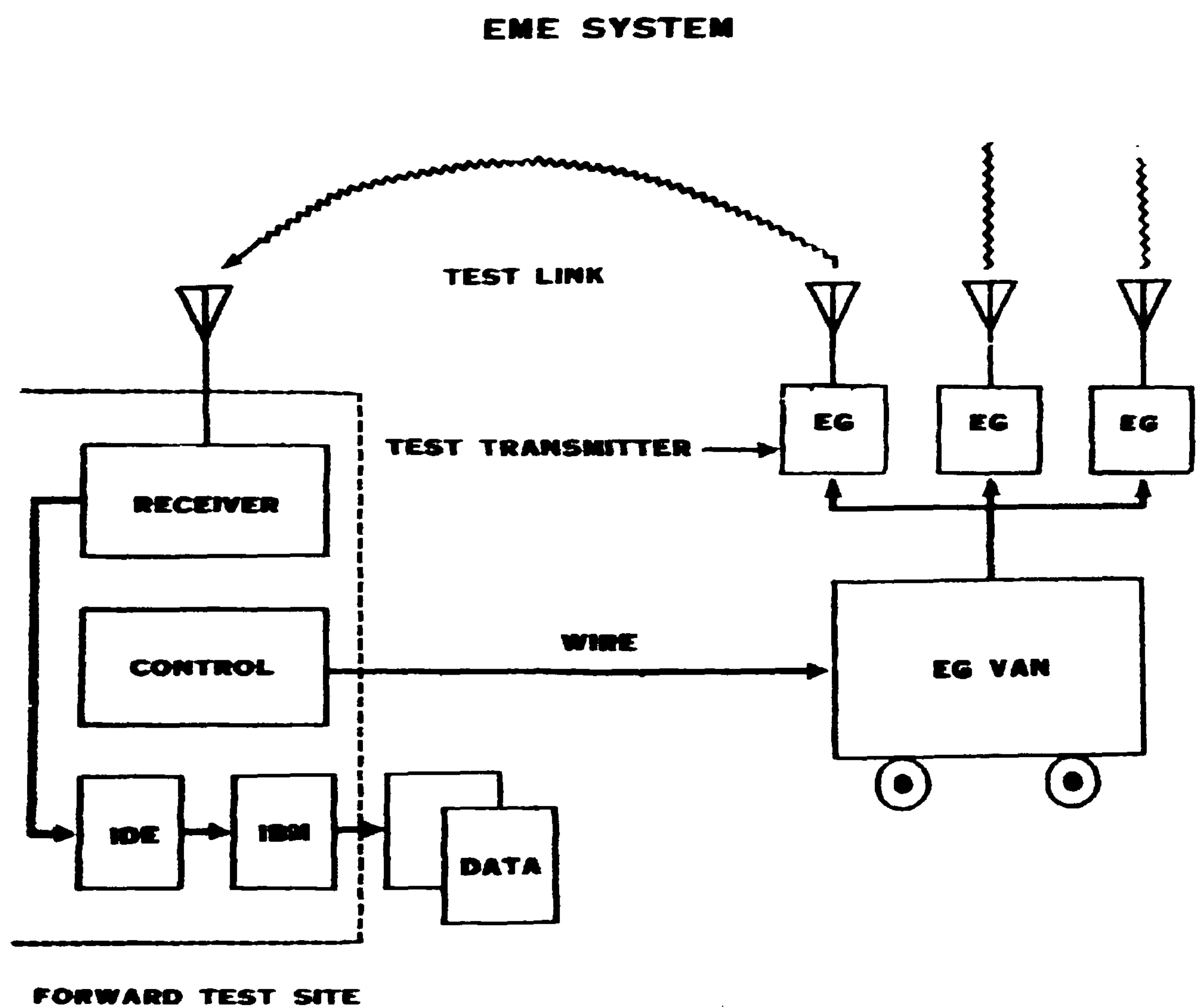


Figure 2. Test cycle of the EMETF

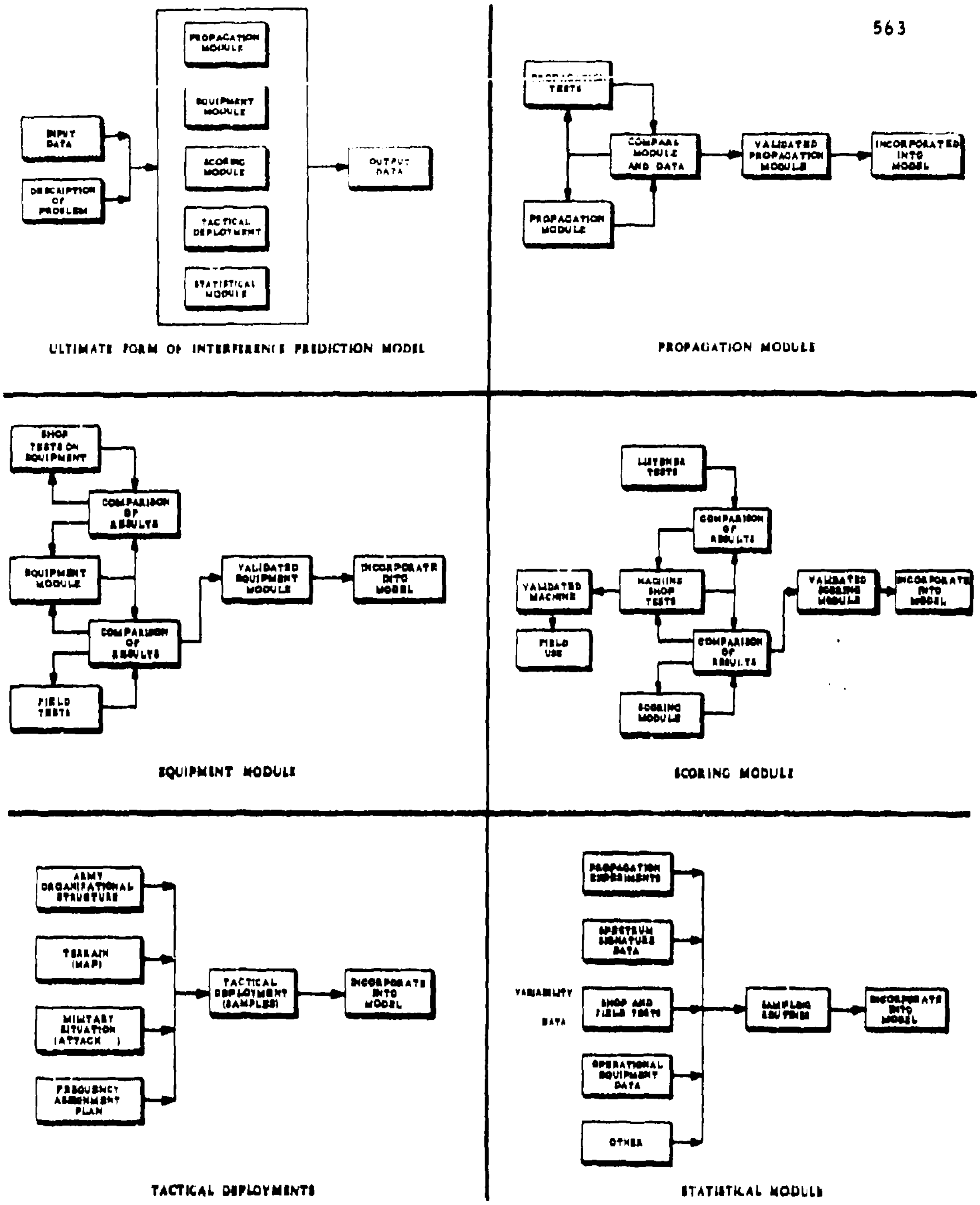


Figure 3. Block diagrams of the Interference Prediction Model and its Modules.

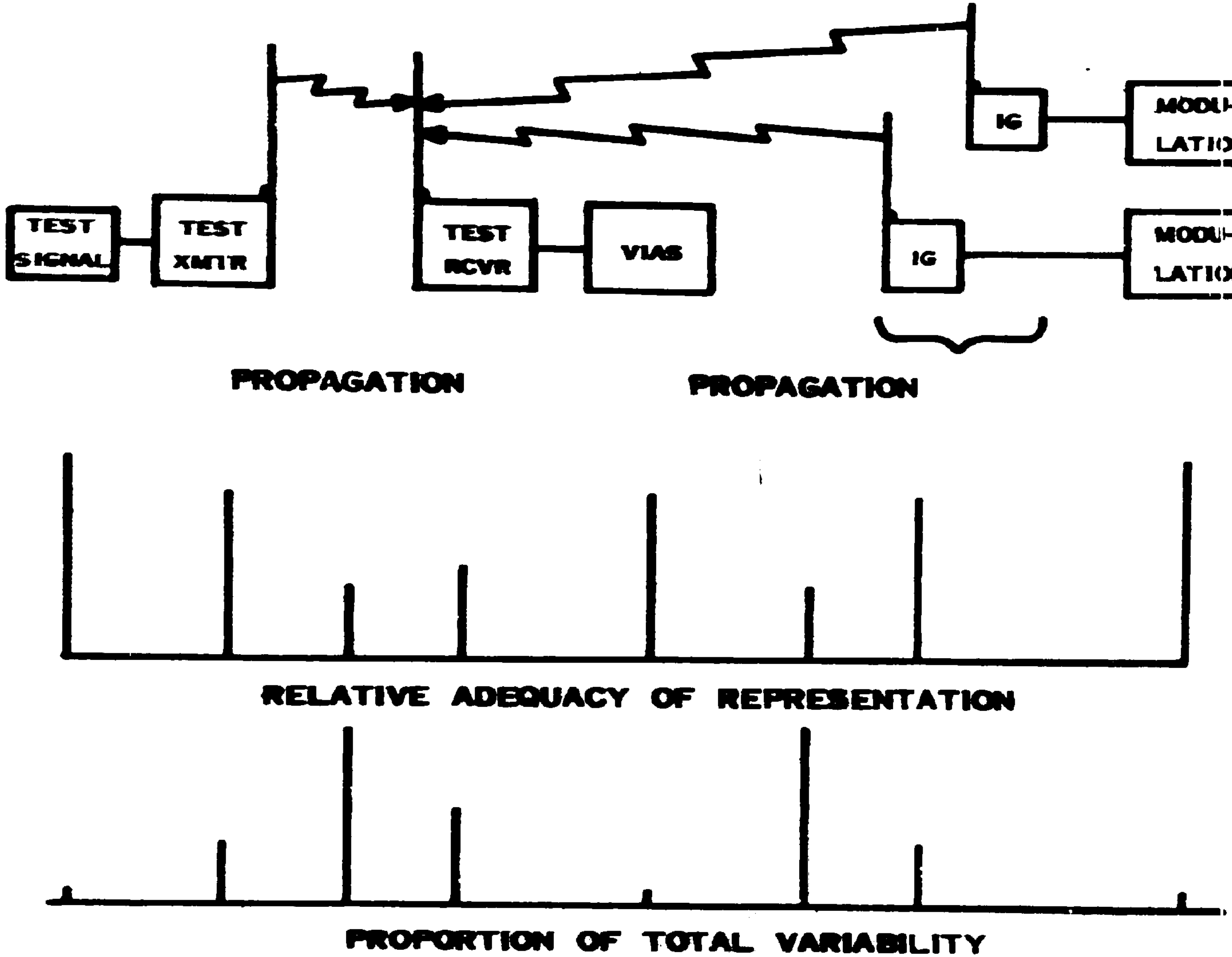


Figure 4. Block diagram of a single test.

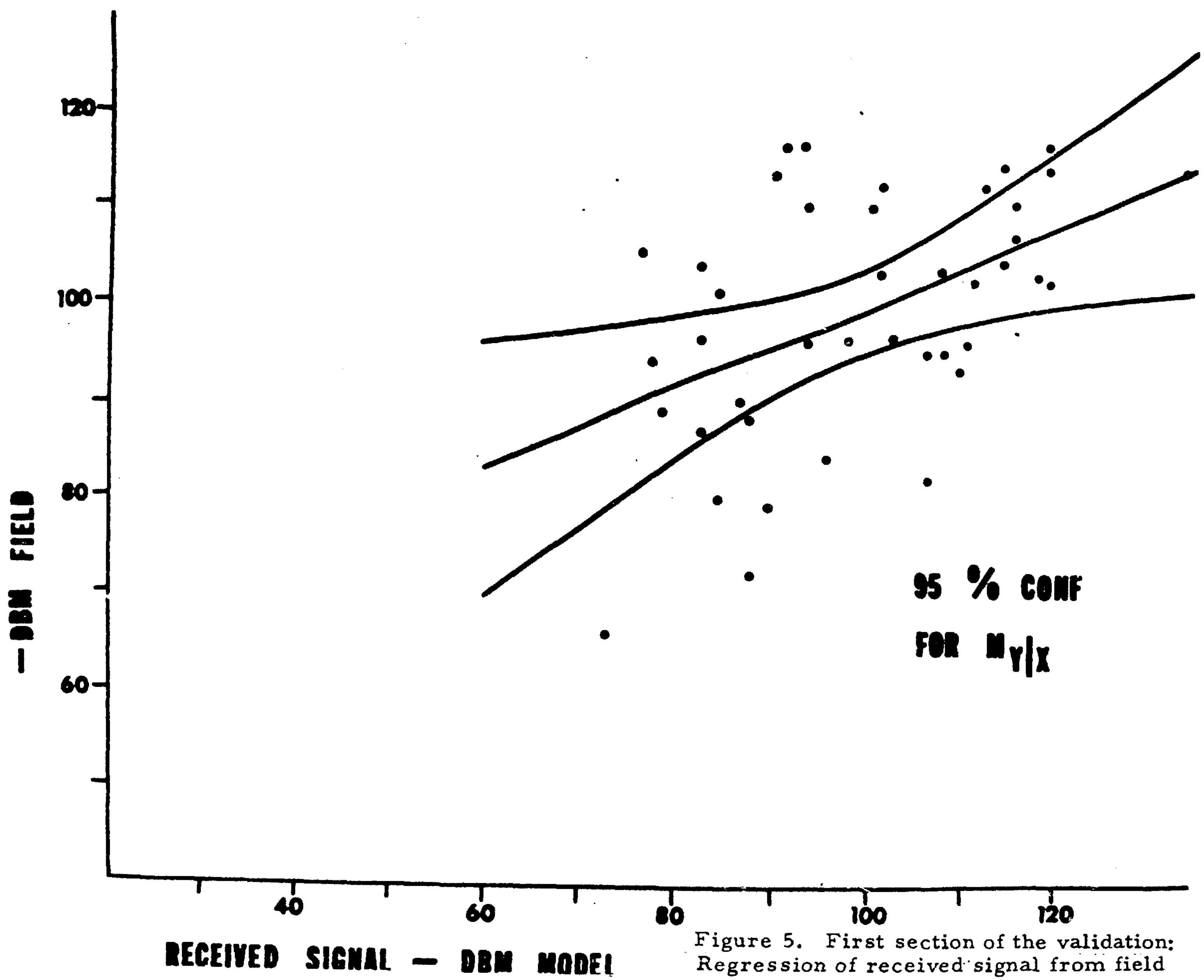


Figure 5. First section of the validation: Regression of received signal from field data against comparable model data.

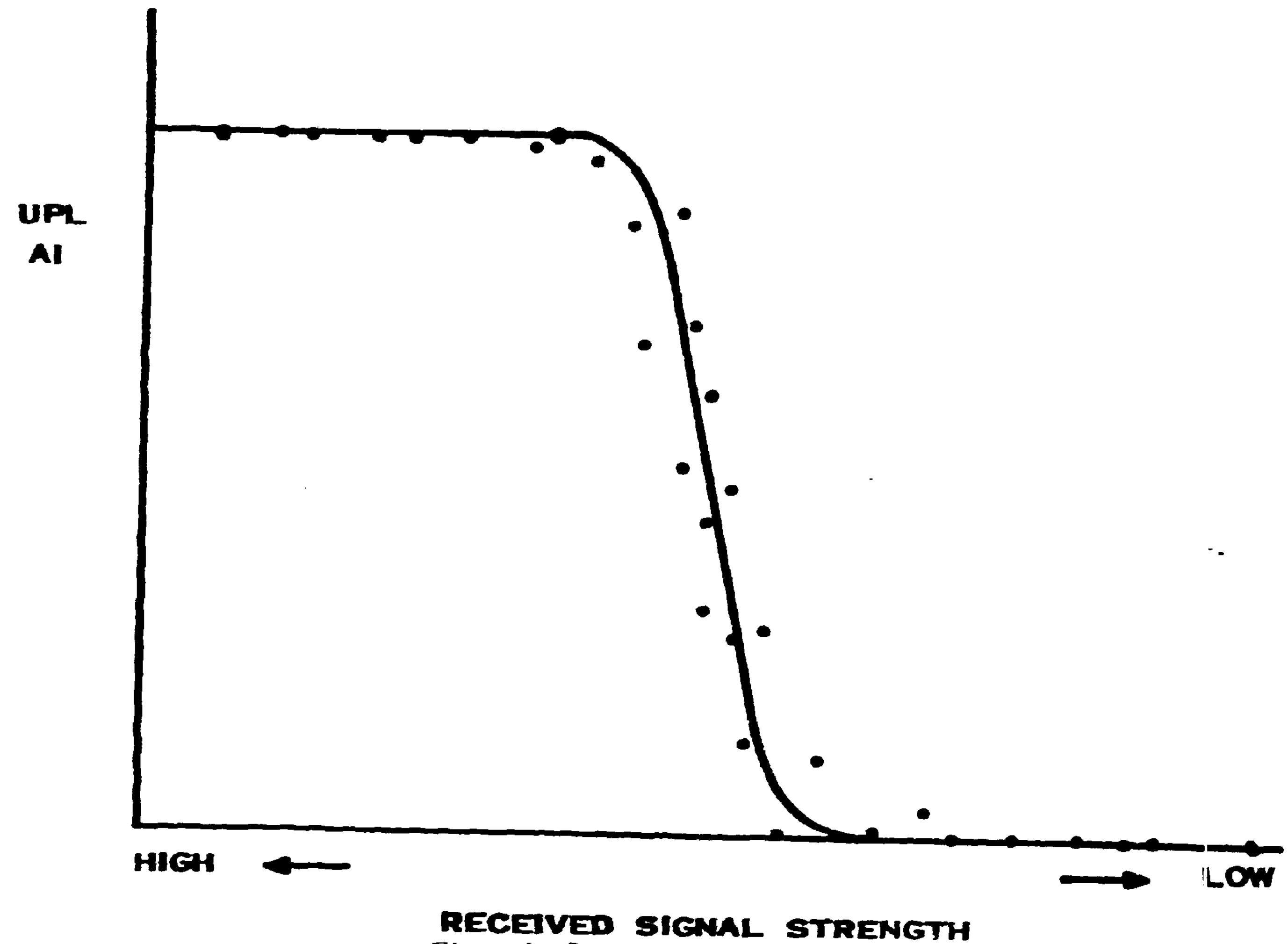


Figure 6. Second section of the validation.
Regression of AI against input signal. Hypothetical data.

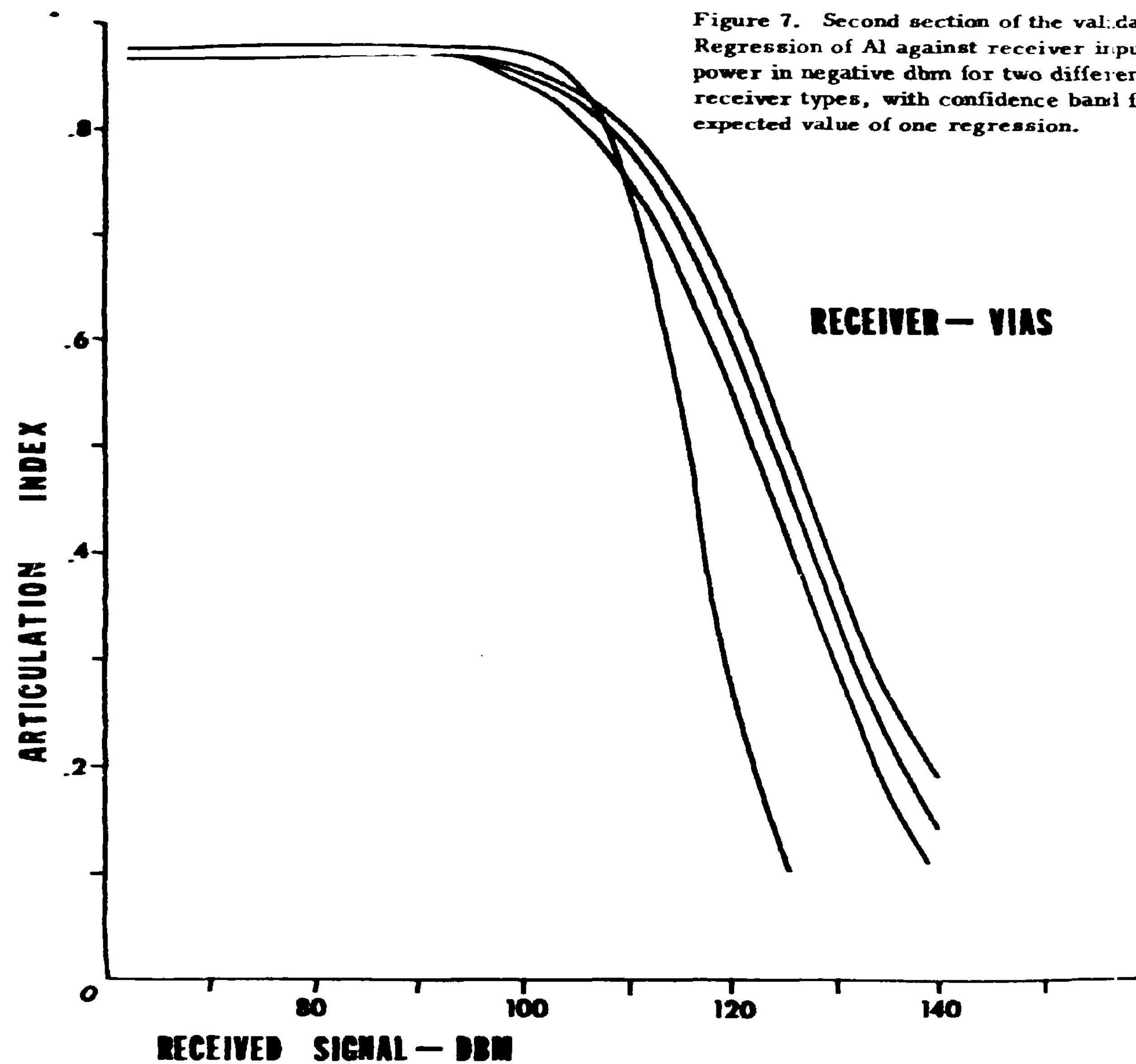


Figure 7. Second section of the validation. Regression of AI against receiver input power in negative dbm for two different receiver types, with confidence band for expected value of one regression.

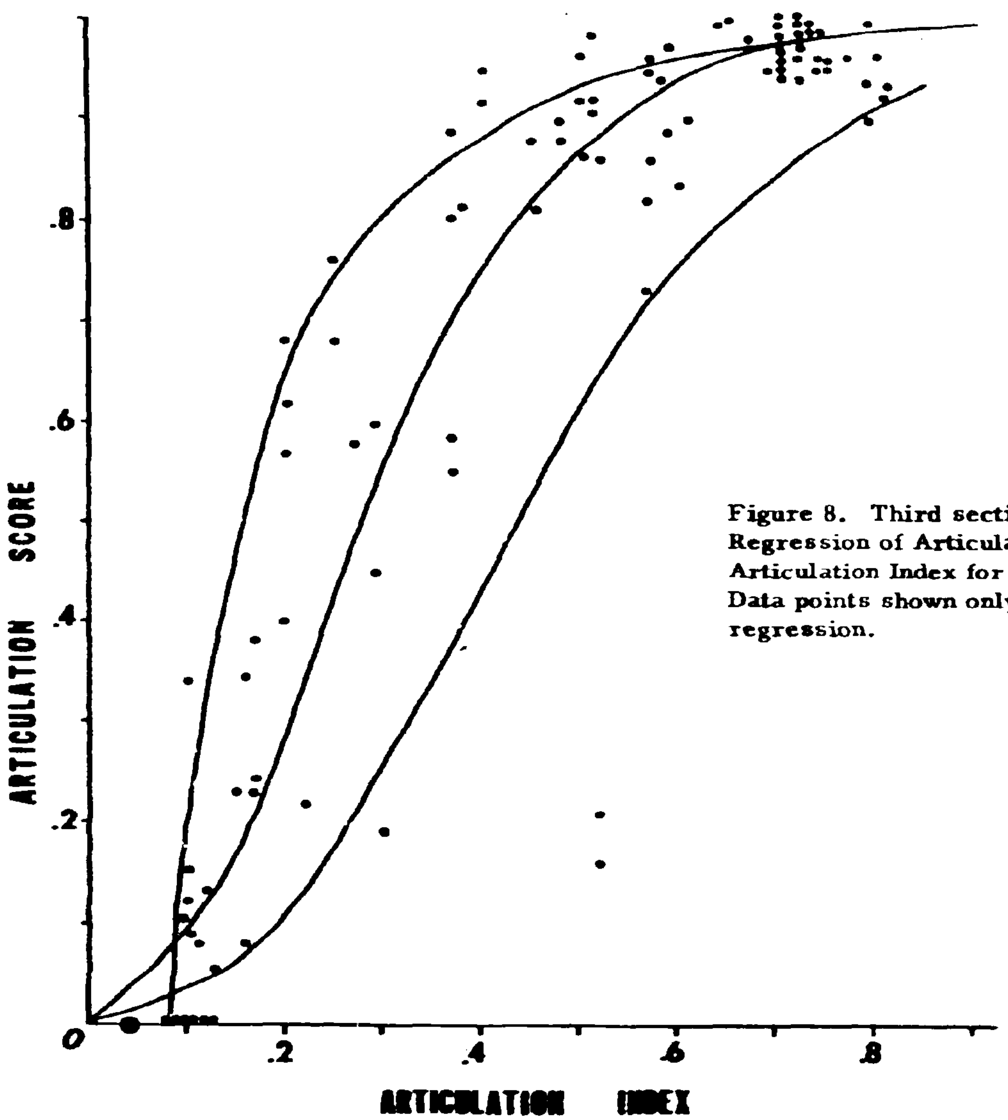


Figure 8. Third section of the validation. Regression of Articulation Score against Articulation Index for three receiver types. Data points shown only for the middle regression.

THE DESIGN OF COMPLEX SENSITIVITY EXPERIMENTS

D. Rothman and J. M. Zimmerman
Rocketdyne, A Division of North American Aviation, Inc.

1. INTRODUCTION. There is a growing tendency among the practitioners of the art of experimental design to allocate more of their efforts to the macroscopic aspects of test planning. This often results in greater benefit than that obtained from intensive improvement of isolated experimental segments. Very little work of this kind has been carried out for sensitivity experiments, however, despite the long history of statistical effort in this field, probably for two reasons. First, most of the major laboratories conducting sensitivity experiments have established over the years their own traditional set of test procedures which are relatively insensitive to variations in experimental objectives. Secondly, the majority of sensitivity experiments have been somewhat restricted in scope, being limited to such purposes as material screening or comparison of properties with those of a standard, and have not usually required extensive experimental planning and expenditures.

Recently it has become clear to many practitioners that there are several newer methods for the design and analysis of sensitivity experiments which deserve more substantial attention, partly because of their intrinsic merit and partly due to the increased complexity and cost of some current programs. It was in connection with one such program that the methods described in this paper were developed, although a substantial portion of the material had been previously formulated under a NASA, MSFC research contract, NAS 8-11061, monitored by Dr. John B. Gayle.

2. FORMULATION OF THE PROBLEM. Consider a sensitivity experiment in which there are n stimulus variables, $x_1, x_2, \ldots, x_n$, and for which the cost for each test is at least approximately known as a function of any combination of these variables. For simplicity, we assume that this cost is no different if the test response is positive (1) or null (0). Given α, suppose that the goal of the experiment is to estimate a specified portion of that $n-1$ dimensional surface S_α on which the probability of a positive response, $M(x_1, \ldots, x_n)$, equals α. Our analysis will be based on a loss function, L, which is made up conceptually of two terms: the cost of tolerating a specified variance in the

estimate of S_α, and the cost of testing. The overall problem is then to find that experimental design which minimizes $\bar{L}$, the value of L averaged over those portions of S_α which are of interest.

The treatment of the problem in this general form requires a carefully worked out technique for the design and analysis of multivariate sensitivity experiments which is readily amenable to the introduction of cost considerations. Although some algorithms for the design of multivariate sensitivity experiments have recently been developed (references 1 and 2), they are extremely complex and do not lend themselves to the implementation of loss minimization. Therefore, a simplification in the structure of the problem is required.

Towards this end, we replace the original multiple stimulus-variable problem by a hybrid regression-sensitivity problem in the following way. We select n-1 of the stimulus variables and consider them as independent variables in a regression model. The remaining variable (say the n^{th}) is considered as a stimulus with a possibly different response function at each combination of the n-1 regression variables. Effectively what we are doing here is replacing the n-variate response function $M(x_1, \ldots, x_n)$ by a univariate function $M(x_n; x_1, \ldots, x_{n-1})$ with parameters $x_1, \ldots, x_{n-1}$. Our program will be to estimate, at a set of specified values of these parameters, that value x_n^α of x_n for which $M(x_n; x_1, \ldots, x_{n-1}) = \alpha$; each point $(x_1, x_2, \ldots, x_{n-1}, x_n^\alpha)$ is in fact on S_α. Then we shall describe the effect of the parameters $x_1, \ldots, x_{n-1}$ by means of an ordinary regression of these variables on the estimates of x_n^α.

In a particular problem, the selection of the single stimulus variable from the original set is usually obvious, being dictated by the nature of the experimental apparatus, preparation of the test specimens, and long standing practice (e.g., in drop tests involving several environmental variables, such as temperature, orientation of the specimen, etc., the height of the impactor would invariably be the single stimulus chosen). In the present case, another important consideration which may affect the choice of the stimulus variable is the relative influence it has on the cost of testing. Our optimization procedure will be based only on the regression variables; that is, we determine the best combinations of the variables $x_1, \ldots, x_{n-1}$ at which to test in order to minimize the loss function of the entire experiment. This macroscopic type of optimization does not itself dictate the local or microscopic design for the stimulus variable (x_n) at each of the regression parameter level combinations. Thus it is important in applying this method to select as the stimulus variable one which affects the cost of testing as little as possible.

It should be pointed out that this general approach to reducing the complexity of the problem is not new. For example, in 1961 Grant and Van Dolah described a procedure for handling multidimensional problems by the use of factorial designs combined with the simple up and down method (reference 3). In our work, however, the aspect of cost minimization has been added, and in addition a quantitative method for describing the efficiency of sensitivity experiments is developed. We treat these two topics in the following sections.

3. MACROSCOPIC COST OPTIMIZATION. The regression model relating the n-1 variables $x_1, \ldots, x_{n-1}$ with the estimates of x_n^a will be written in the form

$$x_n^a = P_o(\underline{x}) + P_1(\underline{x}) + \cdots + P_r(\underline{x}) + \epsilon, \tag{1}$$

where $\underline{x}$ is the vector $(x_1, \ldots, x_{n-1})$, $P_j(\underline{x})$ is a sum of terms of the j^{th} degree in the components of $\underline{x}$ with unknown coefficients, and ϵ is a normally distributed random variable with mean zero and (unknown) variance σ^2. Let N be the number of (not necessarily distinct) values of $\underline{x}$ at which test sequences on the stimulus variable x_n are to be run. The covariance matrix, Q, of the estimates of the coefficients in (1) can be written in the form

$$Q = \sigma^2 R(\underline{x})/N \tag{2}$$

where R is a matrix, independent of σ and N, whose elements involve averages of the components of $\underline{x}$ over the design. In treating particular problems, one determines the optimum proportions of tests to be conducted at each of a certain fixed number of optimum treatment combinations, with N specifying the number by which these proportions are multiplied to obtain an actual design.

We have assumed that the average cost per test depends only on the vector $\underline{x}$ and not on x_n or N (it would depend on N if, for example, there were a setup cost). Let this cost be denoted by $C(\underline{x})$. For the moment we suppose that it is desired to obtain estimates of the function $x_n^{\alpha}(x_1, \ldots, x_{n-1})$ over an a priori specified region U with weighting function $W(\underline{u})$. Our loss function is a linear combination of the weighted average of the prediction variance over this region and the cost of testing. Thus the average loss is

$$\bar{L} = AN^{-1}\sigma^2 \int_U \underline{V}'(\underline{u})R(\underline{x})\underline{V}(\underline{u})W(\underline{u})d\underline{u} + BC(\underline{x})\,N \tag{3}$$

where $\underline{u} = (u_1, \ldots, u_{n-1})$, A and B are appropriately chosen constants, and $\underline{V}$ is a column vector whose components are the linearly independent functions of the components of $\underline{x}$ contained in the quantities $P_j(\underline{x})$, $j = 0, 1, \ldots, r$ of equation (1). For example, in the very simple case when $\underline{x}$ is the scalar u and $r = 2$, we have

$$\underline{V} = \begin{pmatrix} 1 \\ u \\ u^2 \end{pmatrix} . \tag{4}$$

In this situation we have explicitly

$$Q = \sigma^2 \begin{pmatrix} N & \Sigma x & \Sigma x^2 \\ \Sigma x & \Sigma x^2 & \Sigma x^3 \\ \Sigma x^2 & \Sigma x^3 & \Sigma x^4 \end{pmatrix}^{-1} , \tag{5}$$

$$R = \begin{pmatrix} 1 & \overline{x} & \overline{x^2} \\ \overline{x} & \overline{x^2} & \overline{x^3} \\ \overline{x^2} & \overline{x^3} & \overline{x^4} \end{pmatrix}^{-1} ,$$

where $\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$, etc. In the case when N (or the total cost) is not specified in advance we must find that value of N which minimizes (3). Since $R(\underline{x})$ is independent of N, on differentiating this expression one obtains

$$(6) \qquad N_{opt} = \sigma \sqrt{A \int_U \underline{V}'(\underline{u}) R(\underline{x}) \underline{V}(\underline{u}) W(\underline{u}) d\underline{u} / B C(\underline{x})}$$

and the associated value of $\overline{L}$ is

$$(7) \qquad \overline{L} = 2\sigma \sqrt{AB \int_U \underline{V}'(\underline{u}) R(\underline{x}) \underline{V}(\underline{u}) W(\underline{u}) d\underline{u} \cdot C(\underline{x})} \ .$$

Thus, independently of the values of σ, A, and B, it is sufficient to minimize

$$(8) \qquad \overline{L}^2 / 4\sigma^2 AB = C(\underline{x}) \int_U \underline{V}'(\underline{u}) R(\underline{x}) \underline{V}(\underline{u}) W(\underline{u}) d\underline{u}$$

where the right hand member of (8) is proportional to the cost times the average prediction variance or "cost per unit of information". Note that this latter type of loss minimization may be accomplished independently of N and of the cost per unit variance ratio B/A. The value of $\sigma^2 A/B$ is explicitly required only if it is desired to determine N_{opt} from (6). If the total maximum expenditure of the test program is fixed in advance, as is often the case, then N is fixed and the values of σ, A, and B do not affect the minimization of the right member of (8).

When the region U over which the prediction variance is averaged is not specified a priori, the practical solution of the problem becomes more difficult. For example, it may be of interest in some problems to minimize the loss under the circumstances when an estimate of the value of $x_n^a(x_1, \ldots, x_{n-1})$ is to be made by extrapolation to a specified value of x_n^a, rather than to a given value of $\underline{x} = (x_1, \ldots, x_{n-1})$. In such cases, it is generally not possible to formulate the loss function explicitly in as simple a form as we have done since the coefficients in the model (1) are not known in advance. In this situation one may guess at the values of $\underline{x}$ at which the extrapolation is to be made and perform the optimization for a few such possibilities, or, alternatively, a formal Bayesian viewpoint can be taken, an a priori distribution of the extrapolation point made, and the optimization carried out formally in terms of this distribution. We will not pursue this more difficult version of the problem here, although it occurs not infrequently in practice.

When $r = 1$, and the form of the regression and cost models are simple, it is possible to carry out the minimization of (3) in closed form. However, the explicit optimum values of $\underline{x}$ are not always determined by this procedure. For example, we have shown (see reference 4) that when $r = 1$, there are p regression variables, and the cost function C is quadratic, then all that is specified by the minimization of (8) are the means and covariance matrix of the design variables. That is, the minimization of (8) provides

$$p\,\frac{(p+1)}{2} + p = \frac{(p+1)\,(p+2)}{2} - 1*$$

constraints which the optimum design must satisfy. Now generally $k(p+1)-1$ constraints are required to define uniquely a design consisting of k distinct points. When $r = 1$, $p = 1$, for quadratic cost, we obtain $\frac{2.3}{2} - 1 = 2$ constraints; this is one short of the $2(2)-1 = 3$ required for a unique two-point design. When $r = 1$, $p = 2$, we obtain 5 constraints; since three points are required to fit this model, thus requiring $3(3)-1 = 8$ constraints, we get a family of optimum three-point designs with three degrees of freedom.

*Alternatively we may say that, for quadratic cost, the number of distinct elements in the cross product matrix for the design, less one, gives the number of constraints obtained from minimization of (8).

When the cost function is made up of functions of components of $\underline{x}$ which do not already appear in the cross product matrix then one obtains an additional constraint from the minimization of (8). In fact, in the general case of any r and p we have made the following conjecture.

Conjecture: Let m be the number of distinct elements in the cross product matrix, P, corresponding to the polynomial model (1), of degree r. Suppose the cost function C contains functions of the components of the p(= n-1) dimensional vector $\underline{x}$ which do not appear in P (we refer to this as condition I). Then minimization of the loss function (8) yields m constraints for the determination of the optimum design. If the cost only contains functions already appearing in P (condition II) then minimization of (8) provides m-1 constraints.

Since a design of k distinct "points" or treatment combinations requires k(p+1)-1 constraints for unique determination we have immediately the following:

Corollary: Minimization of (8) results in an optimum design consisting of

$$\max \left\{ q, \;] \frac{m+1}{p+1} [\right\}$$ points when condition I prevails and

$$\max \left\{ q, \;] \frac{m}{p+1} [\right\}$$ points when condition II prevails,

where $q = \Sigma_{s=0}^{s_o} \binom{r}{s} \binom{p}{s}$, $s_o = \min\{r, p\}$; q = number of unknown parameters in the model, and] y [denotes the smallest integer larger than or equal to y. The design is unique when the quantity in brackets is an integer. A formal proof of this conjecture may require solving the general minimization problem for (8), a very formidable task. Even the case r = 1 poses serious difficulties (see references 4 and 5). Apart from our verification of the conjecture in the linear case when condition II prevails (reference 5), we have recently solved a particular problem (using a computer search procedure) when U is a single point, r = 2, p = 1, and the cost function is exponential for all stimulus levels above a specified value. A unique three-point optimum design was found. Applying the conjecture and corollary (with condition I), we find in this case that the cross product matrix contains five distinct elements so that indeed five constraints are obtained determining a unique three-point optimum design.

In implementing this conjecture it is convenient to have the explicit relation between m, r and p. For example, for small values of the latter we have the following table.

p \ r	1	2	3
1	2	4	6
2	5	14	27
3	9	34	83

Values of m-1

$$m = \Sigma_{j=0}^{J} \binom{2r}{j} \binom{p}{j}, \quad J = \min(2r, p)$$

Thus, for example, if the regression model is cubic in three variables and condition II prevails, one would expect to find a unique 21-point optimum design. Note that in this case q = 20, so that the number of required points is greater than the number of unknown parameters.

Despite the formidable nature of an explicit closed form minimization of (8) in the general case, numerical minimization procedure may not require excessive effort. For example, the recently conducted study referred to above (r = 2, p = 1) only took a few minutes to run on an IBM 7094 computer.

4. <u>BLOCKING OF THE TESTS AND THE GROWTH OF INFORMATION</u>. Suppose we have obtained an optimum k-point design by the methods outlined above. The order in which these groups of tests are to be conducted is usually dictated by specific characteristics of the particular program. Generally the "least expensive" treatment combination or point (from the point of view of $C(\underline{x})$) will be explored first, then the next, and so on until the most expensive point is arrived at. We will not consider this question further here, but next turn our attention to the design of the individual group of sensitivity tests at each of the, say, q optimum treatment combinations of the regression variables.

Sensitivity experiments are most efficient when they are purely sequential, since in this situation one can reflect carefully on all previous results before selecting the next test level for the stimulus variable. But if the experimenter is required for reasons of economy or manufacturing time limitations to order batches of test materials with specified (not necessarily equal) stimulus levels (as, for example, in solid propellant critical diameter studies), then it is necessary to consider the question of "block-sequential" sensitivity experiments and to evaluate the expected loss of information implicit in this mode of operation relative to the usual purely sequential test procedure.

To discuss block-sequential designs we will require a characterization of the amount of information available before the entire group of tests is conducted (from previous studies, etc.). This prior information will be expressed as that number of equivalent asymptotically optimal tests which would provide the same asymptotic information. Our approach will be based on the use of asymptotic expressions to characterize the growth of information in sensitivity experiments. Attention will be limited to the case in which the response function is a normal cdf, and to simplify the calculations we will assume that the sole aim of the tests is to estimate the median critical stress level (i.e., $\alpha = 50\%$). Our analysis will be carried out without actually specifying the test levels to be used in each of the blocks, although it is known (see reference 6) that for this type of experiment any test sequence converging to the median is asymptotically optimal in terms of efficiency in estimating the median. Evaluation of the validity of the asymptotic theory for small sample size is currently being studied by means of simulation.

Efficiency and Growth of Information. Suppose we have a cumulative normal response function with (unknown) parameters μ and σ. Let $\hat{\mu}$ denote the maximum-likelihood estimate of $\hat{\mu}$. Consider a design with T tests whose goal is the estimation of μ. An asymptotic expression for the variance of $\hat{\mu}$ (as $T \to \infty$) is given by (reference 7)

$$\sigma^2_{\hat{\mu}}(T) \sim C_2\sigma^2/(C_0C_2-C_1^2) \tag{9}$$

where y_i = the level of the stimulus variable on the i^{th} test,

$$
\begin{aligned}
t_i &= (y_i - \mu)/\sigma \ , \\
Z_i &= \frac{1}{\sqrt{2\pi}} \ e^{-t_i^2/2} \ , \\
p_i &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t_i} e^{-u^2/2} \, du \ , \\
q_i &= 1 - p_i \ , \\
V_i &= Z_i^2/p_i q_i \ , \\
C_j &= \sum_{i=1}^{T} V_i t_i^j \ .
\end{aligned}
\tag{10}
$$

Since the goal of the experiment at each of the q optimum regression points is the estimation of μ, it is not unreasonable to restrict attention to designs which are asymptotically symmetric with respect to μ. Then

$$C_1 \sim 0 \ ,$$

$$\sigma_{\hat{\mu}}^2 (T) \sim \sigma^2/C_0 \ .$$

It has been shown (references 6 and 7) that $\sigma_{\hat{\mu}}^2 (T)$ is asymptotically minimized when $t_i = 0$, $i = 1, \ldots, T$; this, minimum value is

$$\sigma_{\hat{\mu}}^2 (T) \sim (\pi/2)\sigma^2/T \ .$$

The asymptotic information after T tests, I_T, may be expressed by the reciprocal of the variance of $\hat{\mu}$, or

(11) $$I_T \sim C_0/\sigma^2 = \sum_{i=1}^{T} Z_i^2/p_i q_i \sigma^2 \quad .$$

Thus the information contribution of a test at $t_i = t$ is given by

(12) $$I(t) \sim e^{-t^2}/2\pi pq\sigma^2 \quad .$$

Since this is maximized at $t = 0$, where we have

(13) $$I(0) \sim 2/\pi\sigma^2 \quad ,$$

the efficiency, defined as the relative information of an individual test at stimulus level t, may be written as

(14) $$E(t) = I(t)/I(0) \sim e^{-t^2}/4pq \quad .$$

The function $E(t)$ is tabulated below for selected values:

Table 1

$\|t\|$	0	.1	.2	.5	.75	1.0	1.5	2.0	3.0	4.0	5.0
$E(t)$	1	.9964	.9856	.9127	.8128	.6888	.4226	.2060	.0229	.00089	.000012

It may be noticed that the efficiency declines rapidly in the range $.75 < |t| < 2$. Tests for which $|t| > 3$ are very inefficient in the long run, although they may provide a large fractional increase in information early in the experiment.

In order to derive an expected value for $E(t)$, we express it in a more explicit form. It can be shown that the following expansions are convergent for all values of t:

$$p = 1/2 + t/\sqrt{2\pi} - t^3/6\sqrt{2\pi} + t^5/40\sqrt{2\pi}$$

(15) $$- t^7/336\sqrt{2\pi} + t^9/3456\sqrt{2\pi} - t^{11}/42240\sqrt{2\pi} + t^{13}/599040\sqrt{2\pi} - \ldots,$$

$$q = 1/2 - t/\sqrt{2\pi} + t^3/6\sqrt{2\pi} - t^5/40\sqrt{2\pi}$$

(16) $$+ t^7/336\sqrt{2\pi} - t^9/3456\sqrt{2\pi} + t^{11}/42240\sqrt{2\pi} - t^{13}/599040\sqrt{2\pi} + \ldots .$$

Therefore

$$pq = 1/4 - t^2/2\pi + t^4/6\pi - 7t^6/180\pi + t^8/140\pi$$

(17) $$- 83t^{10}/75600\pi + 73t^{12}/498960\pi - 523t^{14}/30270240\pi + \ldots,$$

and

$$E(t) \sim e^{-t^2}/(1-2t^2/\pi+2t^4/3\pi-7t^6/45\pi+t^8/35\pi$$

(18) $$- 83t^{10}/18900\pi+73t^{12}/124740\pi-523t^{14}/7567560\pi + \ldots) .$$

We have finally

$$e^{t^2}E(t) \sim 1 + 2t^2/\pi - 2t^4/3\pi + 4t^4/\pi^2 + 7t^6/45\pi - 8t^6/3\pi^2$$

$$+ 8t^6/\pi^3 - t^8/35\pi + 16t^8/15\pi^2 - 8t^8/\pi^3 + 16t^8/\pi^4$$

$$+ 82t^{10}/18900\pi - 304t^{10}/945\pi^2 + 68t^{10}/15\pi^3 - 64t^{10}/3\pi^4$$

$$+ 32t^{10}/\pi^5 - 73t^{12}/124740\pi + 1132t^{12}/14175\pi^2$$

$$- 356t^{12}/189\pi^3 + 704t^{12}/45\pi^4 - 160t^{12}/3\pi^5 + 64t^{12}/\pi^6$$

$$+ 532t^{14}/7567560\pi - 296t^{14}/17325\pi^2 + 599t^{14}/945\pi^3$$

(19) $$- 7808t^{14}/945\pi^4 + 48t^{14}/\pi^5 - 128t^{14}/\pi^6 + 128t^{14}/\pi^7 + \ldots .$$

In general, since μ and σ are unknown, we are uncertain as to just which value of t we are testing at; let this uncertainty be represented by a density $f(t)$ with mean M and variance U. Since we are trying to test at $x = \mu$ $(t = 0)$, and since $\hat{\mu}$ is unbiased and asymptotically normal, we have

(20) $$M = 0$$

(21) $$U \equiv \sigma_{\hat{t}}^2 = \sigma_{\hat{\mu}}^2(T)/\sigma^2 .$$

Then the expected test efficiency is given by

$$\overline{E(t)} \equiv \int_{-\infty}^{\infty} E(t)f(t)dt$$

$$\sim \int_{-\infty}^{\infty} \frac{E(t)}{\sqrt{2\pi U}} e^{-t^2/2U} dt$$

$$= \frac{1}{\sqrt{2U+1}} \int_{-\infty}^{\infty} \frac{e^{t^2} E(t)}{\sqrt{2\pi U/(2U+1)}} e^{-t^2/2[U/(2U+1)]} dt$$

(22) $$= \frac{1}{\sqrt{2U+1}} \int_{-\infty}^{\infty} \frac{e^{t^2} E(t)}{\sqrt{2\pi v}} e^{-t^2/2v} dt ,$$

using the substitution $v = U/(2U+1)$. Because of (19) this integral can be thought of as a sum (with coefficients given by (19)) of the even central moments, M_{2n}, of a normal distribution with variance v. We have

$$M_0 = 1 ,$$

(23) $$M_{2n} = (2n-1)vM_{2n-2} ,$$

from which it follows that $M_2 = v$, $M_4 = 3v^2$, $M_6 = 15v^3$, $M_8 = 105v^4$, $M_{10} = 945v^5$, $M_{12} = 10395v^6$, and $M_{14} = 135135v^7$. Then from (22) and (19) we have

$$\overline{E(t)} \sim \left[1 + \frac{2}{\pi}v + \left(\frac{12}{\pi^2} - \frac{2}{\pi}\right)v^2 + \left(\frac{7}{3\pi} - \frac{40}{\pi^2} + \frac{120}{\pi^3}\right)v^3 + \left(\frac{1680}{\pi^4} - \frac{840}{\pi^3} + \frac{112}{\pi^2} - \frac{3}{\pi}\right)v^4 \right.$$

$$+ \left(\frac{83}{20\pi} - \frac{304}{\pi^2} + \frac{4284}{\pi^3} - \frac{20160}{\pi^4} + \frac{30240}{\pi^5}\right)v^5$$

$$+ \left(\frac{665280}{\pi^6} - \frac{554400}{\pi^5} + \frac{162624}{\pi^4} - \frac{19580}{\pi^3} + \frac{12452}{15\pi^2} - \frac{73}{12\pi}\right)v^6$$

$$+ \left(\frac{523}{56\pi} - \frac{11544}{5\pi^2} + \frac{85657}{\pi^3} - \frac{1116544}{\pi^4} + \frac{6486480}{\pi^5} - \frac{17297280}{\pi^6}\right.$$

$$\left. \left. + \frac{17297280}{\pi^7}\right)v^7 + \cdots\right] / \sqrt{2U+1}$$

$$\cong \left[1 + .636620v + .579234v^2 + .560060v^3 + .548604v^4\right.$$

(24) $$\left. + .539890v^5 + .534223v^6 + .530582v^7\right] / \sqrt{2U+1} \quad .$$

This function is tabulated below for selected values:

Table II

U	0.0	0.2	0.5	1.0	1.5	2.0	3.0	4.0	6.0	8.0	10.0	15.0	20.0	40.0	100.0
$\overline{E(t)}$	1.000	.934	.853	.75	.68	.63	.55	.49	.42	.37	.34	.28	.25	.18	.11

For small values of U, the approximation given by the first two terms of the expansion,

(25) $$\overline{E(t)} \approx 1 - .3634U \quad ,$$

is sufficiently accurate. For large values of U we have $v \cong .5$ and numerical evaluation of equation (24) leads directly to

$$\overline{E(t)} \approx 1.132/\sqrt{U} \quad . \tag{26}$$

At this point we have only an asymptotic formula (24) for computing the expected efficiency of a new test, given $\sigma^2_{\hat{\mu}}$. But denoting by E_0 the prior information in terms of equivalent efficient tests, and by E_i the efficiency of the i^{th} test, we have from our definition of the information after T tests that

$$I_T \sim I_{T-1} + 2E_T/\pi\sigma^2 \quad . \tag{27}$$

Then one can show by an elementary induction that

$$I_T \sim \frac{2}{\pi\sigma^2} \sum_{i=0}^{T} E_i \quad , \tag{28}$$

or

$$\sigma^2_{\hat{\mu}}(T) \sim (\pi/2)\sigma^2 / \sum_{i=0}^{T} E_i \quad . \tag{29}$$

Equations (21), (24), and (29) can be used to asymptotically describe the growth of information in sensitivity experiments.

Of these equations, only (21) is exact. Equation (29) is asymptotically valid as $\Sigma_{i=0}^{T} E_i$ goes to infinity, which will happen if and only if E_0 and/or T become arbitrarily large. Equation (24) holds asymptotically on the $j+1^{st}$ test as $\Sigma_{i=0}^{j} E_i$ goes to infinity, in which case U goes to zero. But note that (24) and (29) do not give unreasonable results even for large values of U and for small values of E_0 and T. Thus we shall attempt to draw tentative conclusions even in the latter cases.

In the following example we see how slowly the individual test efficiencies increase in the course of a purely sequential sensitivity experiment. Note that we never have to specify the individual test levels in this line of reasoning.

<u>Example</u>. If $E_0 = .10$, then from (29), $\sigma^2_{\hat{\mu}}(0) \sim 15.71\sigma^2$. From (21), $U \sim 15.71$, and from (24), $\bar{\bar{E}}_1 \sim .275$. Continuing in this manner, we have $\sigma^2_{\hat{\mu}}(1) \sim 4.19\sigma^2$,

$$U \sim 4.19, \quad \bar{\bar{E}}_2 \sim .486,$$

$$U \sim 1.82, \quad \bar{\bar{E}}_3 \sim .647,$$

$$U \sim 1.042, \quad \bar{\bar{E}}_4 \sim .748,$$

$$U \sim .696, \quad \bar{\bar{E}}_5 \sim .811,$$

$$U \sim .512, \quad \bar{\bar{E}}_6 \sim .851,$$

$$U \sim .401, \quad \bar{\bar{E}}_7 \sim .878,$$

$$U \sim .328, \quad \bar{\bar{E}}_8 \sim .899,$$

$$U \sim .276, \quad \bar{\bar{E}}_9 \sim .911, \text{ etc.}$$

The above asymptotic theory has been tested by means of a computer program for simulating sensitivity experiments. The value of $\sigma^2_{\hat{\mu}}$ given by this theory is much more realistic than the value $(\pi/2)\sigma^2/T$, but still sometimes conservative by a factor of three.

<u>Block-Sequential Designs</u>. Now let us introduce the notion of a "block-sequential" design, in which each block of tests is planned after all previous test results have been analyzed. To "stage" such an experiment means to assign sample sizes to each of a given number of blocks, given the total sample size T. The "optimum" staging of a block-sequential sensitivity experiment is that staging which produces the greatest expected gain in information. Using the asymptotic methodology derived above, we

have computed a table of optimum stagins for 2-block sensitivity experiments for total sample sizes up to 34 and for two different amounts of prior information.

Table III

Sample Size of First Block		1	2	3	4	5	6	7	8
Total Sample Sizes for which the Given First Block is Optimum	$E_0=.02$	2-3	4-7	8-11	12-15	16-19	20-24	25-29	30-34
	$E_0=2.00$	2-3	4-6	7-10	11-14	15-19	20-25	26-31	32-34

For example, if 13 tests permitted were specified at the particular combination of regression variables under consideration, the optimum 2-block design would call for 4 tests in the first block and 9 tests in the second, over the given range of values of E_0.

It is fortunate that the above results are relatively independent of E_0, because this parameter is in practice very difficult to evaluate. For example, if our prior density on μ is uniform in [A, B], then

$$\sigma^2_{\hat{\mu}}(0) = (B-A)^2/12 .$$

But to compute

$$E_0 = (\pi/2)\sigma^2/\sigma^2_{\hat{\mu}}(0) ,$$

we must know σ^2, and such information is almost always unavailable.

Results of the type given in Table III are not completely rigorous even for large values of E_0 and/or T, since we compute expected information in the second stage as a function of expected information in the first stage, rather than in terms of the distribution of this information. But the results are all plausible and of practical value precisely because they are similar for E_0 = .02 and E_0 = 2.00. In addition, the optimum block sizes are obviously right for a total sample size of two, and the fraction in the first block decreases relatively smoothly as the total sample size increases.

It should be noted in passing that the above machinery permits us for the first time to characterise experiments in which the stress variable has an independent "setting" error, such as the projectile velocity in projectile penetration tests. Let this setting error be normally distributed with mean 0 and variance σ_s^2. Then the only change in the above formulae is in the expression for U, which is now

$$U = (\sigma_{\hat{\mu}}^2 + \sigma_s^2)/\sigma^2 .$$

Example: Let $\sigma_s^2 = 2\sigma^2$. Then the asymptotic efficiency of even a purely sequential design in this case is only 63%, since $U \sim 2$ (see Table II).

ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance of Dr. S. R. Webb in formulating the conjecture of Section 3. The computer simulation program was coded by P. M. Hait and B. J. Byars.

REFERENCES

[1] Madeline J. Alexander, "Multivariate Sensitivity Experiments with Non-Interacting Stimuu," Rocketdyne Research Report RR 64-15, July 1964.

[2] Madeline J. Alexander, "Models for the Analysis of Multivariate Experiments with Interacting Stimuli," Rocketdyne Research Memorandum RM 1121-351, 22 February 1965.

[3] R. L. Grant and R. W. Van Dolah, "Use of the Up-and-Down Method with Factorial Designs," Proceedings of the Seventh Conference on the Design of Experiments in Army Research, Development, and Testing, ARODR 62-2, pp. 39-65.

[4] "Statistical Design of Complex Experimental Programs, Final Report, I. Optimum Experimental Designs Obtained by Minimizing a Loss Function," Rocketdyne Report R-3392-1, March 1962.

[5] Madeline J. Alexander, N. R. Goodman, M. O. Locks, and S. R. Webb, "Some Results on Designs for Regression Experiments, Design of Experiments with Autocorrelated Errors Present, and Decision Theory Approach to Complex Experimentation," Aeronautical Research Laboratories Report ARL 63-107, June 1963.

[6] H. Chernoff, "Optimal Design of Experiments," Technical Report No. 82, Applied Mathematics and Statistics Labs, Stanford University, 31 October 1962, pp. 2-3, 8-15.

[7] D. Rothman, "The Inverse Response Problem for the Cumulative Normal Response Function, with Application to the Safety Problem," Rocketdyne Research Memorandum RM 1057-351, 28 July 1964.

FACTORS AFFECTING SENSITIVITY TESTING

James R. Kniss, and Warren S. Wenger
Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

INTRODUCTION. Sensitivity testing is frequently utilized by the army in evaluating the sensitivity and consequently the reliability of percussion primers. Since such primers are used rather extensively in nuclear warheads, missiles and conventional munition systems, their functioning characteristics frequently have an important bearing on the reliability of these systems. However, knowledge of these characteristics of the primers and consequently of their reliability under varying temperature and impact conditions is often rather limited.

In a recent study involving the reliability of a nuclear weapon system, the effect of temperature and firing pin impact velocity on the reliability of initiation of a primer became of interest. This problem arose as a result of the procedures currently being used to test the system in which a particular type of primer is used. These procedures do not include a test of the primer itself, but the firing pin which initiates the primer is tested to determine whether the kinetic energy produced equals or exceeds a specified level of kinetic energy. The test results thus far obtained indicate that the level of kinetic energy specified is not compatible with the sensitivity of the primers. For, although a considerable number of firing pins have failed to produce the specified level of kinetic energy, in subsequent tests none of them has failed to fire a primer. Furthermore, the results of the primer testing that has been done indicate that the required kinetic energy is dependent upon the temperature of the primer. It has also been suggested that the sensitivity of the primers might not be a function of kinetic energy alone, but might also be a function of the impact velocity of the firing pin. If this relationship does exist then any primer test fixture should be designed to simulate the stroke velocity of the firing pin normally used to detonate that type of primer.

As a result of the questions that arose from this testing problem, a test was designed that would (1) measure the sensitivity of the primers under standard conditions, (2) determine the effect of strike velocity upon the kinetic energy required to function the primer, and (3) determine the effect of temperature upon the kinetic energy required to function the primer.

DISCUSSION OF TEST DESIGN. The test was to be conducted using the Bruceton Up-and-Down method of sensitivity testing. This method has been used for years, primarily in evaluating the quality or changes in quality of conventional primers. However, no record could be found of any test conducted where the strike velocity and temperature effects were investigated. Most past tests were conducted at ambient temperature and a single weight ball was dropped throughout. The height at which 50% of the primers would function was estimated along with the variability in this height. The results were used primarily to detect trends due to age and to detect lot-to-lot variability.

For this test the conventional procedures were modified in that four different weight balls and four different conditioning temperatures were used. A conventional type primer (the MK2A4) was used since the primer in question was not available and the MK2A4 is of a similar type. The primers were tested according to the following design:

		Ball Weight (oz.)			
		4	8	12	16
	70	X_{111}	X_{121}	X_{131}	X_{141}
		X_{112}	X_{122}	X_{132}	X_{142}
	25	X_{211}	.	.	.
		X_{212}	.	.	.
Temp. (°F)	-20	X_{311}	.	.	.
		X_{312}	.	.	.
	-65	X_{411}	.	.	X_{441}
		X_{412}	.	.	X_{442}

Where the x's represent the drop height at which 50% of the kth sample of primers conditioned at the ith temperature and using the jth ball weight will function. However, it is obvious that the drop height will be affected by ball weight; and, of course, we are not interested in this obvious relationship.

In order to obtain the relationship in which we are interested, it will be necessary to convert these drop heights to kinetic energy using the following transformation

$$y_{ijk} = w_j x_{ijk}$$

where w_j is the ball weight in ounces and y_{ijk} is kinetic energy in inch-ounces. The kinetic energy obtained will be a function of the sensitivity of the primer and should be unaffected by the strike velocity or temperature if these factors are indeed insignificant.

It is, therefore, possible to hypothesize that if kinetic energy is the only factor affecting the sensitivity of the primer the analysis of the data should reveal no significant effects due to either temperature or ball weight. Should ball weight affect the required energy, it could be further hypothesized that this difference is due to the impact velocity of the firing pin.

At first glance the above design would suggest that a simple two-way classification of variables should be performed. However, it was suspected (and later confirmed) that the homogeniety of variances assumption which is necessary for this type of analysis might not be met.

This lack of homogeniety becomes intuitively obvious when it is considered that the change in kinetic energy per unit change in drop height is greater for the heavy ball. This would imply that the lighter balls yield better estimates of drop heights and consequently the variability associated with such estimates will be smaller for lighter balls.

If the data were analyzed as it is, erroneous results might be obtained as to the significance of the main effects as well as of any interaction that might exist between ball weight and temperature.

It was, therefore, planned to break the results down and first work with ball weight vs. kinetic energy at each level of temperature.

This, of course, does not solve the problem of the lack of homogeniety, and it is necessary to correct for this condition before progressing further. This may be accomplished by computing the within cell variation for each ball weight and attempting to obtain the standard deviation as a function of required kinetic energy over columns, (ball weights) i.e.:

if we can obtain a relation $\sigma = f(\mu)$

then $\int \frac{1}{f(\mu)} \cdot d\mu$ is an appropriate

transformation that will transform the data so that the variability will be independent of the ball weight.

It is now possible to determine the relationship of ball weight to kinetic energy through the use of a simple least square analysis performed on the transformed data. However, it should be understood that the transformed data should be used only for purposes of significance testing and that the actual relationships should be represented by the un-transformed data.

It must be determined whether this relationship differs for each or any of the temperatures. If it does differ, is the difference only in intercept, only in slope, or in both intercept and slope? If it differs only in intercept the differences are constants over ball weight and all the data may be corrected back to ambient temperature so that a final relationship may be represented at ambient temperature (or at any other temperature within the range of the test that may be of interest). If the relationship differs in slope, an interaction between temperature and ball weight is indicated, and the relationship will necessarily be represented for each temperature or range of temperatures over which the slopes are homogeneous.

In any case a final representation of required kinetic energy (to function 50% of the primers) will be obtained and will be of the form $E = a + bw + cw^2 + \ldots$ (since a simple least square fit is being used). However, the relationship of kinetic energy (E) to strike velocity is of primary interest; and, therefore, the ball weight (w) in the above equation must be converted to strike velocity (v).

This may easily be done since a direct relationship exists between weight and velocity for any free falling body (neglecting air resistance, etc.). For example, assuming a linear relationship between required energy and ball weight, the relationship between required energy and strike velocity may be obtained as follows:

Given the relation:

(1) $E = a + bw$

and the equations for free falling bodies:

(2) $S = (1/2)gt^2$

(3) $V = gt$

(4) $E = WS$

where E = Energy (in.-oz.)

a, b = constants

S = Drop height (in.)

g = acceleration of gravity (384 in./sec.^{2})

t = time (sec.)

V = Strick Velocity (in./sec.)

Solving equation (2) for t: $t = \sqrt{2S/g}$ substituting this value for t in (3):

$V = gt = g\sqrt{2S/g} = \sqrt{2gS}$

then substituting for S: $S = E/W$ gives

(5) $V = \sqrt{2gE/W}$

and solving (5) for W: $W = 2gE/V^2$

finally substituting in (1) and solving for E

$E = a + b\,(2gE/V^2)$

gives

$$(6)\quad E = \frac{a V^2}{V^2 - 2bg}, \text{ the desired relation .}$$

The final relationship of kinetic energy to temperature may be obtained in a similar manner. However, in this case there is no reason to believe that the variances will not be homogeneous.

DISCUSSION OF TEST RESULTS. In order to conduct this test, 32 samples of MK2A4 primers were fired, each sample being comprised of 40 primers. Using the Bruceton up and down method, the 32 estimates of 50% points were as follows for each combination of temperature and ball weight used:

50% Points (in inches)
Computed from Up and Down Tests*
of Primer, Percussion, Mk2A4

		Approximate Ball Weight (oz)			
		4	8	12	16
Temp. (°F)	70°	8.7500	5.3289	3.1375	2.4375
		9.4167	5.4250	3.7829	3.0329
	25°	9.7361	4.8289	3.5461	2.6118
		9.4342	5.2500	3.6591	3.1125
	-20°	10.0000	5.5395	3.5417	2.7875
		9.5000	5.6310	3.4934	3.1125
	-65°	9.8000	5.4868	3.9500	3.1000
		10.0109	5.5500	3.6500	3.4539

*A sample of 40 primers was used for each of the 32 tests.

Obviously there is a correlation between ball weight and the 50% points of drop height. But, of course, this is not very useful.

To get a meaningful basis for comparison, these heights were converted to the equivalent values of kinetic energy by multiplying each height by the corresponding ball weight (exact). Therefore, all further analyses were performed using the following values of kinetic energy:

Kinetic Energy (in. -oz.)
for above 50% Points

		Approximate Ball Weight (oz)			
		4	8	12	16
Temp. (°F)	70°	34.5562	42.4163	39.6570	39.7800
		37.1892	43.1813	47.8146	49.4969
	25°	38.4506	38.4365	44.8216	42.6246
		37.2583	41.7883	46.2498	50.7960
	-20°	39.4928	44.0926	44.7659	45.4920
		37.5182	44.8210	44.1555	50.7960
	-65°	38.7029	43.6732	49.9267	50.5920
		39.5358	44.1762	46.1348	56.3676

Since we suspected that the variances within ball weights might not be homogeneous, the individual cell variances were calculated and tested for homogeneity. This test not only confirmed our suspicions, but indicated a rahter acute case of non-homogeneity.

Further investigation showed that the relation between the standard deviation and ball weight could be satisfactorily represented by a function of the form $\sigma = a + bx$. And the required transformation, to correct for the observed non-homogeneity was found to be $y = 2.63 \ln(-13.8 + .38x)$ i.e. by substituting the E_{ijk} above for x in this equation we obtained y_{ijk} with homogeneous variances.

Having obtained these y's, we could then proceed to "determine" the relationship between ball weight and kinetic energy for each of the four test temperatures. A graphical representation of these relationships, together with the data from which they were derived, is given in Figure I.

The dots, of course, represent the data points and the lines the linear relationship derived from these points using least squares methods. Tests, using the transformed data (variances homogeneous), showed the slopes of these lines to be significantly different from zero, i.e., the 50% points of kinetic energy are a function of the ball weight used.

Visual comparison (see Figure II) indicated and a test confirmed that the slopes of these lines did not differ significantly, i.e., there was no reason to believe that the difference in the 50% point resulting from a given difference in ball weight varied with temperature. Or, to state it differently, the results of our analysis did not contradict the hypothesis that the difference in the 50% point resulting from a given difference in ball weight is independent of temperature.

If we accept this hypothesis, it follows that a better estimate of the effect of ball weight on the 50% point of kinetic energy should be obtained by "correcting" the data for temperature and then using the resulting (32) points to obtain a single relationship. This was done, and we obtained the equation K.E. = 33.51 + .81W as our best estimate of the relationship between kinetic energy and ball weight at ambient temperature (see Figure III).

We then obtained the desired relationship, between Strike Velocity and Kinetic Energy, by inserting the values of a and b from the above equation (a = 33.51, b = 0.81) into equation (6). This relationship was determined to be: $E = 33.51\ V^2/(V^2 - 620.28)$. Figure IV shows a graphical representation of this relationship and the points obtained for each of the 32 samples. The velocity and kinetic energy values used in plotting the points shown were "corrected" for temperature.

The data was also analyzed to determine the relationship between the 50% points of kinetic energy and temperature. As before, the cell variances were tested for homogeneity, this time they passed, i.e., no evidence of non-homogeneity was found.

We could, therefore, proceed to determine the relationship between temperature and kinetic energy.

Again using least squares methods, we obtained the relationships (see Figure V) between the 50% points of kinetic energy and temperature for each ball weight. (Comparison between points for 4 oz. and 16 oz. makes it apparent why the test in the first case indicated non-homogeneity of variances.)

Comparison of the slopes (Figure VI) confirmed that they did not differ significantly. Thus again a better estimate of the effect of temperature on the 50% points of kinetic energy should be obtained by "correcting" the data for ball weight and using the resulting (32) points to obtain a single relationship. This was done (with data "corrected" to 16 oz.) and we obtained K.E. = 48.375 - .012 t as our best estimate of the relationship between kinetic energy and temperature (see Figure VII).

To summarize: We found that, for the Mk2A4 Primer, both temperature and strike velocity had a significant effect on the 50% point of kinetic energy, i.e., the kinetic energy required to fire 50% of the primers is a function of the temperature of the primers and the strike velocity of the firing pin as well as of the sensitivity of the primer.

While only the Mk2A4 primer was tested, we would expect similar results for other percussion primers. (If we had reason to believe otherwise we would not have used the Mk2A4 for this test).

Therefore, we feel, these results indicate the desirability of considering the effect of primer temperature and firing pin strike velocity on the kinetic energy required by other primers to assure reliable performance. Also, the desirability in testing primers of simulating the strike velocity of the firing pin normally used to detonate the primer is indicated.

Further, one might infer that investigation of the effect of strike velocity should be considered for sensitivity testing in general.

Kinetic Energy vs. Ball Weight
at Various Temperatures
for Primer, Percussion, Mk2A4

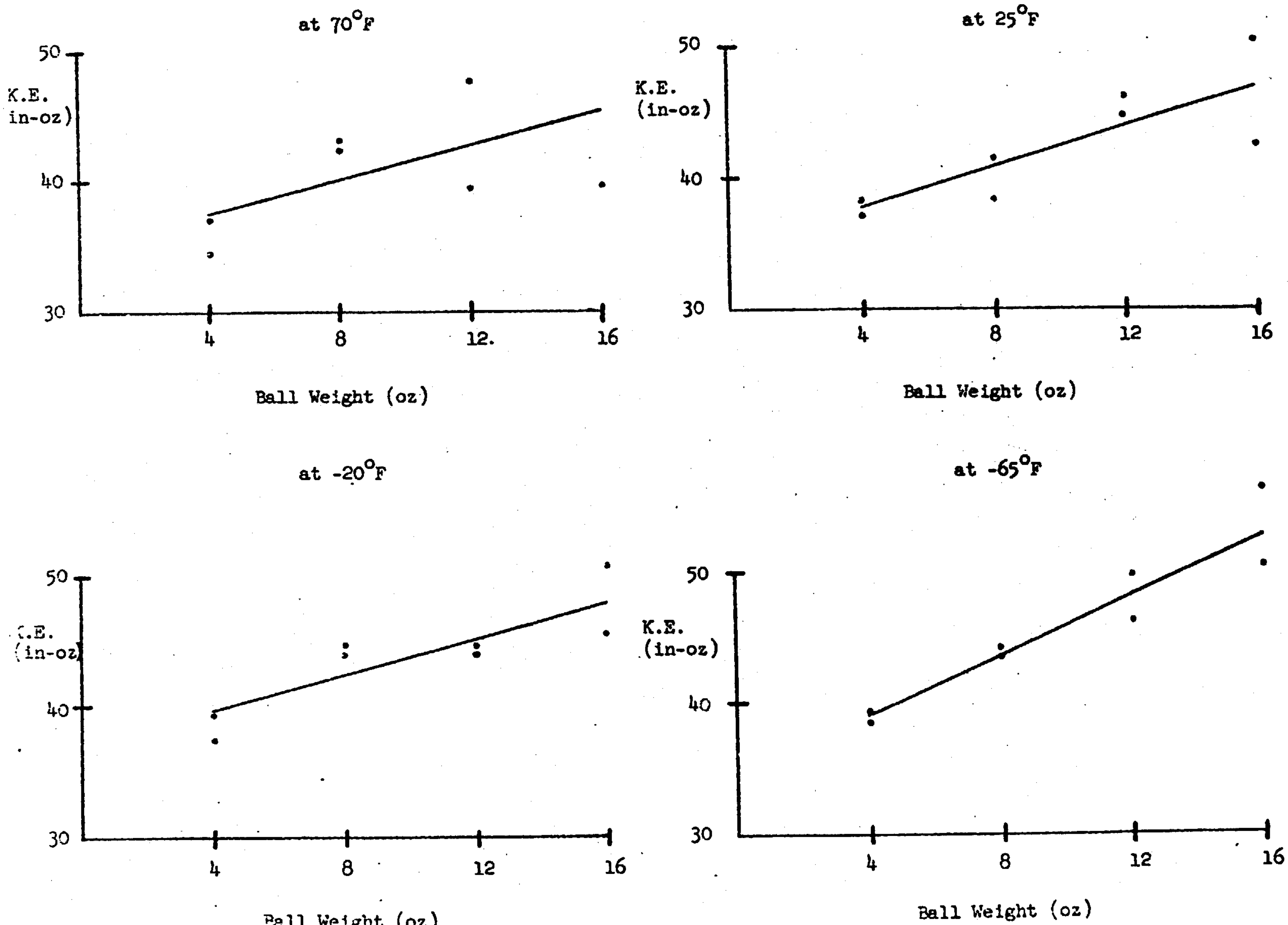


Figure I

Effect of Ball Weight on 50% Point
of Kinetic Energy - for Varicus Temperatures
(for Primer, Percussion, Mk2A4)

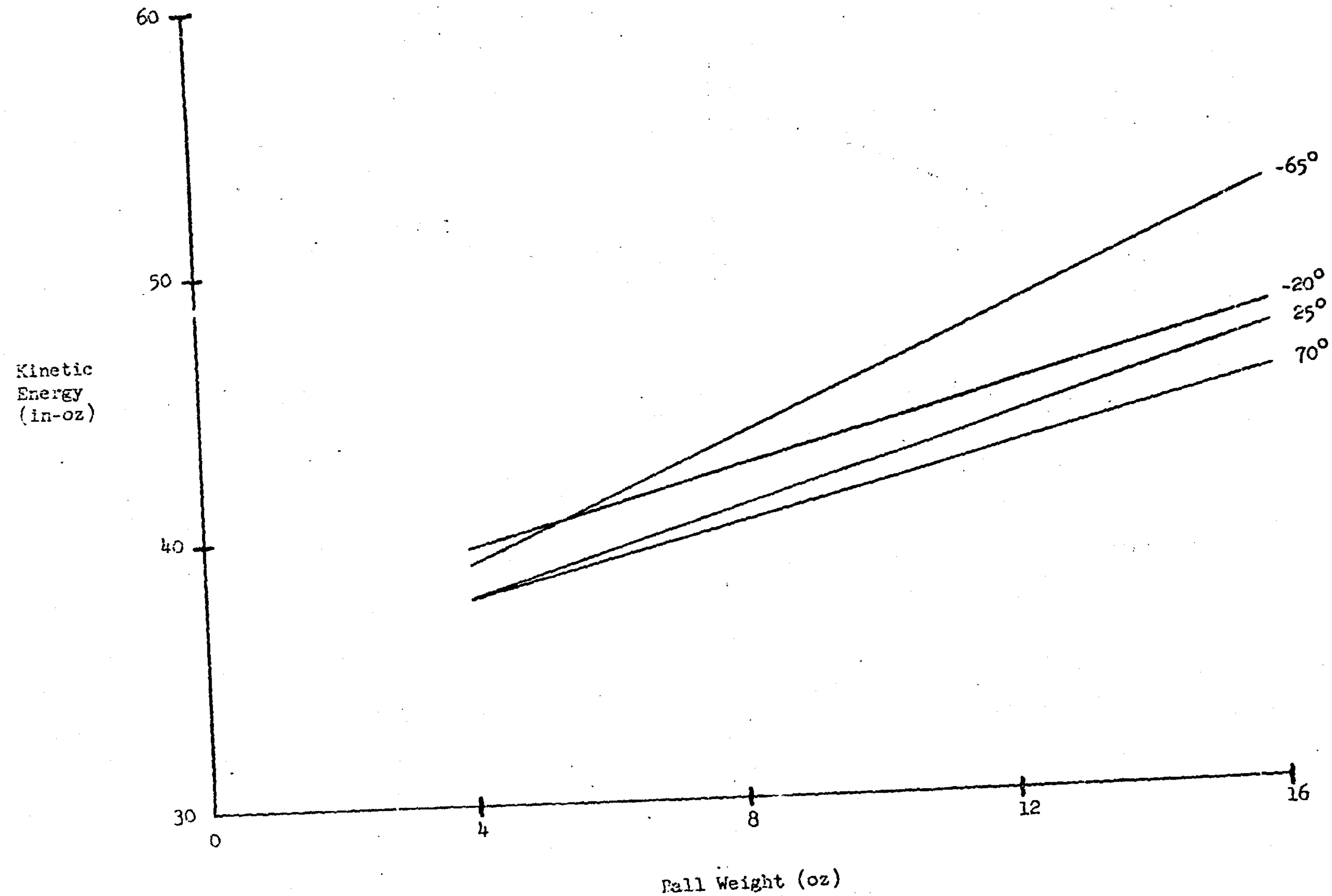


Figure II

Effect of Ball Weight on 50% Point
of Kinetic Energy
(for Primer, Percussion, Mk2A4)

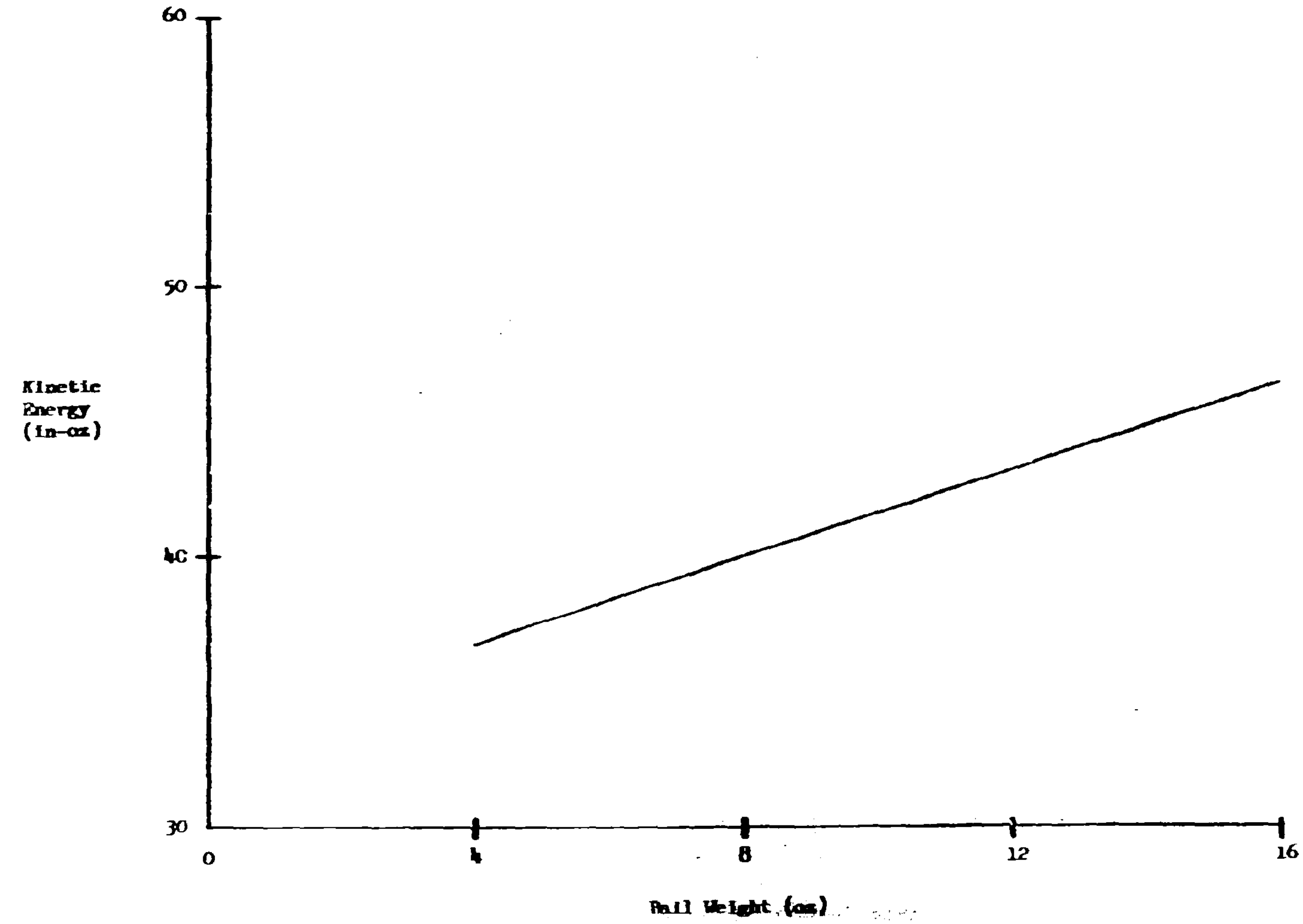


Figure III

Effect of Strike Velocity on 50% Point of Kinetic Energy
(for Primer, Percussion, Mk2A4)

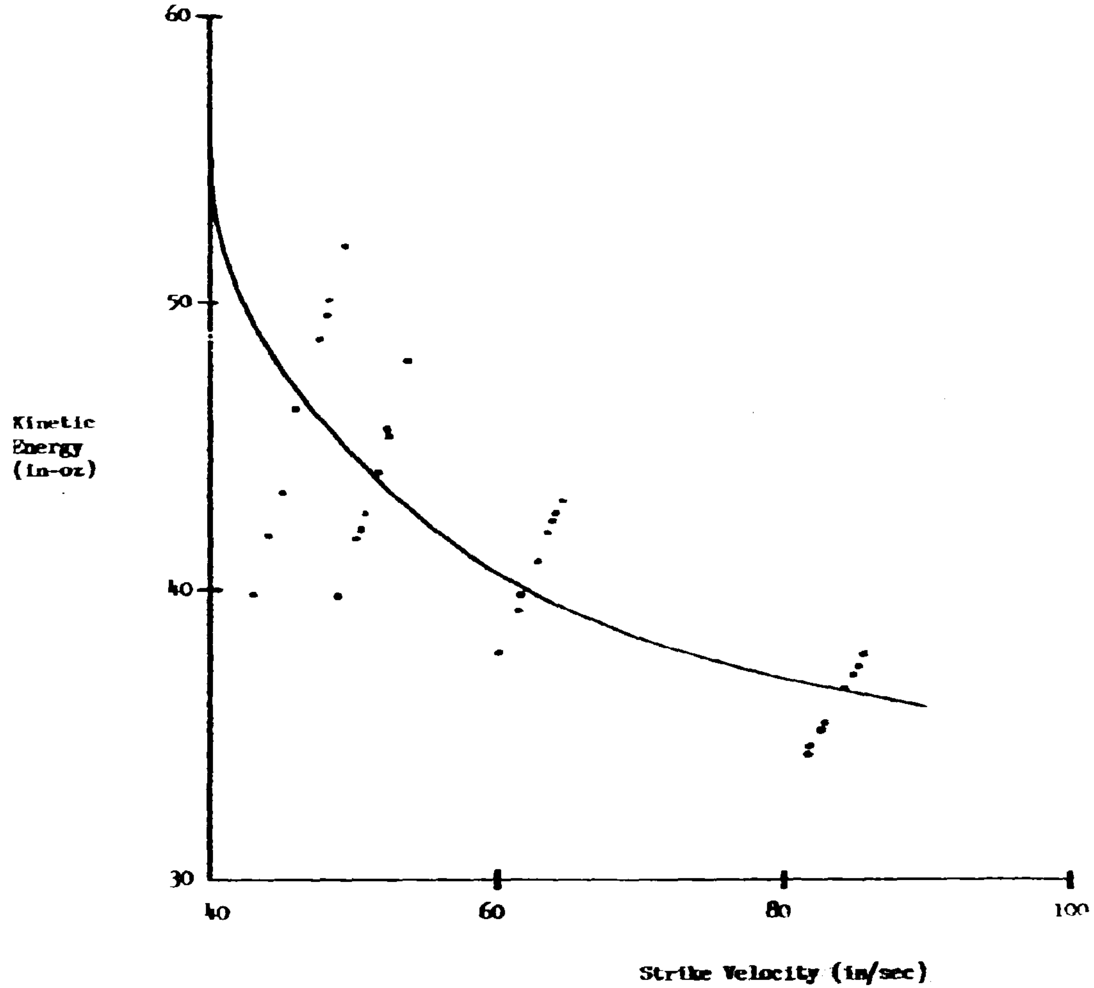


Figure IV

Kinetic Energy vs. Temperature
for Various Weight Balls
(for Primer, Percussion, M[illegible])

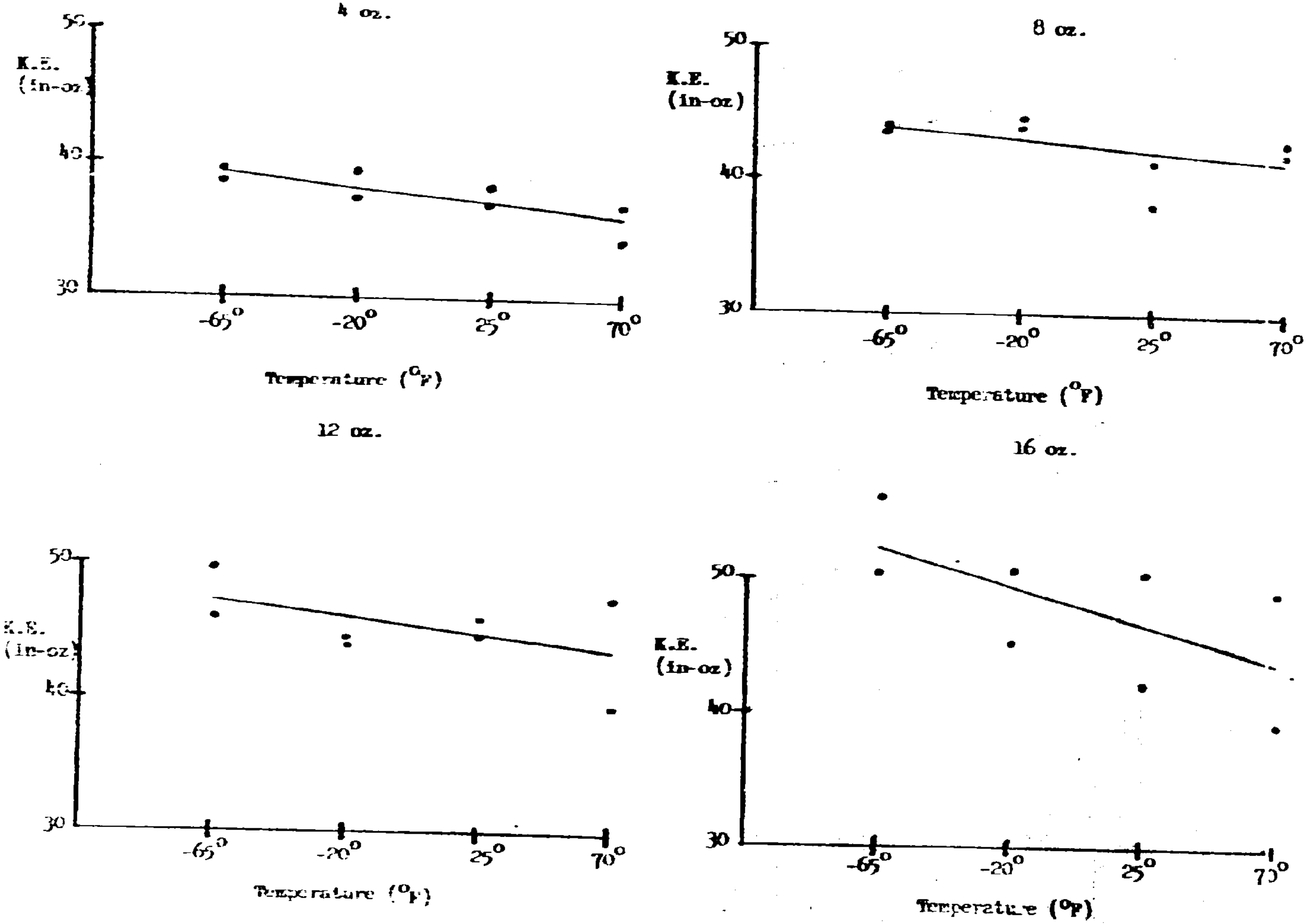


Figure V

Effect of Temperature on 50% Point
of Kinetic Energy - for various Ball Weights
(for Primer, Percussion, M29A4)

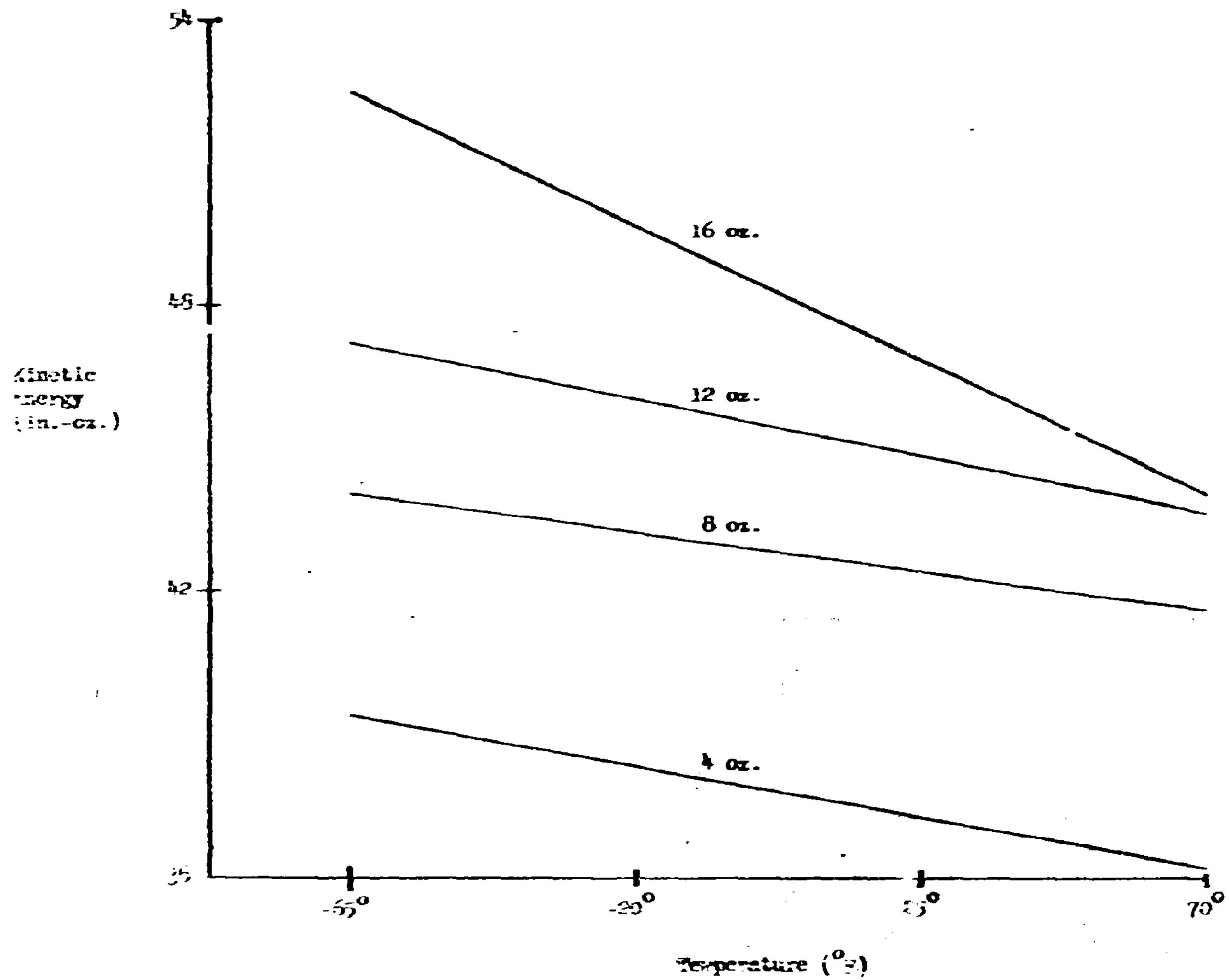


Figure VI

Effect of Temperature on 50% Point of Kinetic Energy (for Primer, Percussion, M42A1)

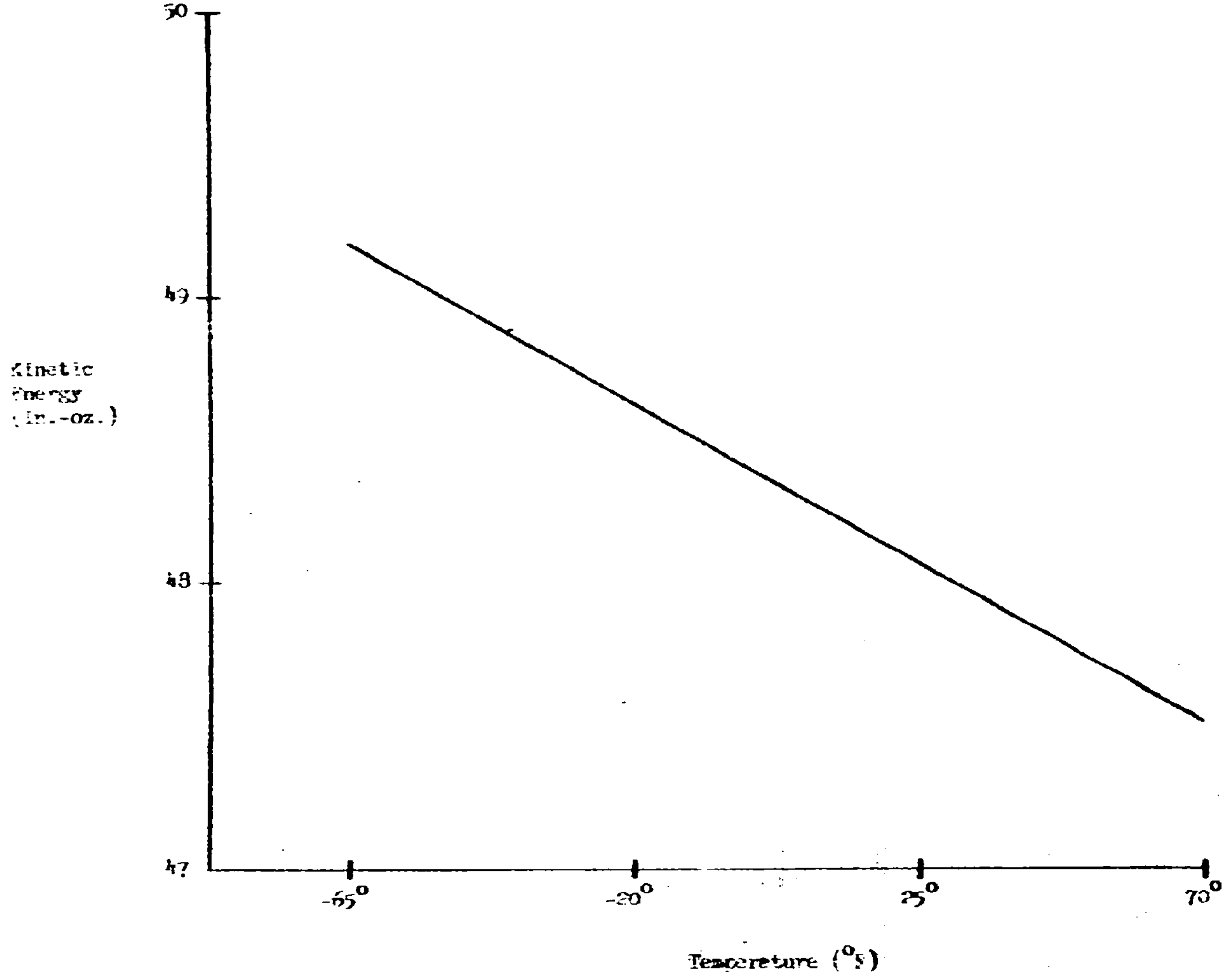


Figure VII

A COMPARISON OF RECONNAISSANCE TECHNIQUES FOR LIGHT OBSERVATION HELICOPTERS AND A GROUND SCOUT PLATOON

Harrison N. Hoppes, Barry M. Kibel, and Arthur R. Woods
Research Analysis Corporation, McLean, Virginia

INTRODUCTION. The Field Experiments Division of RAC is attempting to provide timely solutions to current Army problems involving tactics and doctrine. A major portion of the Division's field activities have dealt with helicopter operations. 1,2,3,4,5,6,7/ During July 1963 a research team from the Field Experiments Division conducted a two-sided, free-play field study with the 2nd Squadron, 4th Cav, 4th Armored Division to evaluate several techniques of helicopter reconnaissance. The results of that study were presented at the 9th Conference on the Design of Experiments in the paper "An Analysis of Helicopter Reconnaissance Techniques."

In November 1963, the Study's Project Advisory Group requested that a winter-phase investigation be carried out. The winter-phase venture measured the reconnaissance effectiveness of helicopters employing three reconnaissance tactics and compared the best of these tactics with the performance of a platoon of M114A1 Command and Reconnaissance Vehicles.

This paper describes the experimental design employed in the winter-phase investigation, summarizes the results obtained, and presents a brief statement of the study's conclusions and recommendations.

EXPERIMENTAL DESIGN. The experimental design is summarized in Table 1. As is indicated the three helicopter reconnaissance techniques studied were: (1) "high,"--flying at treetop level and maximum aircraft speed, (2) "low with pop up,"--nap-of-the-earth flight with emphasis placed on clearing an area before entering by popping up behind terrain masks, and (3) "low with dismount,"--nap-of-the-earth flight allowing the helicopter pilot to land and dismount an observer with binoculars. Single OH-13 helicopters, the vehicle currently used by the light-scout section of the air cavalary troop, 8/ were employed on all helicopter missions.

TABLE 1

Winter-Phase Experimental Runs

Ground Employment	Number of Runs For: Helicopter Tactic High	Low/Pop-Up	Low/Dismount	Ground Reconnaissance Platoon	Total Runs
Stationary	4	4	4	4	16
Moving	4	4	4	4	16
	8	8	8	8	32

The ground reconnaissance platoon generally consisted of five M114A1 scout vehicles. Usually the platoon leader divided the designated area or route into two sectors and coordinated the activity of the pairs of scouts operating in each sector. In performing their assigned mission, scout vehicle commanders frequently sent crew members forward on foot in much the same manner as helicopter pilots employed dismounted observers.

Like the companion study conducted during July 1963, the winter-phase investigation allowed scout elements complete freedom in determining paths of reconnaissance and time required to complete the assigned mission. Helicopter pilots were constrained only by the reconnaissance tactic they were instructed to employ; no restrictions whatsoever were placed on the ground reconnaissance platoon. Scenarios were designed to be tactically realistic and still permit experimental control.

Reconnaissance missions were conducted against static and fluid targets. Scout elements performed area reconnaissance missions against stationary target complexes and route reconnaissance missions against fluid targets. On each area reconnaissance mission scout elements reconnoitered against two target complexes, positioned to guard key terrain features and likely avenues of approach; each target complex consisted of one M113 APC and one or two M114A1's. On route reconnaissance missions target vehicles generally consisted of two M113's simulating the point of an armor column and three or four M114A1's providing route security. Target vehicles were mounted with gun cameras and event sequence recorders. Vehicle commanders were instructed to engage all reconnaissance elements acquired. Scout elements, on the other hand, were told to break contact whenever an enemy vehicle was acquired.

Throughout the paper the term "stationary runs" refers to those experimental runs involving stationary target complexes and "moving runs" to those involving fluid ground targets. Similarly, the term "target vehicle" is used to refer to the ground vehicles against which scout elements reconnoitered; it is never used to refer to reconnaissance vehicles taken under fire.

RESULTS. The winter-phase experimental design discussed above was successfully fulfilled between 20 January and 6 February 1964. A winter environment with snow cover, ground haze, and gray overcast was present on all days of field activity except February 5, 6.

Data, obtained from event sequence recorders and gun cameras mounted on target vehicles and from reconnaissance element sightings reported to a central control point, were analyzed using statistical techniques. Major emphasis was placed on comparing (1) the performance of helicopters vs ground scout teams, (2) the desirability of flying low/dismount vs high vs low/pop-up, and (3) the effects of reconnoitering against stationary vs fluid target complexes. The basic statistical technique used in making these comparisons was the analysis of variance; other common statistical techniques employed were t tests and chi-square tests.

Analyzing the results of two-sided, free-play experiments conducted in sector is often quite difficult. Frequently the outcomes of a given situation differ widely and the number of replications is small. At times experimental variables cannot be controlled as closely as is statistically desirable if troops and equipment are to be utilized when they are available. As a result, no attempt was made to analyze the experimental data in a rigorous manner. The statistical analyses did, however, provide an orderly framework for studying the large amount of data generated during the experiment.

Multiple measures of effectiveness were used in analyzing the experimental data. It was felt that no single measure could adequately consider all facets of the reconnaissance mission. Among the most important measures were those dealing with acquisitions, firings, and length of time required for mission completion. These included: (1) the percent of available targets acquired by reconnaissance elements, (2) the percent of ground targets acquiring at least one reconnaissance element, (3) the total number of times reconnaissance elements were detected, (4) the number of times reconnaissance elements and ground targets saw each other first, (5) the average length of interacquisition advantages scored, (6) the percent of the time the reconnaissance element was heard before it was seen, (7) the average lay time against scout elements (8) the percent of reconnaissance elements acquired that were taken under fire by ground target vehicles, (9) the total number of individual weapon firings at reconnaissance elements (10) the number of simulated rounds fired at scout elements, and (11) the time required to complete reconnaissance missions. Each of these measures has its merits and its limitations. By considering a variety of measures the relative ability of reconnaissance elements to acquire targets, avoid destruction, and provide timely information can be estimated.

Summary data concerning these measures are shown in Tables 2-4. From these data it can be seen that:

1. Helicopters acquired about 60 percent of the available ground targets regardless of the reconnaissance technique employed. Based on the percent of ground targets acquiring a helicopter, the total number of times helicopters were detected, and the net number of acquisition advantages scored against helicopters, the low/dismount tactic was superior to the other two helicopter tactics examined.

2. Based on the number of firings and number of rounds simulated against helicopters, pilots employing the low/dismount tactic also outperformed those using the high and the low/pop-up tactics.

3. On the average it required 10 minutes to complete missions flying at treetop level and maximum OH-13 speed. Low/pop-up missions lasted twice this long and low/dismount missions $3\frac{1}{2}$ times as long.

4. The acquisition performance of a single OH-13 helicopter was quite similar to the performance of a platoon of five M114A1 scout vehicles. Both acquired about the same percent of available targets and both had 8 net acquisition advantages scored against them. Only according to one acquisition measure did low/dismount helicopters and the ground scout platoon differ widely; helicopters employing the dismount tactic were acquired audibly before they were seen 23 percent of the time compared with only 3 percent of the time for ground scouts.

5. On stationary runs the acquisition and firing measures listed in Table 3 indicated that the ground reconnaissance platoon outperformed the single OH-13 helicopter flying the dismount tactic. On moving runs, however, the low/dismount helicopter tactic was more effective than the ground scout vehicles.

6. In terms of all 11 performance measures summarized in Table 3, reconnaissance elements were more effective against the fluid target complex than against the stationary ground complexes studied. Many of the observed differences were quite large. For example, about twice as many acquisitions were made by stationary ground vehicles as by fluid vehicles, about 2½ times as many acquisition advantages were scored by stationary target vehicles as by moving targets, the mean interacquisition advantage against scout elements was twice as long for static units as for fluid, and over three times as many simulated rounds were fired by stationary vehicles as by moving vehicles.

CONCLUSIONS. Based on the summary data presented in Tables 2-4 and on the more detailed statistical analyses conduced for each effectiveness measure, it was concluded that in a winter environment against targets of the type studied:

1. The low/dismount tactic is more effective than the tactics of flying high or nap-of-the-earth with pop up.

2. The overall effectiveness of a platoon of M114A1 scout vehicles is similar to that of a single OH-13 helicopter employing the nap-of-the earth with dismount tactic. The ground scout platoon was more effective on the stationary runs and the helicopter dismount tactic on the moving runs.

3. The performance of both helicopters and ground scouts was significantly better against fluid vehicles with a movement mission than against stationary target complexes.

RECOMMENDATIONS. If it is decided to employ either a ground scout platoon or helicopters on winter-time reconnaissance missions in terrain similar to the type studied, it is recommended that:

1. Ground scouts be used against suspected stationary targets if time permits. If time does not permit, the helicopter tactic of low/dismount is suggested.

2. In reconnoitering against a fluid enemy, helicopters using the dismount tactic should be employed.

REFERENCES

1. RAC-TP-83, "A Preliminary Investigation of Helicopter-vs-Tank Operations (U)," February 1963. UNCLASSIFIED

2. RAC-TP-122, "Helicopter Operations with the Long Range Reconnaissance Patrol During Exercise BIG LIFT (U)," April 1964. SECRET

3. RAC-TP-124, "An Evaluation of the Laying and Tracking Capability of the M114A1 Scout Vehicle, Mainly Against Helicopters (U)." UNCLASSIFIED

4. RAC-TP-139, "A Preliminary Investigation of the Main Tank Gun vs Aerial Targets (U)." UNCLASSIFIED Technical Paper in preparation.

5. RAC-TP-136, "An Investigation of Gunner Tracking Ability Against OH-13 Helicopters employing Evasive Maneuvers (U)." UNCLASSIFIED Technical Paper in preparation.

6. RAC-TP- , "An Evaluation of Helicopter Pop-Up Tactics (U)." UNCLASSIFIED Technical Paper in preparation.

7. RAC-T-433, "Reconnaissance Techniques for Light Observation Helicopters in a Summer Environment -- A Two-Sided Field Play (U)." FOR OFFICIAL USE ONLY

8. Headquarters, Department of Army, FM 17-36, "Armored Cavalry Platoon and Troop, Air Cavalry Troop and Armored Cavalry Squadron (U)," December 1961.

TABLE 2

SUMMARY OF PERFORMANCE DATA FOR THREE HELICOPTER TACTICS, GROUND RECONNAISSANCE PLATOON

Effectiveness Measure	Helicopter Scout Tactic: Low/ Dismount	Helicopter Scout Tactic: High	Helicopter Scout Tactic: Low/ Pop-Up	Ground Scout Platoon
1. Percent of Ground Targets Acquired	59%	61%	55%	65%
2. Percent of Ground Targets Acquiring Reconnaissance Element	59%	81%	81%	51%
3. Number of Times Reconnaissance Element Acquired	31	57	51	51
4. Total Number of Acquisition Advantages Scored By Ground Targets	23	40	39	29
Total Number of Acquisition Advantages Scored By Reconnaissance Elements	15	9	13	21
Net Number of Acquisition Advantages Scored By Ground Targets	8	31	26	8
5. Length of Interacquisition Advantage Against Scout Element (Mean) in seconds	26	24	16	17
Length of Interacquisition Advantage Against Scout Element (Median) in seconds	10	21	10	20
6. Percent of Time Reconnaissance Element Was Acquired Audibly Before Visually	23%	21%	27%	3%
7. Lay Time Against Reconnaissance Element (Mean) in seconds	12	12	12	12
Lay Time Against Reconnaissance Element (Median) in seconds	10	8	11	8
8. Percent of Reconnaissance Elements Detected That Were Taken Under Fire	61%	58%	47%	72%
9. Number of Times Reconnaissance Elements Were Taken Under Fire By Ground Weapons	23	43	32	48
10. Number of Simulated Rounds Fired At Reconnaissance Elements	3424	5016	4880	2556
11. Time Required to Complete Mission (Mean) in minutes	35	10	20	93

TABLE 3

SUMMARY OF PERFORMANCE DATA FOR HELICOPTERS EMPLOYING LOW/DISMOUNT TACTIC, GROUND RECONNAISSANCE PLATOON

Effectiveness Measure	Stationary Runs		Moving Runs	
	Helicopter Low/ Dismount	Ground Scout Platoon	Helicopter Low/ Dismount	Ground Scout Platoon
1. Percent of Ground Targets Acquired	48%	62%	72%	68%
2. Percent of Ground Targets Acquiring Reconnaissance Element	62%	62%	56%	41%
3. Number of Times Reconnaissance Element Acquired	20	31	11	20
4. Total Number of Acquisition Advantages Scored By Ground Targets	18	19	5	10
Total Number of Acquisition Advantages Scored By Reconnaissance Elements	4	9	11	12
Net Number of Acquisition Advantages Scored by Ground Targets	14	10	-6	-2
5. Length of Interacquisition Advantage Against Scout Elements (Mean) in seconds	32	23	7	13
Length of Interacquisition Advantage Against Scout Elements(Median) in seconds	12	23	5	10
6. Percent of Time Reconnaissance Element Was Acquired Audibly Before Visually	43%	3%	10%	0%
7. Lay Time Against Reconnaissance Element (Mean) in seconds	10	11	19	14
Lay Time Against Reconnaissance Element (Median) in seconds	6	8	14	12
8. Percent of Reconnaissance Elements Detected That Were Taken Under Fire	75%	68%	36%	80%
9. Number of Times Reconnaissance Elements Were Taken Under Fire	18	28	5	20
10. Number of Simulated Rounds Fired At Reconnaissance Elements	2984	1296	440	1360
11. Time Required to Complete Mission (Mean) in minutes	49	117	21	70

TABLE 4

SUMMARY OF PERFORMANCE DATA FOR STATIONARY, MOVING RUNS

	Effectiveness Measure	Stationary Runs	Moving Runs
1.	Percent of Ground Targets Acquired	54%	68%
2.	Percent of Ground Targets Acquiring Reconnaissance Element	77%	57%
3.	Number of Times Reconnaissance Element Acquired	125	65
4.	Total Number of Acquisition Advantages Scored By Ground Targets	93	38
	Total Number of Acquisition Advantages Scored By Reconnaissance Elements	21	37
	Net Number of Acquisition Advantages Scored by Ground Targets	72	1
5.	Length of Interacquisition Advantage Against Reconnaissance Elements (Mean) in seconds	27	13
	Length of Interacquisition Advantage Against Reconnaissance Elements(Median)in seconds	13	10
6.	Percent of Time Reconnaissance Element Was Acquired Audibly Before Visually	26%	2%
7.	Lay Time Against Reconnaissance Element (Mean) in seconds	12	14
	Lay Time Against Reconnaissance Element (Median) in seconds	8	12
8.	Percent of Reconnaissance Elements Detected That Were Taken Under Fire	66%	48%
9.	Number of Times Reconnaissance Elements Were Taken Under Fire	106	40
10.	Number of Simulated Rounds Fired At Reconnaissance Elements	12,240	3736
11.	Time Required to Complete Mission (Mean) in minutes	52	26

A STUDY OF PROBABILITY ASPECTS OF A SIMULTANEOUS SHOCK WAVE PROBLEM

A method of solving probabilistic problems without a computer.

Edward C. Hecht
Picatinny Arsenal, Dover, New Jersey

I am going to present a procedure for the rapid solution by desk calculator of an involved probabilistic problem. Figure 1 is a sample computation, showing all the paperwork necessary for one solution. The first two and the last of even these few columns are identical for every computation of this sort.

Unfortunately, although characteristically, the evolution of the simple tool requires a long explanation. I have made up a problem as a vehicle for the explanation, and I hope you will bear with me as I toil through it.

In a certain classified ordnance application, two HE weapons are detonated, and it is a matter of concern whether the shock waves from the explosives arrive at a point between them simultaneously and before the occurrence of a particular event at the intermediate point.

Speaking generally, we have three events, each with its own distribution in time, and we want to find the probability associated with certain spacings and orders of occurrence of the events.

Calling the locations of the two weapons and the intermediate point A, C, and B, respectively, as illustrated in figure 2,

A______________B_____________C

the problem is to determine the probability of arrival of shock waves from A and C at B simultaneously and before the occurrence of an event at B (called hereafter event B). Simultaneous is defined arbitrarily as within 100 micro-seconds. The expected times of the detonation and event B may be the same or different in some ordered manner.

For visualization purposes, it may be considered that the interaction effects of the two shock fronts are to be photographed using a Schlieren technique. The shock waves must meet within the brief angle covered by the camera. The camera's film supply is limited; and event B is the start of film exposure. The 100-microsecond simultaneity period is the time within which the interaction is within the narrow range of the camera. This is not the real problem, but a hypothetical problem that I have invented, since the true problem is classified. I want to emphasize that it is the general method of solution, rather than the problem, that I want to present. To make the problem fit the solution, system failure must be thought of as coincidence of the shock waves at B but before the film has started running.

This paper will develop a procedure for determining the probability of system failure given the expected times of the detonation of the HE weapons, the expected time of event B, and the probability distribution of these times.

In the situation in which this problem arose, it was necessary to find a solution because the probability of the shock waves arriving at B simultaneously and before event B occurred was required to be very small, of the order of 0.001, while the variability of some of the proposed detonators and other components was of the same order as the shock wave travel time from A or C to B. It was necessary to find whether such variability could be tolerated, and, if not, how tight the dispersion had to be. In order to aid the required design decisions, it seemed desirable to get the results in the parameterized form of a plot of system sigma versus probability for selected shock wave travel times.

As a matter of personal preference, I looked for a desk calculator solution, which might later be programmed for computer.

In its simplest form, which I will discuss first, the problem has A and C equidistant from B, so that the shock wave travel times are equal, and all events will be expected to be absolutely simultaneous.

The problem requires finding, for every infinitessimal interval of time, the differential probability of system failure, which is the product of three probabilities -- the probability that A detonates within that infinitessimal interval; the probability that C detonates within 100 microseconds of the interval; and the probability that B has not occurred one shock wave travel time later than that interval. Integrating the product of these probabilities over all time gives us the total probability of system failure. For the example used, the probability distributions of the event times were all taken as normal and the standard deviations as equal for A and C; but these assumptions are not necessary to apply the general method of solution.

Now I will put the problem in more general terms. In any infinitesimal period of time, dt, at a time t', the probability of system failure is the compound probability that event A has happened a constant, predictable time earlier than t', that event C has happened within a stated small interval about a constant, predictable time earlier than t', and that event B has not yet happened at time t'. The constant, predictable times are the shock wave travel times from A and C to B, which may or may not be equal in the general case; and the stated small interval is the simultaneity period, which must be small enough relative to the travel times that events occurring within it may be considered simultaneous.

Integrating this compound probability over all time yields the total probability of system failure.

As illustrated in figure 3, we will call the times of occurrence of events, A, B, and C, t_A, t_B, and t_C, and the probability distributions of these events, $P(t_A \geq t)$, $P(t_B \geq t)$, and $P(t_C \geq t)$. The shock wave travel time depends largely on the travel distance and on the amount of explosive involved, and is considered to be constant. We will call the travel times from A to B and from C to B, t_{AB} and t_{CB}. The short period within which shock wave arrival is considered simultaneous, we will call Δ.

Starting with the probability distributions of the times of events A, B, and C, each of which has somewhat of the appearance of the top curve of figure 4, we proceed as follows to find the probability of system failure.

For convenience in notation and in thinking about the problem, we will tie our general time frame to the time frame of event A. This poses no difficulty since the expected times of events A, B, and C are known; and we would, in any event, have used one of these fixed times as the origin of the general time system.

For a system failure to occur at time t', then, $t' = t_A + t_{AB}$. Also, for the necessary simultaneity, t_C must occur within the period Δ and later than t_A by the difference between their traveltimes to B; or, in mathematical language, as it is written on the figure 3. (Of course, it is an algebraic "later" and, if $t_{AB}-t_{CB}$ is negative in sign, t_C must occur earlier in time than t_A for simultaneity of shock wave arrival at B.)

Finally, for system failure, when the simultaneous shock wave arrives at B at time t', event B must not have occurred. Therefore, $t_B \geq t_A + t_{AB}$.

The probability that event A occurs within any differential period of time, dt, is $d\,P\,(t_A \geq t)$. This differential probability must be multiplied by the probability that event C occurs in an interval, Δ, $\overline{t_{AB}-t_{CB}}$ later than dt, $P\,(t_A + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t_A + \overline{t_{AB} - t_{CB}} + \Delta)$.

The product must further be multiplied by the probability that event B occurs after t', $P\,(t_B \geq t_A + t_{AB})$. Integrating this final product over all $P\,(t_A \geq t)$ is equivalent to integrating over all t, since $P\,(t_A \geq t)$ is a single valued function of t.

The probability that C occurs within the simultaneity interval of any time t is obtained from the probability distribution of the times of event C. In the case of normal distributions, this is easy to do.

The curve of this function versus t has the general form of the bottom curve of figure 4. Then the probability distribution, $P\,(t_B \geq t)$ versus t is modified to $P\,(t_B - t_{AB} \geq t)$ versus t.

These two functions of t are multiplied together to get $P\,(t_B - t_{AB} \geq t)$. $P\,(t + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t + \overline{t_{AB} - t_{CB}} + \Delta)$ versus t.

Since $P\,(t_A \geq t)$ is a single valued function of t, values of this probability can be substitued for values of t to get a plot of $P\,(t_B - t_{AB} \geq t)$

times $P(t + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t + \overline{t_{AB} - t_{CB}} + \Delta)$ versus $P(t_A \geq t)$ (as in figure 5). The area under this last curve is;

$$\int_0^1 P(t_B \geq t + t_{AB})\, P(t + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t + \overline{t_{AB} - t_{CB}} + \Delta)\, dP(t_A \geq t),$$

and this is the probability of a system failure, P_{SF}. The important attribute of this method of solution is that this area may readily be evaluated, without constructing the curves, by use of Simpson's Rule.

If a Simpson Rule division of the area into ten parts gives us enough accuracy (as it well may, depending on how accurately we know the shapes of the distributions involved), we need only find eleven values of t corresponding to $P(t_A \geq t)$ values of 0, 0.1, 0.2, etc., to 1.0; and two of these times are plus and minus infinity. At these extreme times, the simultaneity probabilities are zero. For the intermediate times, the simultaneity probability may easily be looked up in any well-detailed table of areas under the normal curve for the normal distributions assumed in our example.

A sample computation has been shown in figure 1. For any computation using a 10-part Simpson Rule integration, the first two columns will be the same. To get the simultaneity probabilities, it is observed that $\Delta = 0.01 \cdot \sigma_A$; so that for the second time point the probability looked up is that of being between 1.292 and 1.272 standard deviations away from the mean. For the column of $t + t_{AB}$, it is noted that $t_{AB} = 1\sigma_B$ and that $\sigma_A = 2\sigma_B$. Then since $\bar{t}_A = \bar{t}_B$, $\bar{t}_A + k\sigma_A = \bar{t}_B + (2k+1)\sigma_B$. Having found these t_B equivalents, the table look-up is easy. The next column is the product of the third and fifth columns. These products are multiplied by the Simpson Rule factors, 1, 4, 2, 4, 2, etc.; and the sum is multiplied by the class interval of 0.1 and divided by 3 to get the value of the integral (see figure 4), which is the answer sought. Repetition enables the computation to be performed in about 20 minutes.

Taking expected times as equal, and at least two of the sigmas as equal, allows results to be plotted as in figure 6. Other conditions are not much harder to compute, but the results are harder to present.

Many problems besides this fictionalized and hypothetical one are susceptible to this technique of solution. The method is one which, with some familiarity, enables an engineer or mathematician to solve problems involving probability at his desk before or without submitting them for computer solution. And, of course, it is useful for those who have no access to a computer.

SAMPLE COMPUTATION

$\sigma_A = \sigma_C = 10$ M SEC $[= 2\,\sigma_B]$ $\sigma_B = 5$ M SEC $T_{AB} = T_{CB} = 5$ M SEC $[= \sigma_B]$ $\Delta = 0.1$ M SEC $[= .01\,\sigma_A]$

$P\ T_A \geq T$	T	P_{SIMUL}	$T + T_{AB}$	$P[T_B \geq T + T_{AB}]$	$P_{SIMUL}\,P[T_B \geq \cdot\cdot]$	SIMPSON RULE FACTORS
0	$\overline{X_A} + \infty$	0	$\overline{X_B} + \infty$	0	0	1
.1	$\overline{X_A} + 1.282\,\sigma_A$	.0035	$\overline{X_B} + 3.564\,\sigma_B$	.0002	.00000	4
.2	$\overline{X_A} + .842\,\sigma_A$	.0056	$\overline{X_B} + 2.684\,\sigma_B$	.0036	.00002	2
.3	$\overline{X_A} + .524\,\sigma_A$	.0070	$\overline{X_B} + 2.048\,\sigma_B$	.0203	.00014	4
.4	$\overline{X_A} + .253\,\sigma_A$	.0077	$\overline{X_B} + 1.506\,\sigma_B$	.0660	.00051	2
.5	$\overline{X_A} + 0$	.0080	$\overline{X_B} + 1.000\,\sigma_B$	.1587	.00127	4
.6	$\overline{X_A} - .253\,\sigma_A$	.0077	$\overline{X_B} + .494\,\sigma_B$	.3107	.00239	2
.7	$\overline{X_A} - .524\,\sigma_A$	.0070	$\overline{X_B} - .048\,\sigma_B$	.5191	.00363	4
.8	$\overline{X_A} - .842\,\sigma_A$	.0056	$\overline{X_B} - .684\,\sigma_B$	.7530	.00422	2
.9	$\overline{X_A} - 1.282\,\sigma_A$	.0035	$\overline{X_B} - 1.564\,\sigma_B$	.9411	.00329	4
1.0	$\overline{X_A} - \infty$	0	$\overline{X_B} - \infty$	1.	0	1
					$\Sigma = .04770$	

SYSTEM FAILURE PROBABILITY $= \dfrac{0.1\,\Sigma}{3} = .0016$

FIGURE 1

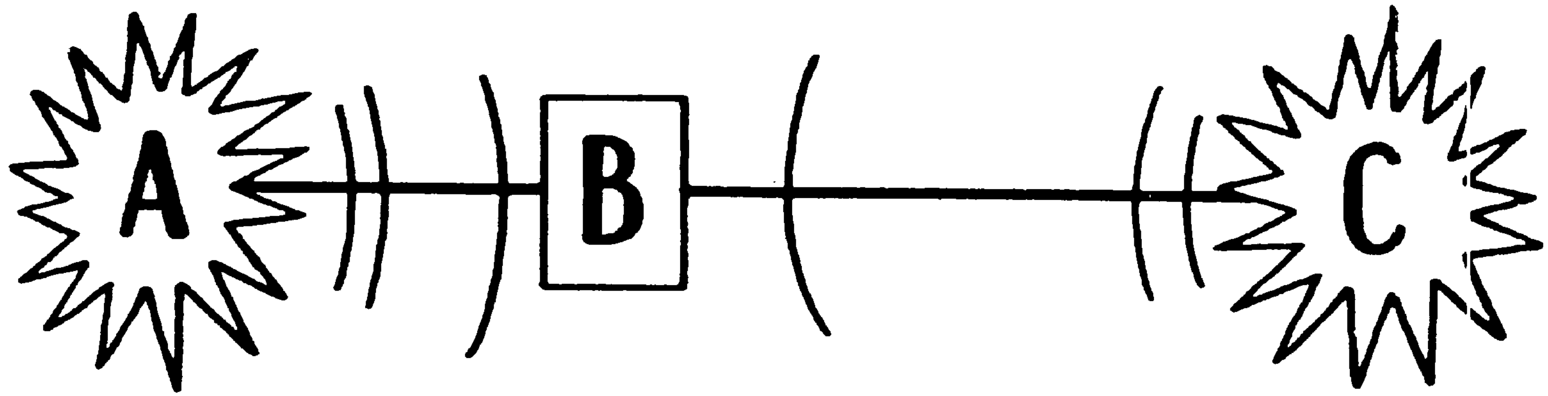


FIGURE 2

Event Times	t_A	t_B	t_C
Probability Distribution	$P(t_A \geq t)$	$P(t_B \geq t)$	$P(t_C \geq t)$
Shock Wave Travel Times	t_{AB}	t_{CB}	
Simultaneity Period	Δ		

Simult. of A and B $\quad t_A + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t_A + \overline{t_{AB} - t_{CB}} + \Delta$

Simult. Shock Wave Arrival

At B Before Event B $\quad t_B \geq t_A + t_{AB}$

Total System Failure Probability

$$\int_0^1 P(t_B \geq t + t_{AB})\, P(\textit{Simult. of A and B})\, dP(t_A \geq t)$$

FIGURE 3

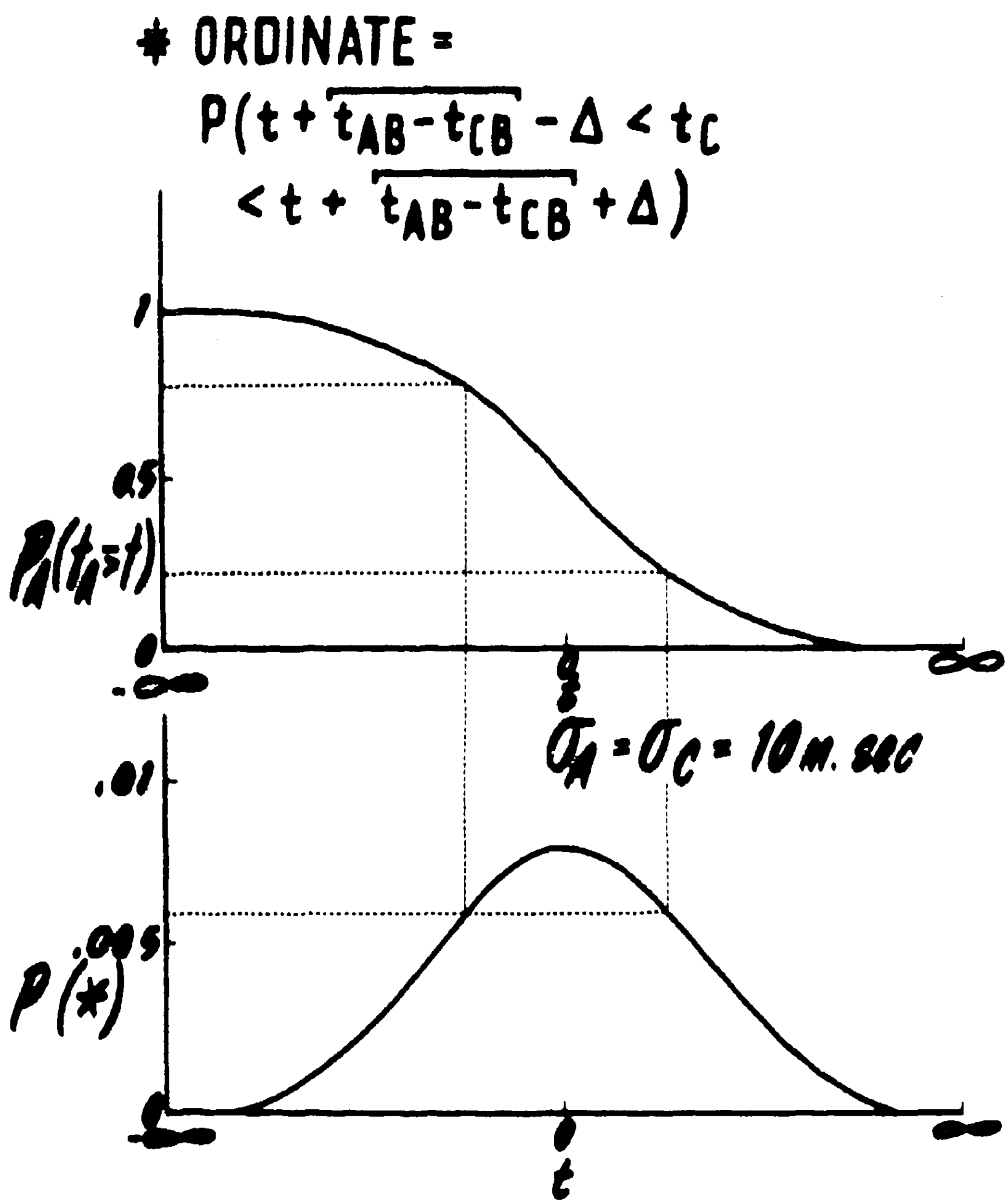

* ORDINATE =
P(t + tAB - tCB - Δ < tC
< t + tAB - tCB + Δ)
1
0.5
PA(tA=t)
0
-∞
0
∞
σA = σC = 10 m. sec
.01
.005
P(*)
0
-∞
0
∞
t

FIGURE 4

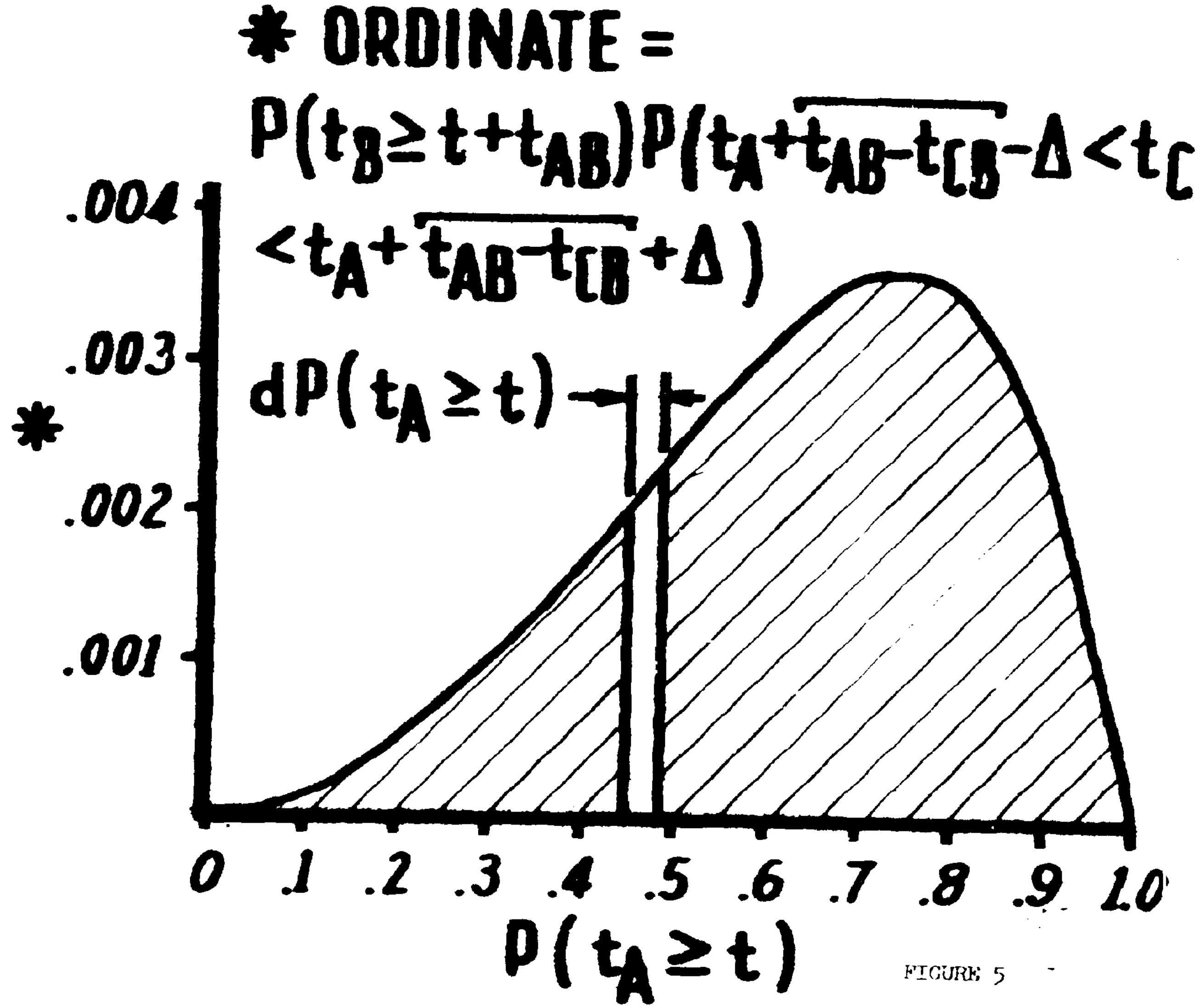
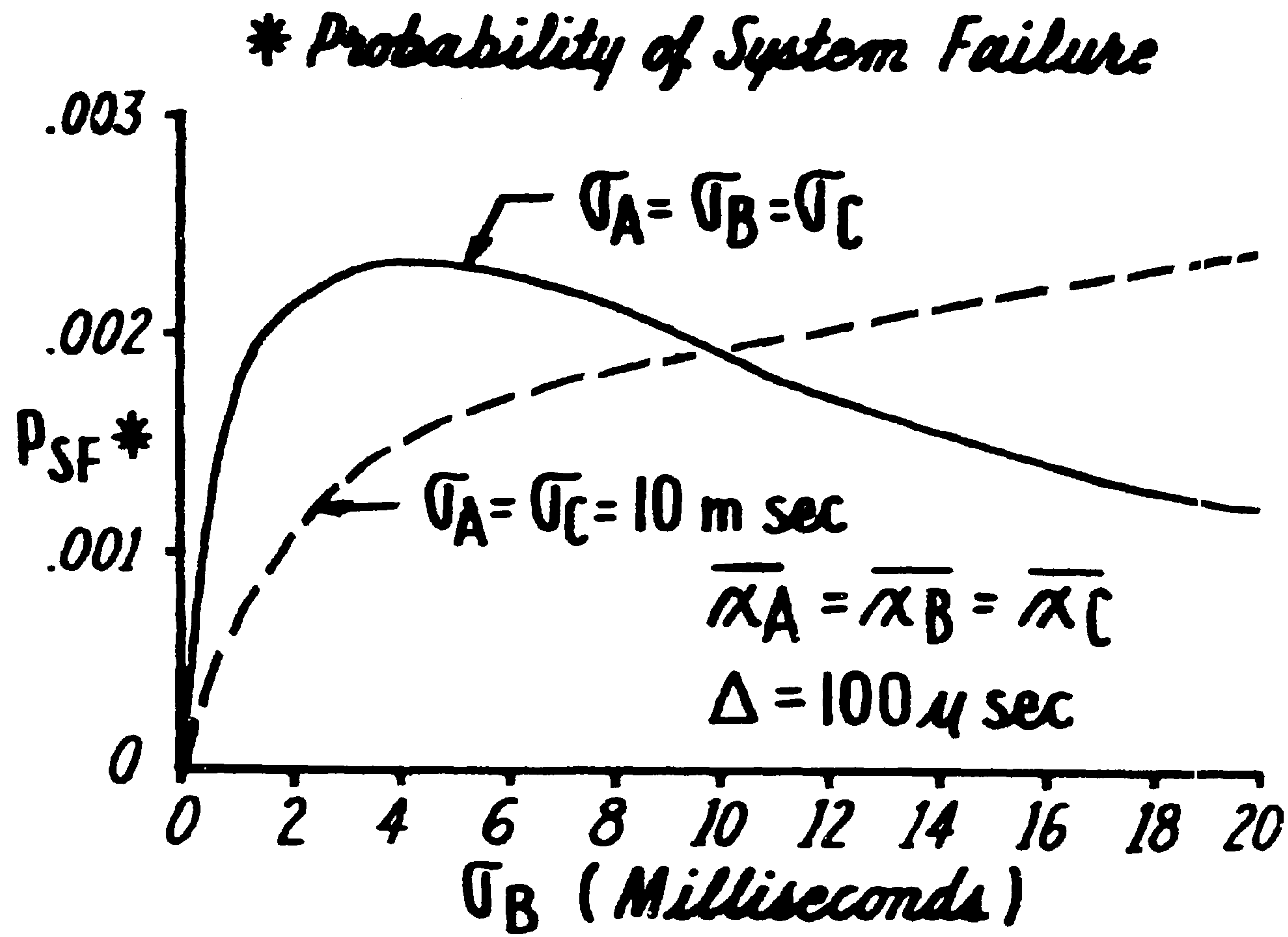

* ORDINATE =
$P(t_B \geq t + t_{AB}) P(t_A + \overline{t_{AB} - t_{CB}} - \Delta < t_C < t_A + \overline{t_{AB} - t_{CB}} + \Delta)$
$dP(t_A \geq t)$
*
.004
.003
.002
.001
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
$P(t_A \geq t)$

FIGURE 5



* Probability of System Failure
.003
.002
.001
0
PSF *
σA = σB = σC
σA = σC = 10 m sec
x̄A = x̄B = x̄C
Δ = 100 μ sec
0 2 4 6 8 10 12 14 16 18 20
σB (Milliseconds)

FIGURE 6

A DATA COLLECTION PROCEDURE FOR ASSESSING NEURO-MOTOR PERFORMANCE IN THE PRESENCE OF MISSILE WOUNDS

William H. Kirby, Jr., William Kokinakis,
Larry M. Sturdivan, and William P. Johnson
U. S. Army Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

INTRODUCTION. While medical clinicians diagnose, treat, and judge the sequels of injury, it is of interest to others such as those engaged in man-task or man-machine system studies to consider the effects of injury from additional points of view. Those concerned with the medical problems are naturally interested in procedures and optimal treatments which prevent or at least minimize the consequences of injury. Those concerned with man-machine system performance problems are interested, in addition, in the ability of the injured or otherwise stressed individual to perform given tasks. Common to both the clinician and the man-machine researcher interested in injury is a need for a better understanding of the mechanisms and responses associated with traumatic pathological dynamics.

We are in the process of developing a methodology for describing and assessing anatomical and physiological pathologies associated with missile wounds. Hopefully, we will be able to express these in terms of a set of affectors and/or effectors such as those found in the nerve-muscle or neuro-motor structure. The reason for this approach is that they may serve as a common denominator for describing injury as well as for describing the task or machine operation requirement. In this presentation we will limit our attention to wounds caused by missiles. The type and amount of data to be collected will probably be influenced by the number of accident cases that enter hospitals which are accessible for study.

This is an interdisciplinary problem area in which clinicians, engineers, mathematicians and statisticians should meet. It is usually the case that such multi-discipline representatives are faced early in the process with certain communication problems. For example, a function to the clinician has one meaning, but to the mathematician it has quite another. A medical researcher may apply a Chi-Square test to a set of data in which the statistician may insist that the application is invalid due to the fact that the data do not conform to a normal distribution. A surgeon may be entirely satisfied that the maximal strength of a grasp is equal in both hands of a patient as determined through a hand squeezing process whereas the engineer is satisfied only if such an assessment is in quantitative terms such as a pressure-time history. Clinicians can really get confused whenthey attempt to understand differences between mathematicians and statisticians.

DISCUSSION. Our first approach considers the body as a system composed of a set of clinical subsystems coordinated to maintain life and control human performance. These clinical subsystems will initially be divided for convenience into a primary and secondary group. The primary group will include the neurological, cardiovascular, respiratory, skeletal, and muscular sybsystems. The secondary group will consist of the gastrointestinal, genitourinary, and endocrine subsystems. While we intend to collect some data associated with the secondary group, initially only the primary group will be considered in detail.

We attempt to describe performance in terms of a simplified set of afferent and efferent (input-output) factors shown in Figure 1. For the present we intend only to recognize the presence or absence of the afferent (input) factors - vision and hearing, and the efferent (output) factor - voice. Essentially then we have reduced our performance descriptors to the first six listed in Figure 1. Actually these descriptors are regional subdivisions of the human body and they will be represented by the (motor) muscles which are located in the respective regions. The neurological or muscle activator network is distributed over these regions and no controlled human actions occur without its activation. It was therefore natural to choose these motor factors as a common denominator to which all performance phenomena and subsystem changes may be related.

Performance Factors

Performance Symbol	Efferent (Output) Factors
e_1	Right upper limb
e_2	Left upper limb
e_3	Right lower limb
e_4	Left lower limb
e_5	Head and Neck
e_6	Trunk
e_7	Verbal communication
	Afferent (Input) Factors
e_8	Vision
e_9	Hearing

FIGURE 1

Using the above rationale we are interested in collecting data from accidental wound cases in order to describe, classify, and relate important missile characteristics, clinical subsystem injuries, and neuro-motor performance phenomena. Concurrently a more comprehensive study of this type of problem is being considered but which is beyond the scope of this presentation.(1)

The Neuro-Motor or Effector Logic. The "terminal" body tissue or structure directly responsible for physical movements as indicated above is muscle. Inasmuch as muscles are innervated by specific peripheral nerves, associated nerves and muscles have become known as "neuro-motor units." Fortunately the nerve-muscle anatomical distribution system has been well established by anatomists in the past.

In order to demonstrate this logic attention is directed to Figure 2 which is a matrix showing the muscles and their actions in the upper limb. This matrix could represent either of the effectors, e_1 or e_2

since they are symmetrical. The numbers along the abscissa are subscripts of the letter "A" in which each subscript represents a specific anatomical action as described in Appendix A. The numbers along the ordinate are subscripts of the letter "M" in which each subscript represents a specific muscle also described in Appendix A. The muscles (m_i, i=1,2,... 61) are arranged in a manner such that the lower numbers represent muscles in the shoulder and in ascending order represent muscles in the arm, forearm, and hand.

The distribution of nerves and their contained fibers which innervate the skeletal muscles is unique. For instance, the large number of nerve fibers which originate from a given source such as a particular spinal cord segment, are dispersed into a multiplicity of branches. As if to provide maximum reliability, many fibers from the same source reach a given muscle by different pathways. On the other hand nerve fibers from the same spinal cord level are known to innervate different muscles. While the nerve pathways are not demonstrated here, some idea of the nerve fiber distribution may be obtained from the left side of Figure 3. The C_i (i = 1, 2, ... 8), T_1, and T_2 represent nerve roots (large groups of fibers) which emerge from the designated spinal cord levels. These are identified in Appendix A. The letter, C, refers to the cervical or neck region and the subscripts refer to the specific locations out of which the bundles of nerve fibers flow. The letter, T, refers to the thoracic or chest region. Only the first two thoracic nerve bundles, T_1 and T_2, are included in Figure 3.

It is interesting to observe from this matrix that a given muscle is innervated by nerve fibers from more than one source. Note, for example, that M_4 (pectoralis major) is innervated by nerve fibers derived from several spinal cord segments, namely, C_5, C_6, C_7, C_8, and T_1. An important implication in conjunction with the previous comments on reliability is that if the spinal cord were severely injured at the level of C_8, muscle, M_4, would not become completely paralyzed inasmuch as it would still receive considerable innervation from fibers above the site of injury namely, C_5, C_6, and C_7. It is also interesting to observe the number of muscles in the shoulder region innervated by a given nerve root such as C_5. They include M_2, M_3, M_4, M_7, M_{10}, M_{11}, M_{12}, M_{13}, M_{15}, M_{16}, M_{17}, M_{18}, and M_{19}.

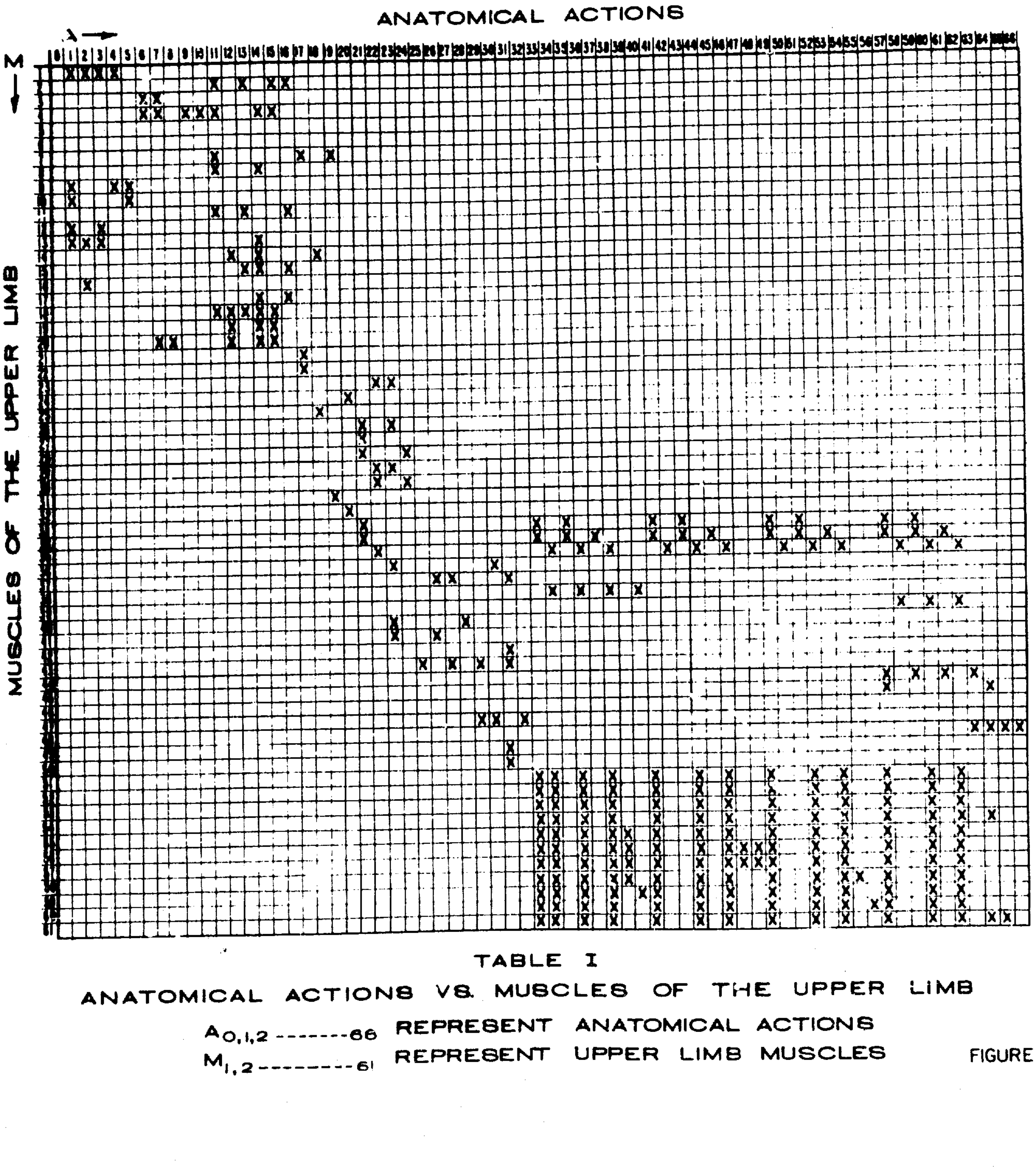


TABLE I

ANATOMICAL ACTIONS VS. MUSCLES OF THE UPPER LIMB

$A_{0,1,2}$ -------66 REPRESENT ANATOMICAL ACTIONS

$M_{1,2}$ --------61 REPRESENT UPPER LIMB MUSCLES

FIGURE 2

T₂	T₁	C₈	C₇	C₆	C₅	C₄	C₃	C₂	C₁	M ↓	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
						X	X			1		X	X	X	X																
				X	X					2												X		X		X	X				
				X	X					3							X	X													
	X	X	X	X	X					4							X	X		X	X	X			X	X					
							X	X		5																					
							X	X	X	6																					
				X	X					7												X						X		X	
			X	X						8												X			X						
	X	X								9		X			X	X															
			X	X	X					10		X				X															
					X					11												X		X			X				
					X	X	X			12		X		X																	
					X					13		X	X	X											X						
		X	X							14													X		X				X		
				X	X					15														X	X		X				
					X					16			X																		
					X					17															X		X				
				X	X					18												X	X	X	X	X					
				X	X					19													X		X	X					
		X	X	X						20									X				X		X	X					

INNERVATING NERVE FIBERS (columns T_2 … C_1); ANATOMICAL ACTIONS (columns A→ 0 … 20); MUSCLES OF THE SHOULDER (rows M ↓ 1 … 20)

TABLE II

NERVE-MUSCLE-ACTION RELATIONS OF THE SHOULDER

A_0, 1, 2 —— 20 REPRESENT ANATOMICAL ACTIONS

M_1, 2 ······ 20 REPRESENT SHOULDER MUSCLES

C1, 2, ··· C_8, T_1, T_2 REPRESENT INNERVATING SPINAL NERVES

FIGURE 3

649

This multiplicity of overlap or redundancy is unique for minimizing the effects of injury. Quantitative force time relationships have not been effectively established at the muscle level to allow us to assign weighting factors to the contributions from a muscle to an associated anatomical action.

Single muscles and their respective anatomical actions combine to characterize regional and joint actions. In the shoulder region, for instance, we think first in terms of the muscles associated with the specific anatomical actions such as flexion, extension, abduction, and adduction. In continuing the generalization process in this region we next consider the shoulder motions of rotation and circumflexion which are derived from the same muscles acting in different combinations and sequences. In applying the same notions to the hand we begin by considering the single muscles associated with finger flexion, extension, abduction, adduction and opposition and then combining these in ways to account for generalized processes such as grasping, holding, and releasing. They, of course, are associated with even more complicated processes associated with performing tasks such as using a screw driver or turning a door knob.

In brief we hope to be able to associate the biomechanical functions with the natural effectors or neuro-motor factors which are responsible for them. One may proceed in man-task study problems in either direction, i.e., he may begin with the knowledge of basic muscular functions and move up the scale to gross movements or he may start with a study of the man-task process in the hope of first identifying useful gross movements and work down to the scale to muscle functions.

A combined mechanical and anatomical orientation appears to have some unique advantages for describing man-machine interaction. For example, we believe that by considering the upper limb as a flexible multi-jointed cantilever with a unique prehensile of grasping device located at the free end, one can develop useful methods for describing physical and physiological factors in relation to man-machine interactions in ways which yield to simplification and measurement.

In our first approach to upper limb biomechanical measurements we are using these anatomical-mechanical notions, i.e., muscle groups associated with hand actions, joint actions, multi-joint actions, liner actions, and combinations of these. Initially we propose to measure only a limited number of these biomechanical or effector functions. For the upper limb we have developed some instrumentation to measure and record force-time histories associated with hand grasping, and flexion and extension actions about the wrist and elbow joints. A brief explanation of this instrumentation is given in Appendix B.

We extend this rationale to include the opposite upper limb, the two lower limbs (considering them also as multi-jointed cantilevers but in terms of their natural anatomical-mechanical functions of weight-bearing and ambulation), and the other effectors (head and neck, and trunk). We believe that this approach will result in useful descriptors for man-task/machine interactions in a manner suitable for describing and assessing changes in performance due to disability regardless of cause.

CLINICAL SUBSYSTEMS. One of the reasons this problem is of interest is that it requires investigations not made in the past. For example, it is known that damage to the cardiovascular subsystem in terms of blood loss is likely to be fatal if the value exceeds approximately 1600 cc to 1700 cc within a short period assuming no replacement. While this is an important upper bound, the effect on one's ability to perform due to hemorrhage of lower orders has not to our knowledge been studied. Hence it is of primary interest for us to collect hospital data on patients who may suffer various degrees of blood loss and to measure the effects on several representative effectors.

Damage to the respiratory subsystem may be assessed in terms of the degree of pneumothorax, rate of O_2 - CO_2 exchange, or, perhaps, in terms of respiration rate and depth. The chosen effectors would be measured at approximately the same time that the physiological measures are taken.

Descriptions of levels of damage may be more difficult for the neurological, muscular, and skeletal subsystems. Presently we are only considering two levels of damage for any substructure (a muscle, a nerve, or a bone), namely, none or complete. For the present we do not expect to make special studies on the gastro-intestinal, genitourinary and endocrine subsystems other than to observe the routine hospital events associated with them. Their respective blood losses are to be considered, however, but viewed as cardiovascular subsystem deficits.

Additional measurements which are not always routine may be added if the results of some of the present research being done elsewhere on traumatic injuries indicates it. For example, one of us is engaged in human shock research in which certain relations between clotting time and levels of shock have been generated for 30-odd humans who were in shock due to blood loss.(3) Interesting observations of adrenal function in combat and wounded soldiers have been made to some extent by others. (4) (5) These suggest possibilities for associating "stress" levels and performance.

It is with these ideas in mind and the notion that feasible relations between the effectors and the body's clinical subsystems do exist that we wish to construct useful data collecting procedures. Hopefully, early insight following some of this data collection will allow us to get some ideas concerning these relations. Since we expect such potential relationships to change with time as a result of injury, we believe that data collection will have to be made at various time periods throughout the clinical course.

Having set up these ideas as guidelines for the data collection process, we must consider some of the practical aspects of the problem. It is important to review the clinical procedures used in evaluating and treating wound cases in hospital accident rooms, operating rooms, and recovery wards in order to appraise the available and/or recorded data in terms of type and quantity. Another point of interest concerns the logic for selecting and running certain clinical tests and not others. There may be many cases in which certain useful clinical data could be made available for specific purposes such as ours but which may not be sought ordinarily by a clinician inasmuch as these data do not in his judgment add any useful information for his purposes. It is also important to be sure that the acquisition of data from a distressed patient does not interfere with his well-being.

Medical Records and Clinical Data. Hospital medical records reflect traditional procedures for recording information and events associated with the professional care and treatment of patients. While time and space preclude any extensive discussion of the meaningfulness of the comprehensive clinical and laboratory data as interpreted by physicians and/or other interested discipline representatives, a few observations are presented.

Clinicians evaluate patients, their care and theraphy, according to the description and history of the presenting complaint as well as other pertinent past patient (and family) history, physical examination, laboratory test results, and progress evaluations. In general clinical information is classified either as subjective or objective information. Subjective information is associated with what the patient or others tell to the examiner. Examples of this would include the patient's interpretation of local or general muscular weakness, walking difficulty, ------ dragging toe of shoe, stumbling or falling, sphinteric disturbances (inability to hold urine), changes in local or general sensation ---- fixed or radiating pain, temperature, tactile discrimination, deep sensation (muscle, bone, vibratory sense), and abnormal sweating. Objective information concerns what the examiner learns from his own observations such as range of limb movement, contractures, diminished size, strength of muscles against resistance, tremors, etc. Thus, it can be seen that except for laboratory tests and some clinical items such as blood pressure, pulse rate, and the electrocardiogram, quantitative measures are minimal in the traditional records.

Records of emergency cases are often initiated in the hospital's accident room and accompany the patient throughout his hospital course. A form of time history of his care, diagnosis, treatment, and progress is recorded for permanent file. We will consider a few examples of hospital records.

An accident room record is shown in Figure 4. This 21 year old male's record shows very little information. This is not unusual in the typical busy accident room. The only history and physical data recorded in this case are blood pressure and pulse. The immediate treatments and the results of laboratory blood tests were recorded. It is highly likely that additional observations of blood pressure and pulse were made with the passage of time but not entered in the record. A certain amount of physical examination was probably performed and not recorded. While the admission time and date were recorded, a detailed history of the presenting complaint is not shown. Such information might have included approximate time of the shooting incident, estimates of external blood loss, and conscious or unconscious behavior of the patient until arrival at the hospital, and ballistic factors, e.g., type and calibre of gun, distance between firer and patient, and angle of target to firer.

UNIVERSITY HOSPITAL
CENTRAL CITY

DATE 9-24-62 ARRIVAL 7:40 P
DEPARTURE A.P.M.

NAME Abbott Robert AGE 21 SEX M COLOR C BIRTHDAY RELIGION

ADDRESS 1022 King Rd PHONE

NEAREST RELATIVE ADDRESS

TIME OF ACCIDENT A P M AUTO COMP. B.C. M.C. OTHER

EMPLOYER & ADDRESS

INSURED BY COMPANY PHYSICIAN

ARRIVED VIA Ambulance E.R PHYSICIAN Dr Jones

COMPLAINT 2 Gunshot wounds One in chest and one in back

DATE & TIME OF ONSET OF ILLNESS OR INJURY Store - Pratt + 10th St - Amsterdam Co.

ORDERS	PHYSICIANS RECORD Police Officer Brown
On adm	BP 130/86 P-100
1000 cc Dextrose 5% in Saline	
500 cc Dextran	
Plasma to follow (300 cc)	
Tetanus toxoid 1cc IM	
✓ Hb 14.5	
✓ Hct 42	
✓ WBC 9,600	
✓ T-match (6 units)	
✓ T + Hold	
✓ BUN	
✓ CO_2	
✓ Sugar	

FIGURE 4

DISPOSITION DR'S SIGNATURE M.D.

Another accident room record is shown in Figure 5. This 15 year old male with multiple gun shot wounds has much more information on the physical examination than the previous case. Another example is shown in Figure 6. One must judge that in this latter case the patient was moved without delay to a ward bed for further control and treatment.

In examining a number of similar records, we believe that information which is useful for the clinician is, for our purposes, both inexact and insufficient. It is not enough for us to know that "the neurological and extremities are negative." It is our opinion that we need to have some measured data associated with the neuro-motor activity of the effectors. For example, the prehensile or grasping function of an upper limb although not injured may be weakened due to a severe blood loss in an artery located in another part of the body. Neurologists and neurosurgeons have a variety of clinical tests which they use based on certain known relationships in neuro-anatomy and physiology. As a matter of interest, it has been stated that in no other branch of medicine is it possible to build up a clinical picture so exact as to localization of pathological conditions.[2] However, if such tests are not performed and/or the results not recorded for patients of interest to us, possible relations between the effectors and clinical subsystem changes are not attainable.

It may be noted also that many classical neuro-motor tests are not strictly quantitative as evidence by such observational descriptions as "partial paralysis" or "some loss of sensation." Conclusions drawn from a variety of tests may vary considerable as they depend on the judgment of different examiners and the impressions expressed by patients.

In the ward the situation is quite different from that found in the accident room. Here the patient is in an environment in which physicians, nurses, and technicians are available to care and control the patient under more favorable conditions. Records are kept in much more detail. The frequency of observations and the application of procedures and treatments are, of course, greater when the patient is in serious condition. Except for emergency procedures, these observations and treatments for patients under intensive care are usually performed about every one of two hours. It takes this long for many therapies to take effect.

UNIVERSITY HOSPITAL
CENTRAL CITY

659

DATE 1/29/63

ARRIVAL 3°° P.M.

DEPARTURE A.P.M.

NAME Jones, William AGE 15 SEX M COLOR N BIRTHDAY RELIGION

ADDRESS 102 N. Charles St Jamestown PHONE None

NEAREST RELATIVE father – Joe ADDRESS Same

TIME OF ACCIDENT A.P.M. AUTO COMP B.C. M.C. OTHER

EMPLOYER & ADDRESS

INSURED BY COMPANY PHYSICIAN

ARRIVED VIA Jamestown Rescue Amb 42 E.R. PHYSICIAN

COMPLAINT Gunshot wound (R) arm & Hand

DATE & TIME OF ONSET OF ILLNESS OR INJURY

ORDERS	92-20 150/90 PHYSICIANS RECORD
Hct, Hgb, WBC,	
Type & Cross	Pt brought into AR c̄ bullet wounds
4 units	Drawn
Bun, Sugar	Inspection: Three openings – two in (R) upper
500 cc D/W	arm – in biceps muscle
	Third in (R) Chest wall about level of
	T5 and axillary line
#14	
Nasogastric tube	Sensorium Clear
inserted	Vital Signs BP 140/80 pulse: 80 resp 16
	Chest
	On (R) side thought to be decreased in (R) hemithorax
	with dullness to percussion over both lung fields
	to about level of T6
	No Subcut Emphysema
	Cor Sinus Rhythm No (m)
Tetanus 1cc	Abd B Sds present
	Tenderness in epigastrium
	No liver impairment
	Extrem: palpable pulse in (R) arm
	With no sensory deficit
	Imp: Bullet wound
	R/O (R) hemopneumothorax

FIGURE 5

DISPOSITION

DR'S SIGNATURE [signature]

UNIVERSITY HOSPITAL
CENTRAL CITY

601

DATE 11-12-62 ARRIVAL 6 A.P.M.
DEPARTURE A.P.M.

NAME Brown, Walter AGE 52 SEX M COLOR W BIRTHDAY 9/?/? RELIGION
ADDRESS 1505 N. Bond St. PHONE
NEAREST RELATIVE Mabel Brown (wife) ADDRESS same
TIME OF ACCIDENT A.P.M. AUTO COMP B.C. M.C. OTHER
EMPLOYER & ADDRESS
INSURED BY COMPANY PHYSICIAN
ARRIVED VIA Pittsville Vol. Amb. E R PHYSICIAN Dr. Jackson
COMPLAINT was robbed + shot - left chest @ intersection of Rts. 9 + 49
DATE & TIME OF ONSET OF ILLNESS OR INJURY 2 miles west of Pittsville

ORDERS	PHYSICIANS RECORD
	Gunshot wounds left chest with pulmonary hemorrhage
	Xray = fragmented bullet in substernal chest wall on left with parenchymal hemorrhage mid lung field. No Pneumothorax
	Cauterize shallow wound
	admit for observation
	T. J. Johnson

FIGURE 6

DISPOSITION DR.'S SIGNATURE M.D.

Whenever a patient undergoes a special procedure, operation, or, if an autopsy is performed, a detailed account of the events is made which becomes a part of the permanent medical record. The first detailed physical examination is often not performed until the patient arrives in the ward following evaluation and treatment in the accident room. Additional laboratory tests and special diagnostic procedures are usually initiated after admission to the ward. A considerable amount of clinical information is recorded which can be useful for our purposes. Large gaps invariably exist, however, due to reasons mentioned previously as well as the thoroughness of the work and recording of interns and residents. Patient care is naturally oriented toward healing and care processes. This is to say that attention is on the progress of the patient's subsystems and behavior as a whole while he is in a resting state. He is not considered as a component in a man-machine system.

A Given Wound Patient Record. In order to get an idea of some of the events which may take place in an accident room and a ward, a medical record of a patient admitted with a bullet would is briefly discussed.

Accident Room.

1st. Day: A 56 year old man was shot in the chest with a 32 Calibre pistol "at close range." He was taken to a hospital and arrived at 5:36 a.m. on the day of the accident.

On admission to the accident room, the following were studied or measured immediately and recorded:

Blood Pressure
Pulse rate
Heart sounds
Breath sounds
Hemoglobin
Chest X-ray

Treatment was also ordered and initiated immediately. It consisted of a fluid replacement program according to the following sequence:

Circulatory expander started (500 cc)
Whole blood (500 cc)
Saline (500 cc)
Dextrose in water (1000 cc)

The patient was also sedated and prevented from taking any fluid or food by mouth (as a precaution in case of the need for operation).

The remaining accident room events which occurred according to the medical record were:

Follow-up blood pressure, pulse rate, and respiration checks at 8:00 am, 8:25 am, 9:00 am, 10:00 am, 12:00 noon, 4:00 pm and 6:00 pm.

The patient was admitted to the ward at 6:30 pm.

Ward The admitting intern made the first detailed investigation at 7:30 pm. It is shown as follows:

Intern's Admission Note: "At 4:00 am patient was shot by wife at close range with a .32 Calibre pistol."

Bullet entered left chest and escaped via left back.

Patient did not lose consciousness.

There was no chest pain, coughing increase, or hemoptysis (coughing of blood).

Physical Examination: Well developed, well nourished, alert, cooperative, no distress (two?) gun shot wounds, no powder burns.

Entrance: approx. 5 mm left anterior axillary line, 8th interspace

Exit: Post axillary line, 10th interspace

Chest: Expands well bilaterally
Lungs - left posterior: basilar dullness, about 3 cm above the base; tactile fremitus (a vibration imparted to the hand placed on the chest) somewhat impaired left posterior base.

Blood Pressure: 120/60 - no murmurs

Abdomen: Slightly distended
Generalized superficial tenderness

Neurological and Extremities - negative

Impressions: (1) gunshot wound, left chest
(2) left hemothorax
(3) rule out perforated bowel or spleen
(4) Cardiac arrythmia - bigemini (paired pulse beats)

Additional medications were given at 8: 00 pm.

2nd. Day: Throughout the second day the following vital signs were observed every hour.

Blood pressure
Pulse rate
Respiration rate
Temperature

Nourishment was still given by intravenous fluids.

3rd. Day: Vital signs were observed every four hours instead of the hourly schedule of the previous day. A liquid diet was prescribed in place of the intravenous fluids.

After three more days the patient was placed on routine care. Except for a thoracentisis (extraction of fluid from the chest cavity via needle and syringe) on the 8th, hospital day (250 cc of bloody fluid was removed), the patient recovered and was discharged on the 12th. day.

Inasmuch as no operating room procedures were required, there was no opportunity to get a pathological description of the internal path of the bullet. In operative cases wound tract descriptions are usually available in varying amounts of detail. In many instances some estimates of hematoma (trapped blood) volume, degree of bone fracture, and other gross abnormalities are noted in the operating room reports. In some cases, missiles or missile fragments are removed whereas in others the additional risk associated with the removal process is such that the fragment(s) are left in the body. X-ray studies, of course, are informative in such cases.

In this case the calibre of the gun was known and some indication of distance between gun and target was given. Occasionally additional information is given which is quite important such as an account of the patient's response following the wounding process (e.g., "the patient ran to the doorway and down the steps before fainting," "the patient screamed and fell to the floor unconscious," etc.). Estimates of blood loss, if noticed, and elapsed time between the shooting incident and arrival at the hospital accident room are of basic interest to us.

From a cursory review of medical records concerning missile wounds, one is impressed with the fact that the large majority of such cases are non-lethal. In survival records it is observed that patient handling, treatments and control vary considerably depending on the case. However, the general care patterns do not appear to be radically different. The type and amount of data recorded on the other hand varies widely. Lethal cases, of course, usually have an autopsy report providing one was permitted by next of kin or ordered by a medical examiner.

Remarks on Hospital Medical Records. In brief several points concerning information and medical records for patients entering the hospital with gunshot wounds are presented.

1. There is considerable variation in type and content of reported information. This is particularly so in the accident room portion of the record.

2. While it may be assumed that a multitude of observations on a patient's subsystems, appearance, manner of behavior, speech, etc., are made by clinicians which may influence diagnosis, treatment, and control decisions, such information is not usually so extensively recorded.

3. Patient information is difficult to handle and retrieve inasmuch as it is not organized according to subsystem variables over time. Some of the measures which are considered as standard include

vital signs such as temperature, blood pressure, pulse rate, and respiration rate. These are usually plotted on graph paper by the nurses. Clinical descriptions and observations are often matters of judgment which do not yield to measurement under present state-of-the-art.

4. Facts and observations pertaining to shooting incidents and especially anatomical, physiological, and psychological events which occur between the incident and arrival at the hospital are usually very brief when included.

5. What is or is not recorded probably depends on the amount of training, experience, and judgment of the respective hospital staffs and residents.

6. Usually no psychological or psychiatric examinations are made on patients suffering from missile trauma. However, this is not done as a routine on traumatized patients in general.

Proposed Data Collection. We have described what we wish to do, a method for going about it, the type of data we think we need as a first approximation, and the entent of its availability in the (emergency) hospital. It is apparent that our requirements call for more complete and more frequent clinical observations and measures in addition to the new information in support of our special interests. This information may be thought of as:

1. Ballistic information (Appendix C).
2. Postwounding behavioral information (Appendix D).
3. Pathological information in terms of
   a. Wound tract information (Appendix E).
   b. Clinical subsystem information (Appendix F).
4. Effector measurement information (Appendix G).
5. Special studies for
   a. Information on normals (Appendix H).
   b. Anthropometric information (Appendix I).

It is appreciated that there are many practical problems in organizing and administering an effort of this kind. Since wounded patients are classified as surgical patients it seems logical that the heads of the participating surgical divisions and their supporting resident and nursing staffs should be sufficiently interested in a program such as this. Modifications in the program of data collection are anticipated once sufficient feedback information is developed.

Hopefully, as mentioned in the early part of this discussion, useful relations may be forthcoming early in the process between missile parameters and subsystem changes and between subsystem changes and the (biomechanical output) effectors. Our mathematical and statistical colleagues tell us that this effector-wound relationship is a stochastic one for two reasons. First, there is a distribution of wounds within a given class or category and hence there is a random variation with a certain distribution function of the corresponding effector. Second, the effect of even identical wounds sustained by different individuals also possesses a certain distribution.

There are many obvious problems involved in analyzing the data to be collected. Unfortunately there are no known data previously collected from which normals or non-injured standards can be established. It is, of course, impossible to obtain pre-injury effector and subsystem "normals" for the hospital cases. Since we are primarily interested in males of military age, the choice of civilian accidents for data collection will be so restricted. However, useful information may be gathered for cases in which injured patients would return following complete recovery. Otherwise independent samples will have to be taken from normal personnel in order to obtain pre-injury distributions of the appropriate subsystem and effector parameters. There is also a need for anthropomorphic data. Therefore they are included under special studies (Appendix I).

Sources of Error in the Procedures. The following are some of the more apparent errors which we are told should be considered:

1. Errors in determining the group normal for each effector measure inasmuch as (a) subjects cannot be evaluated before wounding and (b) a statistical group normal must be established with which to compare the individual disability.

2. Error arising from individual deviation from the group normal.

3. Error arising from interaction between the effectors. These errors are minimized by attempting to isolate the effectors, but they cannot be totally eliminated.

4. Errors arising from variation in the measuring and assessing techniques of the medical and technical evaluators.

5. Errors in the mechanical devices used to quantitate the efficiencies and errors associated with positioning the derives on the subjects.

6. Clinical laboratory errors.

7. Multi-clinic errors especially where more than one hospital is chosen for data collection.

SUMMARY. In an initial effort to quantify changes in human performance due to missile injury, we propose to collect certain clinical information, associated missile ballistic factors, and measure a select group of neuro-muscular responses. Such responses are chosen as potential descriptors for man-machine performance problems. As mentioned, a man-machine system model for incapacitation evaluation is being developed.[1] The acquisition of detailed human clinical data is essentially restricted to that available in the emergency hospital. Hopefully, such information will permit the establishment of new and useful relations early in the process.

REFERENCES

1. A Two-Person Zero-Sum Game Application for Incapacitation Evaluation, W. H. Kirby, Jr., and C. Masaitis. Ballistic Research Laboratories (in process).

2. Practical Neurological Diagnosis, by R. G. Spuring, Charles C. Thomas - Publisher, January 1950.

3. The Effect of Hemorrhagic Shock on Clotting Time in Humans, Attar, A., Kirby, W. H. Jr., Masaitis, C., Cowley, R. A. (in process).

4. Studies of Adrenal Function in Combat and Wounded Soldiers, Howard, John M. et al. Annals of Surgery, March 1955.

5. The Hemostatic Response to Injury. A Study of the Korean Battle Casualty. Annals of Surgery, March 1955, Scott, Russell, Jr., Crosby, W. H.

APPENDIX A

I. Anatomical Actions

Subscript codes for anatomical actions of the upper limb are given only for the shoulder region.

Subscript Code	Anatomical Actions
A_0	Actions of the Parts other than upper limb
A_1	Rotation of scapula
A_2	Adduction of scapula
A_3	Raising of scapula
A_4	Lowering of scapula
A_5	Moving of scapula forward
A_6	Moving of shoulder forward
A_7	Lowering of shoulder
A_8	Drawing of shoulder backward
A_9	Raise shoulder
A_{10}	Adduct shoulder
A_{11}	Flexion of arm
A_{12}	Extension of arm
A_{13}	Abduction of arm
A_{14}	Adduction of arm
A_{15}	Medial rotation of arm
A_{16}	Lateral rotation of arm
A_{17}	Flexion of forearm
A_{18}	Extension of forearm
A_{19}	Supination of hand by forearm
A_{20}	Pronation of hand by forearm

APPENDIX A (cont'd)

II. Muscles of the Upper Limb

Subscript codes for the muscles of the upper limb are given only for the shoulder region.

Subscript Code	Name of Muscle
M_1	Deltoid
M_2	Trapezius
M_3	Subclavius
M_4	Pectoralis major
M_5	Sternocleidomastoid
M_6	Sternohyoid
M_7	Biceps
M_8	Coracobrachialis
M_9	Pectoralis minor
M_{10}	Serratus anterior
M_{11}	Supraspinatus
M_{12}	Levator scapula
M_{13}	Rhomboid minor
M_{14}	Triceps
M_{15}	Infraspinatus
M_{16}	Rhomboid major
M_{17}	Teres minor
M_{18}	Subscapularis
M_{19}	Teres major
M_{20}	Latissimus dorsi

APPENDIX A (cont'd)

III. Spinal Nerves of the Upper Limb

Subscript codes for the spinal nerves which innervate the upper limb muscles are given in association with the shoulder region only.

Subscript Code	Spinal Nerve
C_1	First cervical nerve of spinal cord
C_2	Second cervical nerve of spinal cord
C_3	Third cervical nerve of the spinal cord
C_4	Fourth cervical nerve of the spinal cord
C_5	Fifth cervical nerve of the spinal cord
C_6	Sixth cervical nerve of the spinal cord
C_7	Seventh cervical nerve of the spinal cord
C_8	Eighth cervical nerve of the spinal cord
T_1	First thoracic nerve of the spinal cord
T_2	Second thoracic nerve of the spinal cord

APPENDIX B

Brief Description of Instrumentation for Measuring and Recording Force-Time Histories for Hand Grasping and Flexion and Extension in the Wrist and Elbow

The device is a bipod-mounted pistol grip (4-1/2 x 2 x 1 inch) spring loaded with three internally mounted piezoelectric crystals. It is expected that this will be a measure of grasping ability for sudden short grasps and/or prolonged holding. The lead zirconatetitinate crystals used in the prototype model have been dead-weight tested to 500 pounds with no appreciable non-linearity in electrical response.

Output from the pressure transducers is fed into a battery operated charge amplifier which is also used to drive a 2-channel pen recorder. The amplifier and recorder are housed in a portable carrying case weighing approximately 10 pounds and with an overall dimension of 12" x 8" x 6".

For flexion data, removable swivel screws are attached to each end of the hand grip. The swivel screws rotate in a U-clamp which, in turn, is affixed to the table or a rigid surface. The felt-backed nylon cords which are affixed to the pistol grip are grasped in the palm of the hand or slipped over the wrist. Pulling the cord orients the measuring device parallel to the direction of the flexion force. Removing the cord and pushing on the face of the grip will produce extension measurements.

APPENDIX C

Ballistic Missile Information

Weapon

Caliber: ____________

Type: ____________

Brand Name: ____________

Disposition of Weapon: ____________

Missile

Recovered: Yes ______ No ______

Brand Name: ____________

Disposition of Missile: ____________

____

Weight: (grams or grains) ____________

Dimensions: ____________

Situation

Distance between weapon and victim: ____________

APPENDIX D

Behavioral Information

Activity at moment of wounding:
(check)

Running ________

Standing ________

Sitting ________

Lying ________

Other (explain) ________ ______________________________

Remained conscious:

Yes _____ No _____

If no, was unconsciousness,

Immediate? _____

Later? _____

If later, about how long __________ mins.

Conscious responses:

A. Psychological:

Highly excited? ______

Stayed calm? ______

B. Physical:

		Remarks
Started fighting	______	______________________
Ran	______	______________________
Walked	______	______________________
Stood up	______	______________________
Fell to ground	______	______________________

APPENDIX E

Clothing and External Wound Tract Information

Description of Clothing Damage on Victim:

	Material	Description of Damage (Hole size, etc.)
Overcoat:		
Jacket:		
Shirt:		
Undershirt:		
Pants:		
Shorts:		
Other:		

Wound Extrance:

	Location(s) of Penetration(s)	Dimensions of Penetrations
(a)		
(b)		
(c)		

Wound Exit:

	Location(s) of Exit Hole(s)	Dimensions of Exit Hole(s)
(a)		
(b)		
(c)		

Clinical Subsystems

Name or Code No. ______________ Age: ____ Race: ________

Hospital No. ______________ Sex: ____

Date of Admission: ________

Date of Discharge: ________

Accident Room:

Time of Shooting: ______________

Time of Arrival: ______________

Clinical Observations	TIME Acute Phase: Every hour	Control Phase: Every 4 hours	Recovery Phase: Once a day
Temperature:			
1. Cardiovascular Subsystem:			
Blood Pressure			
Pulse Rate			
Heart Sounds			
Electrocardiogram			
Laboratory Measures:			
Hemoglobin			
Hematocrit			
Red Blood Count			
White Blood Count			
Eosinophil Count			
Coagulation Time			
Sodium			
Potassium			
Chloride			
Proteins (Total)			
Estimated Blood Losses			
Volume Replacement			
Special Studies			

Clinical Observations (Cont'd)

2. Respiratory Subsystem:
   - Rate of Respiration
   - Breath Sounds
     - Right lung
     - Left lung
   - Palpation
     - Right lung
     - Left lung

   Laboratory Studies
   - X-Ray
   - Special Studies

3. Neurological Subsystem:
   - State of Consciousness
   - Mental Responses
   - Sensory Responses
   - Motor Reflexes
   - Cranial Nerve Responses

   Laboratory Studies
   - X-Ray
   - Electroencephalogram
   - Spinal Fluid

   - Special Studies

4. Skeletal Subsystem:
   - Fracture(s)
   - Dislocation(s)

   Laboratory Studies
   - X-Ray

   - Special Studies

<u>Clinical Observations</u> (Cont'd)

5. Muscular Subsystem:
   Penetration/Perforation
   Edema or Hematoma
   Flacidity or Rigidity

   <u>Laboratory Studies</u>
   X-Ray

   Special Studies

6. Gastrointestinal Subsystem:
   Vomiting
   Hematemesis
   Condition of Bowel Sounds
   Organ Palpation

   <u>Laboratory Studies</u>
   X-Ray

   Special Studies

7. Genitourinary Subsystem:
   Presence/Absence of Blood in Urine
   Urine Volume
   Urine Analysis

   Special Studies

8. Endocrine Subsystem:

   Special Studies

PROCEDURES:

THERAPIES:

APPENDIX G

Effector Measurements (Upper Limbs Only)

	Admission	+4hrs.
Hand Grasping:		
Sudden maximal effort:	______	______
Prolonged grasp (loading to be specified):	______	______
Grasping follower exercise (to be specified):	______	______
Wrist Flexion:		
Sudden maximal effort:	______	______
Prolonged flexion (loading to be specified):	______	______
Wrist Extension:		
Sudden maximal effort:	______	______
Prolonged flexion (loading to be specified):	______	______
Elbow Flexion:		
Sudden maximal effort:	______	______
Prolonged flexion (loading to be specified):	______	______
Elbow Extension:		
Sudden maximal effort:	______	______
Prolonged elbow extension (loading to be specified):	______	______

APPENDIX H

Information on Normals

The formats anticipated for normals are duplicates of those used in Appendix F and Appendix G, i.e., wherever clinical information is sampled there is an assumed need for normal values. These values will either be drawn from similar populations and/or from those victims who survive the injuries and are considered to be normal again.

APPENDIX I

Anthropometric Information

Age: ______ Sex: ______ Weight: ______ Race: ______

Physical Measurements:

Height: ______

Weight: ______

Head circumference (forehead level):	______
Vertical distance (top of head to bottom of mandible):	______
Neck length (tip of hyoid bone to suprasternal notch):	______
Neck circumference (at midpoint of neck length):	______
Chest circumference (at nipple line):	______
Abdomenal circumference (umbilical level):	______
Hip level circumference (level of iliac crest):	______
Sternal notch to symphasis publis:	______
Spinous Process of C-7 to coccyx:	______
Width of shoulders (acromion to acromion):	______

	Right	Left
Acromion to radial epicondyle:	____	____
Midarm circumference:	____	____
Radial epicondyle to radial styloid process:	____	____
Radial styloid process to tip of middle finger:	____	____
Midforearm circumference:	____	____
Wrist circumference:	____	____
Anterior iliac spine to upper patella:	____	____
Upper patella to sole:	____	____

APPENDIX I
(cont'd.)

	Right	Left
Midcalf circumference:	______	______
Circumference at patella:	______	______
Midleg circumference:	______	______
Ankle circumference:	______	______
Length of foot:	______	______

PROBLEMS IN THE DESIGN OF STATISTICS - GENERATING WAR GAMES

William H. Sutherland
Research Analysis Corporation, McLean, Virginia

A newspaper headline last week said "Helicopters Crash in War Games." Now since I am billed as presenting a problem to you under the title "Problems in the Design of Statistics Generating War Games," the headline makes me hasten to tell you that the war games I'm concerned with are not at all like the newspaper version, and that whatever statistics that kind of war games generates in the form of crashes, such statistics have little in common with the kind I wish to talk to you about.

The kinds of war games that are of concern are played for Army research purposes. They are two-sided, somewhat formal exercises-played indoors using maps and often using computers. They are of a size and complexity measureable in tens of man months of play and analysis effort (an expensive kind of effort as operation research studies go, but seldom, I suspect, having costs comparable with the kind of war games in the newspaper headline). In our games records are kept of the details of play, and from these records statistics are derived. The statistics of the title, then, concern not real-world helicopters or troops or guns, but do concern the helicopters or troops or guns which the gamers have in mind's eye as they play. Battle results are found by applying rules, not bullets, and sometimes the battle results are made to depend on random numbers.

The players usually have a good deal of freedom of action tactically, and it is in this sense that I take the liberty of considering war games to be experiments. (Certainly the games do have this in common with experiments: we never know how they are going to come out.)

However, they are unlike most experiments in that one would not ordinarily expect to be able to repeat the initial conditions with strict exactness. If one were to try, with say the same players, they would no longer be in the dark about their adversary, because of what they had learned in the first game. If one tried a second set of players, they would necessarily have different tactical experience inside their heads to begin the game with.

Now after telling you that the tool--war gaming--is one that in one sense defies replication, let me ask you about a problem that war gamers face, which it seems to me, can be discussed partly in terms of what replication would show if indeed replication were possible. The problem presented itself in the course of a Research Analysis Corporation study on the use of war gaming as a research tool. It is: when should the war game designer use random numbers in the game, and when should he avoid using them? As you will see, the guidelines we have are only qualitative and what appear to us to be common sense. It would help if we had better guidelines, and this is the problem which I ask the panel to consider. So much for the introduction. What I have to say is in three sections: (1) the reasons for using random numbers; (2) the appropriateness of using average values versus random numbers; and (3) the effects of random numbers on interpreting game results.

1. REASONS FOR USING RANDOM NUMBERS. Games often make use of random numbers as a way of deciding details of combat. The use of random selections from previously determined or estimated probabilities serves two main kinds of function: to represent the chance nature of warfare and to keep players from having an inappropriate knowledge of their opponents.

As for any model, the chance nature of warfare can of course be only imperfectly represented. Only a few of the most important chance events can be selected to be part of a game. I suppose that this statement will seem a little like Alice-in-Wonderland to this audience for many of the papers presented in this conference the problem seems to have been to nail down chance events which wander uninvited into the problem - here we are in effect dragging them in by the heels. Of these chance events the use of random selections can be considered as a means of at least roughly representing the consequences of groups of causal sub-events. These sub-events result in the "main" event which is being represented. The sub-events--the direct causes--may be impractical or impossible to know, or they may in the game be unnecessary to know. As an analogy, the causes for the small deflections of a bullet fired from a gun in a test stand may well be impracticable to know; they depend on such matters as changes in air density along the patch of successive bullets or slight and unknown asymmetries in the loading of the powder. For many practical purposes, it is not necessary to know the causes. So also in war games a spread of outcomes for an event can be used without specifying the causes. Thus random numbers can be used in games to represent events whose causes are not thoroughly understood or sometimes to simplify considerations which would otherwise be in too great a detail considering the scale of the game. For example, in a recent game the outcome--success or failure--of a minor raid against a logistics installation behind enemy lines was represented by a simple draw. Detailed consideration of the complex of factors for and against the raiders was not appropriate considering the probably minor effects of the raid on the overall game outcome.

The second class of use for random choice from distributions is to keep the opposing players suitably in the dark as to the exact capabilities of their opponents, and thus add to the realism of the game. Without random factors a player might be able to work the model or formula backwards and find enemy strengths. So randomness makes the players' decisions more like what they would be in real life.

Parenthetically, it has been observed that such use of random numbers, by making the incidents of a game less predictable, makes play more interesting to the participants. This contributes to the intensity and involvement which seems to characterize games, and which, one hopes, may occasionally result in an otherwise routine tactic being replaced with a brilliant one which may alter the concept being studied.

2. <u>APPROPRIATENESS OF AVERAGE VALUES VS RANDOM SELECTIONS.</u> Suppose that in a particular theater rain, if it occurs, has a strong effect on operations, but that it occurs only say three percent of the time, and without any repetive pattern (i.e., in a way that is reasonably represented as random). A single game is being played, and the random number generator indicates that it is to rain on some important day of the game. Is it appropriate to play that day according to rainy day rules? To do so would make the play for that day "non-typical"; to disregard the weather could be criticized as being unrealistic. So one subquestion to ask about random numbers is--when is it appropriate, for matters like weather effects to use "typical" values (i.e., dry weather); when to use average values (i.e., 97% dry); and when should the variations be randomly selected?

No hard and fast rules come to mind immediately, but certain observations on the subject can be made. The first is elementary. It is that there is little point in using the variations--which of course complicate the game--unless the factor itself is important to the game. If in the example rain had small significance, average values for its effects would certainly be sufficient.

Beyond being important in its general effect on the game output, though, further aspects of its importance need to be examined. One is whether the variation itself has an impact. If the effect being considered does not tend to do what I call "weight" the game results, then there may be no point in using any but an average value even though the overall effect is large.

Let me explain. One concept of the reason for playing war games, as opposed to say OR analytic studies is that the play permits examination of certain interactions that none of the other methods seem able to examine. Specific aspects of the interactions between

weapons,
tactics, and
environment,

can be studies in a combined manner in a game.

Let us take an example. As part of a recent war game at Research Analysis Corporation two competing antiaircraft systems were compared in the role of defending against armed helicopters. The particular tactics used by the helicopters included hiding behind available ridges and then coming in fast. The particular hilly terrain of the game limited the range at which the helicopters could be acquired as targets. For these tactics, then, and this terrain the more sophisticated and expensive weapon was not usable at the long ranges at which it was effective. The less complex weapon did nearly as well. The statistic used for the comparison was simply the number of helicopters shot down per helicopter sortie. In general, then, the game may be looked at as a method of examining the interactions between the weapons and the terrain, and the weapons and the tactics, or for that matter, any of the three as affected by the other two. Quantitative weights are implicit in the play of the game and can be thought of as contributing to the relative use and relative effectiveness of the weapons or tactics under the game conditions. The game provides information and insights concerning such interactions, on the basis of such implicit weighting. The combining of tactics, weapons capabilities, and surrounding, may thus be thought of as being a product or output beyond the previous knowledge of the individual inputs. These are the experimental results with which we work. The game, in my viewpoint, is a particularly suitable and unique tool for Army use, partly because it can study such interactions.

To return to random numbers, there is, I suggest, little to be gained in using them to produce variability in some computational input or output unless such a range of numbers is tied to this weighting. The nature of the tie-in is, as far as I know, an unexplored area and one to which you may be able to suggest approaches.

3. INTERPRETATION OF GAME RESULTS - NUMBER OF RANDOM NUMBERS. But in order to do so with better insight, we might also look at the effects of using random numbers on the interpretation of game results. As I indicated, any single play would have to begin a little differently from any other play, and would then continue differently, even though the general circumstances of the game were similar. This comes about because the decisions of (all too) human players are involved, but also because of random number use. While the overall variability in results, which makes interpretation difficult, cannot really be separated into the two causes, we can for the time being assume that the players could be given only limited choices, and we can concentrate on the variability caused by random numbers. In effect this is what does happen in some computer simulations. Consider the number of times a random number is chosen for a particular purpose in the course of a game. In practice this varies greatly from one game to another. Some use random selections literally thousands of times in one play: others have few, and one two-sided exercise that was called a game did not use any. Let us examine a little the consequences of the use of different numbers of random selections. If the purpose, for example, was to find the battle outcome for a single highly important battle, and the random choice was made just once, then two quite different possible results could happen. In statistical terms, the spread of potential results would be large. (Incidentially what would be learned from such a game would be little.) On the other hand, if a long series of battles were fought, each of which

was equally important, and in which the probability of the different possible results did not change between battles, then statistically the relative spread of possible overall results would be small. Of course war games really do not present a picture of a large number of exactly similar events decided on a probability basis in the way described. Still, they presumably behave somewhat as if they did, and thus to a limited extent the spread of results can be discussed as a statistical matter. The limitations include this: that the measure of outcome is indeed a statistic--"winning" or "losing" is probably not such a measure. Secondly, for some mathematical considerations the measure must be an average. Certainly not all the statistics with which we are concerned are averages. But, to the degree that the measure of outcome used is a statistical average, the greater the number of random events that are applied directly to this outcome, the smaller the standard deviation of the result. The hypothetical universe of means we are speaking of gets relatively narrow as more choices are put in. Its standard deviation would vary inversely with the square root of the number of random choices made. Thus certain aspects of those games which repeatedly make a very large number of draws may be thought of as giving the same results as if average values had been used in the computations.

A complication involved in warlike simulations is that the entities one deals with (units or weapons) may be destroyed or eliminated. Thus one has a decreasing set of pieces to work with, and the relationship between the opposing pieces can well be changing as the work progresses. Thus, if a statistic is used like "the average number of Red entities destroyed per Blue counter-entity," such an average could be quite different at the beginning of a game than toward the end, when perhaps each side would at least need to do more hunting to find the enemy.

What should a game designer do then? Should he try to cut down the quantity of random numbers his game is to use, or try to increase them in the hope of reducing the deviation of possible results? No general rule seems evident at this time.

THE FUTURE OF PROCESSES OF DATA ANALYSIS*

John W. Tukey
Princeton University and Bell Telephone Laboratories

I am here to speak of the near-middle future of the processes of data analysis. This can only be done by indirection, since any processes that serve as examples must be drawn from the present of near future, but we can use such examples to illustrate what may be hoped to be broadly-applicable principles of continuing importance.

In general, the future of processes of data analysis is rosy, but it is not yet clear how fast the sun is rising. The modern computer has offered us many opportunities -- far more than we have seized -- and there have been many more opportunities for innovation that do not require a computer than we have seized. Looking at the last decade or two, it is clear that we have made much progress -- but we cannot be content with the rate at which we have gone. Will we do enough better in the future? Will we try to find approximate (or even crude) answers to more pressing problems, or exact answers to problems of limited (or non-existent) relevance? Who can say? (For a more historical and less specific discussion, see Tukey 1965.)

1. SOME PRINCIPLES. To point toward the near-middle future, we begin by stating a number of broad principles concerning the processes of data analysis (a phrase that ought to be construed as including the thoughts of the analyst of data as well as his manipulations) which we expect to retain their importance:

Two major aspects of such processes will continue their great importance:

(1) the essential erector-set character of data-analysis techniques, where any 2, 3 or 4 techniques are likely to be combined without warning,

(2) the steadily decreasing cost (and a so-far only slowly increasing ease) of computation, which is reflected in an ever-increasing

*Prepared in part in connection with research at Princeton University sponsored by the Army Research Office (Durham).

emphasis on computer usage and an ever-growing role of computer-unique contributions and processes.

Disproportionately rapid expansion will continue to repair past deficiencies in:

(3) graphicality and informality of processes of analysis,

(4) graphicality and incisiveness

(5) flexibility and fluidity

(6) empirical discovery of techniques

(7) focusing and parsimony.

In support of these improvements, our conceptual frameworks will give more and more attention to:

(8) doing the approximately right, rather than the exactly wrong (including dropping tight specifications as rapidly and generally as we may).

(9) using umbra-penumbra model pairs and other simultaneous (rather than alternative) model combinations.

(10) making the relation of estimator and target a two-way street.

And the day will yet dawn when:

(11) there will be one or more programming systems appropriate to data analysis.

As they stand, these principles are mainly unexplained words, requiring both examples and discussion to make then more understandable.

2. A GROUP OF EXAMPLES. Let us turn, then, to a group of examples, to instances of specific areas where progress is current. (I am sure my selection is biased, but this is only to be expected.) These

examples have not been selected to match the principles in a one-to-one way, instead each has been chosen to illustrate a few principles, with an attempt to illustrate each principle more than once. (Unfortunately, principle 11 cannot be adequately illustrated.)

Table 1 lists the examples and indicates their closest relations to specific principles.

I

THE EXAMPLES

3. NEWER APPROACHES TO TYPICAL VALUES. It has long been recognized that samples from distributions whose tails straggle more than those of a Gaussian were not well summarized by an equally weighted arithmetic mean. The procedures suggested by Jeffreys (1938, see also Newcomb 1886) for large samples have been occasionally implemented (e.g. Hulme and Symms 1939). More recently (i.e., Tukey 1960, Hodges and Lehmann 1962, Tukey and McLaughlin 1963) it has been recognized that what is needed is not so much a large-sample technique carefully bent to fit the particular distribution at hand, but rather techniques which provide relatively high efficiency over a wide range of distributions -- techniques that are (approximately) robustly efficient as well as being (approximately) robustly valid.

Much is being done in this area, and we shall soon have not only a body of asymptotic theory (Lehmann 1963a, b, c, 1964, Huber 1964, Buckel 1964) but an array of directly useful techniques (Hodges and Lehmann 1962, 1963, Høyland 1964, Dixon and Tukey 196?).

The problem of typical values in the plane, and in higher dimensions, is not so simple, since there is no obvious affine-invariant generalization of the notions of order statistics, which have played a central role in the one-dimensional case. Gentleman (196?) is tackling this problem from the point of view of minimizing p-th power deviations. Elashoff and Bickel (1964") are investigating Winsorizing and trimming. Soon we may expect working tools for this case, too.

Extensions to many other problems are obviously needed, and can be expected to occupy both asymptotic theorists and practical-technique designers over a considerable period.

TABLE 1

Title of example	Principles to which closely related										(Section in which discussed)
	1	2	3	4	5	6	7	8	9	10	
Newer approaches to typical values	x					x			x		(3)
New dissections of factorial tables	x						x	x			(4)
Spectrum-like techniques		x	x	x		x				x	(5)
Unrestrained monotone transformations	x	x		x	x						(6)
Ordered plotting			x	x							(7)
Hanging rootograms			x								(8)
Deomnibusing							x	x			(9)
The jackknife								x		x	(10)
$\widehat{\text{Var}}$ for weighted means	x								x		(11)

4. NEWER DISSECTIONS OF FACTORIAL TABLES. For a long time only two dissections of data arranged in a two- (or more-) way table were common in data analysis. Both of these were almost always left implicit rather than made explicit. I refer, of course, to the additive decomposition, whose two-way form is

$$y_{ij} = y_{..} + (y_{i.} - y_{..}) + (y_{ij} - y_{..}) + (y_{ij} - y_{i.} - y_{.j} + y_{..})$$

that underlies the analysis of variance for crossed and nested factors, and the multiplicative decomposition, whose two-way form is

$$n_{ij} = n_{++} \left(\frac{n_{i+}}{n_{++}}\right)\left(\frac{n_{+j}}{n_{++}}\right)\left(\frac{n_{ij}n_{++}}{n_{i+}n_{+j}}\right)$$

that underlies the chi-square test for independence in contingency tables.

The cases where the labels of the columns, or the labels of the rows, or both, are at least ordered (and perhaps even relevantly quantitative) are important and deserve much attention. They are not, however, part of the subject we wish to discuss here. Our immediate concern is with decompositions other than the usual ones which can be carried out on any two-way (or more-way) table.

Among the earliest of these was the separation of one degree of freedom for non-additivity (Tukey 1949) in which the "row" and "column" parts of the usual decomposition were used to identify and separate part of the "interaction" parts. Further discussion (Tukey 1955, Scheffé 1959, Elston 1961), some generalizations (Ward and Dick 1952) and various modifications of [illegible] Harter and L[illegible], Mandel 1959, Mandel 1961) followed.

The apparent needs of specific data analysis produced an extension along the lines of the "vacuum cleaner" (Tukey 1962) which does not function well in practice without the aid of some preliminary preparation (e.g. FUNOR-FUNOM, see Tukey 1962). This is only one of a branching family of alternatives that are still unexplored.

Some directions in which we ought to go are clear, but the details of tools and formulations are far from settled. We need to dissect a two-way table in more parts than the four indicated above. It will sometimes suffice to have:

(a1) An over-all contribution.
(a2) Column contributions.
(a3) Row contributions.
(a4) Unusual cell contributions.
(a5) Routine cell contributions.

As well as being important on their own, such dissections clearly have a close relation to the problems of Section 3.

Except for the smallest tables, it is likely to be necessary to go further, dissection row and column behavior into the unusual and the routine, just as for cell contributions. In either case, we will be prepared for both of these extremes:

(b) row and column effects clearly visible above a "noise" of routine cell contributions,

(c) a few cells deviating widely from all the others, which show no pattern of variation (including none by row or column).

We will be prepared for either extreme, since we shall be prepared for any mixture of these extremes.

We are here at a very early stage in the gaining of understanding. We have had some experience in the identification of unusualness, but we undoubtedly have much to learn. Once we are in reasonable shape for two-way tables, there are many ways to go.

5. SPECTRUM-LIKE TECHNIQUES. The application of Fourier methods to data gave rise to useful results in the simplest cases (e.g. Whittaker and Robinson 1924, Bartels 1940). The modern era in this area begins with the recognition (Bartlett 1950, Tukey 1950) that "white noise" is almost always a foolish null hypothesis, and that "white noise plus a few sharp lines" was an equally poor alternative hypothesis. Attention was first directed toward such questions as

consistent estimation of spectrum density (which the writer finds quite uninteresting, since he never saw even an approximately infinite amount of time series data) and variability (under Gaussian assumptions) of estimates of averaged spectrum densities. Later developments have emphasized the importance of keeping close touch with the average value of one's spectrum estimates and the advisability of introducing a variety of new techniques in order to approach the specific problems that are important in the specific application. (See Technometrics 1961 for a general introduction, including complex demodulation, see Akaike 1962 for misbehavior of the autocovariance function, see Akaike and Yamamouchi 1962 for practical problems in the use of cross-spectra, see Hasselman, Munk and MacDonald 1963 for the bispectrum, see Bogert, Healy and Tukey 1963 for the cepstrum, cross-cepstrum, pseudoautocovariance, and related concepts, etc., see MacDonald and Ward 1964 for interesting prediction-studying techniques.)

The two beliefs, both quite erroneous in the writer's view, that have contributed the most to delays and inadequancies in the use of spectrum analysis have been:

(a) A belief that, in using spectra, one ought to be concerned only with Gaussian situations.

(b) A belief that, in using spectra, one ought to be concerned only with stationary situations.

It is true that average value and spectrum only complete the specification of an ensemble of time series if we know more, say that the ensemble is Gaussian. This is, however, no more than the analog of the (equally correct) statement that average value and variance only complete the specification of a distribution when we know more, say that the distribution is Gaussian. (We do not, however, confine our use of variance to Gaussian distributions.)

In the absence of a suitably mathematical formulation and treatment of spectra for nonstationary ensembles, there has been an unfortunate tendency for some workers to feel that spectrum techniques should only be applied to situations of apparent stationarity. In practice, this can be quite foolish, as Munk and Snodgrass's discovery (1957) through their nonstationarity, of the weak long-period ocean waves arriving on our

Pacific Coast from the Indian Ocean and beyond, illustrates. In theory, it is at best dubious, since if our universe should repeat itself every 10^{11} to 10^{12} years, the whole universe (with all its time series) may perhaps be thought of as stationary -- and who can deny such a possibility?

It may prove fortunate that a mathematical formulation of the non-stationary case is now at hand (Priestley 1965) which tells us to do for slowly changing spectra just what we have done for plausibly constant spectra.

In addition to the new types of quantities being introduced and used, we are in the middle of a change in the actual computing techniques used to process the data. Where once some subsequence of:

(c1) taper
(c2) prewhiten
(c3) form mean lagged products
(c4) apply lag window
(c5) Fourier transform
(c6) hann or hamm, etc.

was relatively standard, alternative approaches, involving more computations linear in the observations before the formation of squares of products involving the data, are in use or contemplation. Techniques using complex demodulation appear to involve very real advantages, and are already in routine use (M. D. Godfrey 1964*). Now that complete Fourier transformation for N observations requires only a few times $N \cdot \log_2 N$ multiplications rather than N^2 (Cooley and Tukey 1965), we may well see computational techniques develop which start by complete Fourier transformation of the entire data. (The spectrum-analytic character of these techniques will be revealed by what happens next to the Fourier coefficients and how the ultimate quantities are interpreted.)

In economics, spectrum analysis is currently being applied to the problem of seasonal adjustment, and as a consequence economists are again thinking about the difficult question of what seasonal adjustment is really supposed to do.

So far as one can now see, spectrum-like analysis is going to continue to ramify and develop at a substantial rate.

6. UNRESTRAINED MONOTONE TRANSFORMATION. The fight between those who feared the loss of knowledge that comes from analyzing unwisely expressed data and those who feared serious biasing of levels of significance and confidence would come from expressing the data in the way in which it seemed to like to be expressed is an old one, but one that has never reached the front pages. Partly this has been because changes in modes of expression have seemed unimportant. Partly, I fear, it has been because those who realized that, in practice, 100% and 200% improvements in efficiency come more frequently from such changes than from almost anything else the analyst of data can do once the data is taken, have not advertised this fact sufficiently.

Those who have sought better modes of expression have traditionally chosen some simple family of transformations, often $z = (y+c)^p$, and have tried to choose the few parametric constants wisely in each particular instance. (For a clear exposition of a highly developed form of this approach, see Box and Cox 1964.) As the techniques have become more explicit, the hope of their wider application has increased steadily.

All this continues to be important, but the pressure of a real need for better multidimensional scaling has brought about a computer-aided revolution. The work of Shepard (1962, 1963) and Kruskal (1964a, b) has shown how much can often be gained by letting the computer choose whatever monotone transformation of the original value will lead to the simplest analysis. The impact upon multidimensional scaling and factor analysis is already substantial. Kruskal's reanalyses (196?) of Box and Cox examples show that even a 3x3x3 experiment may be big enough for such an analysis to be fruitful. We can hope for similar progress in many other areas (although semi-classical results on "Maximalkorrelation" show that we cannot do it everywhere).

7. ORDERED PLOTS. The classical example of plotting observed values rearranged in increasing order is the use of "probability paper" to show the apparent Gaussianity, or absence thereof, of a sample of observations. This example is classical, but it is still surprising how many statisticians have had little contact with the technique.

The arrival of the half normal plot (Daniel 1959) introduced a major change into the analysis of unreplicated and fractionated 2^p experiments. The idea that a set of contrasts could be used to show forth the unusual size of its largest values, if any of their sizes were truly unusual, is not a difficult one. It is perhaps surprising that it took so long to appear.

Later, the more general technique of "gamma plotting", in which two parameters require estimation rather than one, was developed and applied in a variety of directions (Wilk and Gnanadesikan 1961, 1964a, Wilk, Gnanadesikan and Huyett 1962).

Today, the problem of adapting these techniques to the general analysis-of-variance situation, where different mean-squares have different numbers of degrees of freedom is being actively attacked with interesting results (Wilk and Gnanadesikan 1964b, Wilk, Gnanadesikan and Lauh 1964).

As a consequence of this, the writer is convinced that we shall see a partial return of the pendulum, which has now swung from analyses guided only by the natural order of lines (and the relations between average mean squares) of the analysis of variance to analyses guided only by the relative sizes of the mean squares. I, for one, believe that we ought to expect attention to both considerations in well thought-through analyses, though in ratios differing widely from instance to instance. (Given a complete 2^{12}, for instance, whose 4095 contrasts behave exactly like a Gaussian sample, I would regard the fact that the 12 largest contrasts were the 12 main effects as nonaccidential and highly significant.)

Here, too, we can, I believe, lift the curtain of the future a little When I try, I see signs of plots of gaps (= spacings) among the ordered observations appearing alongside -- and even in partial replacement of-- the more classical plots of the raw ordered observations. Time will tell.

8. HANGING ROOTOGRAMS. This example is included to show that even among the simplest of graphical techniques there can be new and useful techniques.

The histogram, with its columns of area proportional to number, like the bar graph, is one of the most classical of statistical graphs. Its combination with a fitted bell-shaped curve has been common since the days when the Gaussian curve entered statistics. Yet as a graphical technique it really performs quite poorly. Who is there among us who can look at a histogram-fitted Gaussian combination and tell, reliably, whether the fit is excellent, neutral, or poor? Who can tell when the fit is poor, of what the poorness consists? Yet these are just the sort of questions that a good graphical technique should answer at least approximately.

How can we do better? If we have observed n_i cases in the ith class, we know that the variance of n_i is reasonably proportional to its average values (at least so long as n_i is not a large fraction of the total number of cases, n_+).

If we are to do a reasonable job of assessing fit, we deserve to have roughly constant variance. We can do this by replacing n_i by $\sqrt{n_i}$, as we are well aware in other contexts. We can do the same here, at least for the case of classes of equal width. We have only to take the square root (of the height) of the fitted curve at the same time that we take the square roots of the counts.

Because of the simple identity:

$$\sqrt{\text{one Gaussian density}} = (\text{constant})\,(\text{another Gaussian density})$$

the picture will look much the same -- in the large -- a family of rectangles compared with a Gaussian curve, but now variability is nearly constant (at the price of giving up the principle of "equal area for equal count" which has real uses in other directions but few if any in connection with goodness of fit).

But we are still comparing the ends of a row of rectangles with a curve, something the human-eye-and-brain combination is less than perfect at. How do we improve matters here? We have only to say, carefully and precisely, what we have always done, in order to learn what we might better do. Classically, we have taken a stack of rectangles, fixed one end of each on a horizontal line and compared the other ends with a curve. It is not a great step to say: "Let us take our stack of rectangles, fix one end of each on a curve and compare the other ends with the straight line. Why did we not do it long ago?"

While we are about it, we might as well turn the picture over, letting the curve hang down, supporting the rectangles. This third change completes our path to the "suspended rootogram" in which the eye can do so much more for us. (Some viewers prefer to stop at the "hanging rootogram" stage.) Figures 1, 2, 3, 4 [at the end of this article] show successive stages in the progress from conventional histogram to hanging rootogram.

Theré are other simple things to do in the graphical area, as we shall learn as we take care to realize that graphs can and should, among other things, be used for diagnosis as well as naive exhibition.

9. DEOMNIBUSING. The first step in data analysis is often an omnibus step. We dare not expect otherwise, but we equally dare not forget that this step, and that step, and other step, are all omnibus steps and that we owe the users of such techniques a deep and important obligation to develop ways, often varied and competitive, of replacing omnibus procedures by ones that are more sharply focused.

The replacement of group comparisons by multiple comparisons has been one of the outstanding phenomena of the last decade and a half. It has raised many deep issues on which we are far from being completely agreed -- whose discussion would take more space than we can here provide. So we note here only that a full account of the short-cut methods using ranges both in numerator and denominators is at last appearing (Kurtz, Link, Tukey and Wallace 1965, 196?).

We note also that progress has also been made on the deomnibusing of contingency table chi-square.

The detection of differences in the effects of ordered treatments -- under circumstances where the effects, if any, may be expected to be directly -- or antithetically -- ordered has at last engaged the attention of technique manufacturers. Two competing approaches exist, about which all protagonists will agree that either one is to be preferred to the unwise use of a flabby group comparison. One procedure is developed in a framework of successive testing (Bartholomew 1959, 1961a, b); the other in a framework of single contrasts of maximizing the least sensitivity (Abelson and Tukey 1959, 1963). (The writer notes a continuing preference for the latter, based on what he regards as good reason. Again space bars further discussion.)

Still more recently, there is progress in the deomnibusing of "goodness of fit" tests, which have always had so omnibus a character. For small samples, or compulsory heavy grouping, we need not merely sum the squares of standardized (or, better, Studentized) deviations to find a chi square. As has long been known (e.g. Cochran 1954) we can introduce any convenient set of orthogonal comparisons, and evaluate the results as separately or jointly as we wish. In doing this, it should be our hope to concentrate the effects of fitting the curve to which the data is compared as thoroughly -- and into as few comparisons -- as reasonably may be.

In larger samples, particularly in the absence of grouping, one can go a long way toward the separation of "badness of fit" into three parts:

(a1) underestimated badness of fit, where the almost inevitable fitting of parameters has concealed any true badness,

(a2) systematic badness of fit, where the deviations are both interpretable and indicative of inadequacy of shape of model,

(a3) irregular badness of fit, often an indication only of inadequacy of simple random sampling -- no evidence of inadequacy of distributional shape.

Once this is accomplished, the introduction of the ideas underlying ordered plotting allows us to break new ground, to -- reasonably and sensibly -- inquire as to goodness of fit for many kinds of nonrandom samples without preassumption of what kind of nonrandomness is involved. Early trials of such techniques have had quite illuminating results (Quandt 1964, 196?).

These are only the beginning. Deomnibusing of all our usual omnibus procedures will do much to occupy both technique-manufacturers and philosophy understanders in the years just ahead.

10. THE JACKKNIFE. The "jackknife" procedure allows almost any of us to set approximate confidence limits on almost any results calculated from data which go a reasonable way toward revealing the variation whose likely effects are to be spanned by the confidence interval.

In its simplest form, the jackknife procedure assumes that

(a1) we have data, and a fixed procedure for extracting an interesting number (or numbers) from the data,

(a2) this procedure can be applied to varying amounts of data,

(a3) the data can be divided into r "pieces" of roughly equal "size",

(a4) this can be done in such a way as to make the differences from piece to piece "adequately reflect" the sorts of variation whose effects are to be spanned by the confidence interval,

(a5) the prototype case of "adequately reflect" is the sampling of r "pieces" from a very large collection of pieces, whose combined processing would, by definition give the right answer,

(a6) the results of the processing are <u>not</u> narrowly estimated, in the sense that no one piece has (and no very few pieces have) a dominating effect upon the result.

Given all this, to some reasonable approximation and according to some reasonable belief (which is all that one can ever truly demand) the analyst treats his data as follows:

(b1) Let y_{all} be the result of processing all the pieces of data together.

(b2) Let $y_{(i)}$, read "y-not-i", be the result of processing all but the ith piece of data (hence processing r-1 out of r of the pieces together).

(b3) Let y_{*i}, read "y-pseudo-i", be given by

$$y_{*i} = ry_{all} - (r-1)y_{(i)}$$

(b4) Let $y_* \pm t_{r-1} \cdot s_*$ be the mean of the y_{*i} and the confidence interval generated by a naive application of Student's to the y_{*i} (as if they were a sample).

The procedure is simple, the approximation is usually satisfactory, and the technique is applicable in very diverse and complex circumstances.

Happily this technique has begun to receive attention from some of those fitted to pinpoint some of its weaknesses and difficulties. In particular, there has been inquiry into the asymptotic behavior of the technique, especially where condition (a6) fails (Miller 1964). It is to be hoped that there will be more such studies -- and that their results will be correctly evaluated from the point of view of practice.

The cases where (a6) is most likely to fail are those in which a single order statistic, a median, a maximum, a minimum, or a few order statistics play an unusually important part. In some of these, particularly where medians and other inner order statistics are concerned, we have other means of assessing the stability of our answers that are adequately robust. In these cases we should clearly use these alternate procedures.

In others, often those involving maxima, minima, and ranges, it is clear that a properly assessed uncertainty for the quantity of interest will inevitably depend on such matters as the actual shape of the underlying distribution or distributions. Here robustness is impossible, and so is certainty of validity for any confidence procedure. It will often be true that the best that we can do is to use the jackknife in such situations, even though we know it may be fallible. It is usually better to have some idea of the uncertainty of our values rather than none. (No confidence interval will ever be computed from data in such a way as to include all possible sources of variation, since no body of data allows all possible sources to reveal themselves. A little more inadequacy will not be fatal.)

In cases where (a6) is not in question, the situation is rather similar. If there is available a robust special confidence procedure clearly applicable to the case at hand, by all means use it. Otherwise use the jackknife.

11. ESTIMATED VARIANCES FOR WEIGHTED MEANS.

(a) Given n uncorrelated observations y_i with the same average value and fixed finite variances

for which also

(b) the variances σ_i^2 of y_i are all equal, say to σ^2, it is well known that

$$s^2 = \frac{1}{n-1} \Sigma (y_i - \bar{y})^2$$

where

$$\bar{y} = \frac{1}{n} \Sigma y_i$$

is an unbiased estimate of σ^2 and that

$$s_1^2 = \frac{1}{n} s^2$$

is an unbiased estimate of the variance σ^2/n of $\bar{y}$.

It is known, but not to enough people, or clearly enough, that (b) has no part in the relation

$$(*) \qquad \text{ave } s_1^2 = \text{var } \bar{y}$$

which is a consequence of (a) alone.

We have here a simple example of an umbra-penumbra situation in which two models, one encompassing the other, are wisely considered simultaneously. The penumbra or outer model, here defined by (a) above, suffices for the validity of s_1^2 as an estimated variance $\bar{y}$ (in the sense that (*) then holds). The umbra or inner model, here defined by (a) and (b) together, ensures the optimality of s_1^2 as the unique quadratic function of the y_i that (i) satisfies (*) and (ii) minimises its own variance among the quadratics that do this.

The pattern here: "validity in the outer model, optimality in the inner" is but one of many possible patterns for simultaneous model pairs. It is however, one of the most important ones, one that needs much more explanation.

Suppose, for instance, that our concern is not with $\bar{y}$ but with

$$y_C = \frac{\Sigma c_i y_i}{\Sigma c_i}$$

where the $c_i \geq 0$ are fixed. Can we use the values of the c_i to determine a quadratic function $Q(y_1, y_2, \ldots, y_n)$ so that

(**) $$\text{ave } Q = \text{var } y_C$$

provided only that (a) holds and the c_i are as assumed? Certainly we can do this. We can, indeed, press right on, and find a Q which (i) satisfies (**) under (a) and (ii) minimizes its own variance under (a) and (b) combined. Nay more, we may replace (b) by

(c) the variances of the y_j are in known ratio, in that

$$\text{var } y_i = d_i \sigma^2$$

where the d_i are fixed and known.

For each choice of $\{d_i\}$, there will be a Q satisfying (**) under (a) and minimizing var Q under (a) and (c). This Q will, in fact, be different for different choices of $\{d_i\}$.

We could write down, in closed form, expressions for these Q's, but their detailed form is of far less concern to us than the facts that

(d1) we can have any of many umbras with a single penumbra

(d2) which umbra we choose can, sometimes, turn efficiencies topsy-turvy without affecting validity

(d3) the equally weighted mean seems to have no unusual roles; it appears to be just another weighted mean, the one, perhaps, for which certain formulas look simplest.

We need, and are inevitably going to get for ourselves, a very much wider collection of instances where the consequences of umbra-penumbra model pairs have been worked out, much to our illumination and advantage.

II

THE RELATION OF EXAMPLES TO PRINCIPLES

12. HOW THE EXAMPLES ILLUMINATE THE TWO MAJOR ASPECTS. The first (erector-set) principle is, according to Table 1, illustrated by:

(a1) newer approaches to typical values: where Winsorizing is combined with Student's t; where techniques developed for single samples are expected to be used directly or indirectly in simple and multiple regression and in all sorts of analyses of variance involving replication within cells,

(a2) new dissections of factorial tables: where we try to use both factorial and idosyncratic dissections at the same time; where we expect to build each new kind of dissection into more and more complex patterns,

(a3) unrestrained monotone transformation: which is rapidly propagating itself in cooperative combination with a wide variety of other techniques,

(a4) internally estimated variances for weighted means: where we learn how to do, knowingly, for weighted means what we have so long done, often unknowingly, for equally weighted means.

While not one of these is as striking as Cuthbert Daniel's unpublished injections of 2^{j-m} fractional factorial analysis into the calculations of multiple regression, or as striking as the technique-combinations that are, in practice, appropriate to a variety of complex bodies of data, they do offer solid illustrations.

The second principle (computation cheaper, more used, and more vital) is certainly well-exemplified above. Consider:

(b1) spectrum techniques: where hand-calculator work would be worth while for some of the most crucial instances, but where the cost of hand computation, if it had to be paid, would keep us from many of the useful and illuminating studies we actually make.

(b2) unrestrained monotone transformation: where iterative computer algorithms lead us easily to ends not at all reasonably accessible by hand computation.

(b3) new dissections of factorial tables: where, when the still more effective techniques arrive, their feasibility will depend greatly, perhaps absolutely, upon the availability of computers.

(b4) ordered plotting: where much of the real push forward seems to be associated with the use of modern computer both to calculate and to make such plots.

(b5) deomnibusing: where, while the deomnibusing of goodness of fit is, in many instances, feasible without a modern computer, it is the availability of computer procedures that will make such techniques popular.

(b6) the jackknife: where one of the great advantages is that a well-written computer problem to do y_{all} can also be used to do y_{*j}, so that the cost of jackknifing is only a little more running time, but no extensive effort in programming or debugging.

While these are matters of technique manufacture rather than technique use, many of the new approaches to typical values are only accessible because of the modern computer (as when Monte Carlo techniques are required to find critical values, even when the underlying distribution is Gaussian).

There can be little doubt of the importance of the second principle.

13. HOW THE EXAMPLES ILLUMINATE THE FIVE AREAS OF RAPID EXPANSION. The first area where rapid expansion is trying to repair past deficiencies, principle 3, involves graphicality and informality, where the graph is used as an effective, but very informal, way of connecting the data to the human judgments that are going to be made about it, that constitute the reasons for its analysis. Here, Table 1 points out:

(a1) spectrum-like analysis: where a far larger share of judgments than many might suppose are in fact based on graphical presentations, informally examined.

(a2) ordered plotting: where a formal significance testing procedure is largely replaced by an informal judgment made by those who look at the plot.

(a3) hanging rootograms: where we have striven to learn graphically and far less formally by seeking an approach to goodness of fit where a moderately wise man's eye will tell most of the story.

The second area, principle 4, involves graphicality and incisiveness and is quite distinct from the first, although the areas share graphicality and appear together in many techniques. At issue here is that grand property of many graphs: revelation of the unexpected through the simultaneous revealability of many possible deviations from neutrality. Table 1 directs our attention to:

(b1) spectrum-like techniques: where little peaks have often revealed new phenomena, as in Munk and Snodgrass 1957, or as in the detection and evaluation of the natural modes of vibration of the earth;

(b2) unrestrained monotone transformations: where the graph of the final monotone transformation is often quite revealing: where the structures resulting from multidimensional scaling often show unexpected properties.

(b3) ordered plotting: where we have learned that half-normal plots expose many kinds of interesting behavior other than the stray large values to detect which the plot was invented (Daniel 1959).

The third area, principle 5, involves flexibility and fluidity and deserves discussion in two rather separate pieces. Flexibility here refers to the existence of a wider variety of frameworks for analysis and inference, thus offering, on the average, a better match to the needs of the problem. Fluidity refers to the ability of single analytical procedures to respond in a very wide variety of ways to the apparent character of individual bodies of data. Clearly a continuous graduation from flexibility to fluidity is possible -- indeed many stages along this gradation will be realized -- but for all that it still will pay us to try to make the distinction between choices made in advance by the analyst and choices made en passant by the data.

We have not done much to illustrate flexibility directly in our examples, yet most of them illustrate it indirectly, thus:

(c1) Our newer approaches to typical values are not yet focused into one form -- as they might some day be, where they completely fluidized. (As I trust they never will be, since I expect that the analyst will always be very often able to add information about the underlying distribution over and above that contained in a single small sample.) Shall we use trimmed means, Winsorized means, or Hodges-Lehmann median differences? If we trim or Winsorize, how far? We have not yet provided the user with the information most helpful in choosing answers to these questions, but we have begun to provide him the flexible kit of tools from which to choose.

(c2) There will be more than one choice among new dissections of factorial tables.

(c3) One can rightfully say that the modern phase in spectrum-like analysis comes from expanding our kit of tools beyond the serial correlation function and the periodogram (neither of which was really helpful).

(c4) The jackknife is a great aid to flexibility; in most situations it removes that grim complaint "But if we do that how can we compare the result with chance fluctuations?" and allows much freer choice of technique.

(c5) The ability to estimate variability for all weighted means with the same robustness as for the equally weighted case is a similar contributor to flexibility.

For the moment, the unrestrained monotone transformation is the outstanding example of complete fluidity. We face a challenge in finding others.

The fourth area of growth, principle (6), involves the empirical discovery of techniques, as opposed to their theoretico-mathematical discovery. Here again we deal with a matter of amount rather than kind. It seems likely that no technique was developed solely empirically, without any "theoretical" insight at all, though many have been developed without any trace of a mathematical mode. At the other extreme, techniques based on rigid mathematical models, clearly-specified criteria, and vigorous optimization only gain credibility from some empirical support, whether of their hypotheses or of their functioning in practice.

Accordingly, our examples will tend to be ones that show a greater empirical content than most, ones whose developments are separated in amount, rather than kind, from those of most techniques. Table 1 cites:

(d1) newer approaches to typical values: where "trimming" and "Winsorization" came into being at least as much because of how they worked in practice as for any insight or theoretical argument: where the matching of denominators to numerators has come about by empirical comparisions based on tables of order statistic moments; where the critical values have often to be determined by Monte Carlo.

(d2) spectrum-like techniques: where one source of modern lag-windows was Hamming's observation that the points of an estimated spectrum for a single particular set of data would be improved by hanning; where the pseudoautocorrelation was suggested by a diffuse analogy with the cepstrum, and only the fact that it seemed to work made it plausible.

Though not quite a technique of data analysis, the near constancy of standardized 5% distances for Pearson curves (Pearson and Tukey 196?) is based upon Charles P. Winsor's wholly empirical discovery of the near constancy of the standardized 5% distance for chi-square.

The fifth growth area, principle 7, is one of focusing and parsimony. Some books on probability and statistics reveal that every sample (or other grouping of observations) is unusual in some way. (If only by how closely it matches a copy of itself.) It is rare, however, that the discussion carries on to the logical conclusions: First, that it is important to be restrictive in the kinds of unusualness to which one pays attention. Second, that one escapes this difficulty when one can focus all one's attention upon a single numerical aspect, or on a very few numerical aspects. Third, that once a fair number of such aspects are involved one is in a situation very like the unrestricted case and that just how one divides his attention is of great importance. One can be wisely parsimonious with one of one's most valuable possessions, by focusing one's attention where this is most likely to be profitable.

Table 1 directs our attention to:

(e1) new dissections of factorial tables: where, instead of merely giving a single number to an inchoate mass of "interaction", we are striving to attend to very particular aspects, such as the single very unusual cell or indications that some other mode of expression will lead to a better approximation to additivity.

(e2) deomnibusing: in each of whose specific instances we are trying to improve our focusing, to learn about something identifiable and thereby to increase both the value of our knowledge and the chance of gaining it.

14. HOW THE EXAMPLES ILLUMINATE THE AREAS OF INCREASED ATTENTION. The first principle of increased attention, principle 8, calls for greater attention to being approximately right rather than exactly wrong. The hardest part of this, at least for the mathematician, is to admit that one is proceeding approximately -- even though it is hard to see how one can ever do better in the real world.

Table 1 directs our attention to:

(a1) new dissections of factorial tables: where we are seeking to ask the questions of greatest importance to us, even though their asking tends to destroy the neat, nice, manageable, null hypothesis which was the formal foundation for the classical asking of less useful questions; where our conclusion levels are going to become approximate; where there will be, for a period of years at least, no formal criterion to insulate us from the very real difficulties of picking a good technique.

(a2) deomnibusing: where we are again very willing to be approximate in the answering of more meaningful questions.

(a3) the jackknife: where by admitting that an approximate conclusion procedure can serve us, we have brought a very much wider range of techniques into the fold for which confidence, and significance, statements are at hand for use when appropriate.

The second principle of increased attention, principle 9, calls for more use of model-pairs and other "it might be A and it might be B and we must think about both together" approaches. The use of pairs of models as alternatives, as in the Neyman-Pearson account of hypothesis testing, is classical. (Pearson (1939) points out how much Student had to do with the recognition of its importance.) It is remarkable, by contrast, how little attention has been paid to pairs of models simultaneously considered. Perhaps this is because, in many instances, the use of simultaneous model pairs inevitably attracts attention to the deficiencies of a technique. In an optimality-validity umbra-penumbra situation, for example, emphasizing the validity of the technique in the penumbra cannot help reminding us that it is not optimum throughout.

Table 1 draws our attention to:

(b1) newer approaches to typical values, where Gaussian and crudely Gaussian underlying distributions provide umbra and penumbra that are used in varied ways: relatively good efficiency for the Gaussian and validity for all symmetric distributions; critical values set for the Gaussian (and approximately valid elsewhere) and moderately high efficiency (except for unseizable opportunities) anywhere near the Gaussian: etc.

(b2) internally estimated variance for weighted means: where the whole discussion is on an umbra-penumbra basis.

The third principle of increased attention, principle 10, calls for making the relation of estimator and estimand a two-way street. (See Tukey 1962, p. 10 and references cited there.) The mathematician wants the problem to come before the solution. But a good solution can often be recognized as such before we have identified one or more of the problems it solves. And a good solution may be good because it solves a problem other than the one as whose solution it is customarily derived.

Table 1 directs our attention to:

(c1) spectrum-like techniques: where much has been gained by asking what spectrum estimates actually do estimate, rather than by asking for asymptotic results which demand unreal amounts of data, (c2) the jackknife: where the estimator is defined by a process, selected by what wisdom the analyst possesses, and the estimand follows after it, like the tail of a kite.

In each of these three areas of increased attention, if one goes through the uncited examples carefully, one will find each principle recurring again and again, though usually less explicitly. If one looks at the three areas in the right way, they seem to blur and move together into one.

If we look at all the principles, the same blurring appears, though not as obviously. There is a sense in which all these principles are "sisters under the skin".

III

THE CONCLUSION

15. SUMMARY. If we ask of the near future of processes of data analysis, one can predict three essentials:

(d1) greater realism,

(d2) greater effectiveness,

(d3) greater use of computers.

REFERENCES

(The final digit "?" indicates an unpublished paper, while an * identifies a personal communication.)

R. P. Abelson and J. W. Tukey 1959. Efficient conversion of nonmetric information into metric information. Proc. Social Statist. Section (Amer. Statist. Assoc.) 1959, 226-230.

R. P. Abelson and J. W. Tukey 1963. Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. Annals Math. Statist. 34, 1347-1369.

H. Akaike 1962. Undamped oscillation of the sample autocovariance function and the effect of prewhitening operation. Annals Inst. Statist. Math. (Tokyo) 13, 127-143.

H. Akaike and Y. Yamamouchi 1962. On the statistical estimation of frequency response function. Annals Inst. Statist. Math. 14, 23-56.

J. Bartels 1940 (2nd ed. 1951). Periodicities and harmonic analysis in geophysics. Chapter 16 (pp. 545-605, opening vol. 2) of Geomagnetism (by S. Chapman and J. Bartels) Oxford University Press.

D. J. Bartholomew 1959. A test of homogeneity for ordered alternatives I, II. Biometrika 46, 36-48 and 328-335.

D. J. Bartholomew 1961a. A test of homogeneity of means under restricted assumptions. J. Roy. Statist. Soc. Ser. B, 23 239-271 (discussion 271-281).

D. J. Bartholomew 1961b. Ordered tests in the analysis of variance. Biometrika 48, 325-330.

M. S. Bartlett 1950. Periodogram analysis and continuous spectra. Biometrika 37, 1-16.

P. J. Bickel 1964. On some robust estimates of location. (Abstract) Ann. Math. Statist. 35, 1403.

B. P. Bogert, M. J. R. Healy and J. W. Tukey 1963. The frequency analysis of time series for echoes: cepstrum, pseudoautocovariance, cross-cepstrum and saphe cracking. Chapter 15 (pp. 209-243) of (Proc. of the Symp. on) Time Series Analysis (ed. M. Rosenblatt). New York, John Wiley and Sons.

G. E. P. Box and D. R. Cox 1964. An analysis of transformations. J. Roy. Statist. Soc. Series B 26, 211-243 (discussion 244-252).

W. G. Cochran 1954. Some methods for strengthening the common χ^2 tests. Biometrics 10, 417-451.

J. W. Cooley and J. W. Tukey 1965. An algorithm for the machine calculation of complex fourier series. Mathematics of Computation 19 (in press).

C. Daniel 1959. Use of half-normal plots in interpreting factorial two level experiments. Technometrics 1, 311-341.

W. J. Dixon and J. W. Tukey 196? Approximate behavior of the distribution of Winsorized t (Trimming/Winsorization 2). To be submitted for publication.

R. M. Elashoff and P. J. Bickel 1964*. Personal communication.

R. C. Elston 1961. On additivity in the analysis of variance. Biometrics 17, 209-219. (Also abstract 688 at page 166.)

W. M. Gentleman 196? Robuts affine location estimation by minimizing pth power deviations. Ph. D. thesis (in progress) Princeton University.

M. D. Godfrey 1964*. Personal communication about programs in use at the Princeton University Computing Center.

H. L. Harter and M. D. Lum 1958. A note on Tukey's one degree of freedom for non-additivity. (Abstract 474). Biometrics 14, 136-137.

K. Hasselman, W. Munk and G. MacDonald 1963. Bispectra of ocean waves. Chapter 8 (pp. 125-139) of (Proc. of the Symp. on) Time Series Analysis (ed. M. Rosenblatt). New York, John Wiley and Sons.

J. L. Hodges, Jr. and E. L. Lehmann 1962. Rank methods for combination of independent experiments in analysis of variance. Annals Math. Statist. 33, 482-497.

J. L. Hodges and E. L. Lehmann 1963. Estimates of location based on rank tests. Annals Math. Statist. 34, 598-611.

A. Høyland 1964. Evaluation of Hodges-Lehmann estimates (Abstract). Annals Math. Statist. 35, 938.

P. J. Huber 1964. Robust estimation of a location parameter. Annals Math. Statist. 35, 73-101.

H. R. Hulme and L. S. T. Symms 1939. The law of error and the combination of observations. Roy. Astron. Soc. Monthly Notices 99, 642-649.

H. Jeffreys 1938. The law of error and the combination of observations. Phil. Trans. Roy. Soc. Lond. A. 237, 231-271.

J. B. Kruskal 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29, 1-27.

J. B. Kruskal 1964b. Nonmetric multidimensional scaling: a numerical method. Psychometrika 29, 115-129.

J. B. Kruskal 196? Analysis of factorial experiments by estimating monotone transformation of the data. To appear in J. Roy. Statist. Soc. Ser. B.

T. E. Kurtz, R. F. Link, J. W. Tukey and D. L. Wallace 1965. Short-cut multiple comparisons for balanced single and double classifications, Part 1, Results. Technometrics 7 (in press).

T. E. Kurtz, R. F. Link, J. W. Tukey and D. L. Wallace 196? Short-cut multiple comparisons for balanced single and double classifications, Part 2, Derivations and approximations. Submitted to Biometrika.

E. L. Lehmann 1963a. Robust estimation in analysis of variance. Annals Math. Statist. 34, 957-966.

E. L. Lehmann 1963b. Asymptotically nonparametric inference: an alternative approach to linear models. Annals Math. Statist. 34, 1494-1506.

E. L. Lehmann 1963c. Nonparametric confidence intervals for a shift parameter. Annals Math. Statist. 34, 1507-1512.

E. L. Lehmann 1964. Asymptotically nonparametric inference in some linear models with one observation per cell. Annals Math. Statist. 35, 726-734.

N. J. MacDonald and F. Ward 1964. The prediction of geomagnetic disturbance indices. 1. The elimination of internally predictable variations. J. Geophys. Res. 68, 3351-3373.

J. Mandel 1959. The analysis of latin squares with a certain type of row-column interaction. Technometrics 1, 379-387.

J. Mandel 1961. Non-additivity in two-way analysis of variance. J. Amer. Statist. Assoc. 56, 878-888.

R. G. Miller, Jr. 1964. A trustworthy jackknife. Annals Math. Statist. 35, 1594-1605.

W. H. Munk and F. E. Snodgrass 1957. Measurements of southern swell at Guadalupe Island. Deep Sea Research 4, 272-286.

S. Newcomb 1886. A generalized theory of the combination of observations so as to obtain the best result. Amer. J. Math. 8, 343-366.

E. S. Pearson 1939. "Student" as statistician. Biometrik 30, 210-250 (especially pp. 242-243).

E. S. Pearson and J. W. Tukey 196? Approximate means and standard deviations based on distances between percentage points of frequency curves. (In preparation.)

M. B. Priestley 1965. Evolutionary spectra and nonstationary processes. J. Roy. Statist. Soc. Series B 27 (to appear).

R. E. Quandt 1964. Old and new methods of estimation and the Pareto distribution. Econometric Research Program Research Memorandum No. 10. Princeton University (submitted to the J. Amer. Stat. Assoc.).

R. E. Quandt 196? Statistical discrimination among alternative hypotheses and some economic regularities. J. Regional Science (in press).

H. Scheffé 1959. The Analysis of Variance (477 pp.) New York, John Wiley (especially pp. 129-134).

R. N. Shepard 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika 27, 125-140. II. Psychometrika 27, 219-246.

R. N. Shepard 1963. Analysis of proximities as a technique for the study of information processing in man. Human Factors, 5, 33-48.

Technometrics 1961. (Papers on spectrum analysis by G. M. Jenkins (133-166, 229-232), E. Parzen (167-190, 232-234), J. W. Tukey (191-219), T. H. Wonnacott (235-243) and N. R. Goodman, S. Katz, B. H. Kramer, M. T. Kuo (245-268)) Technometrics 3, 133-268.

J. W. Tukey 1949. One degree of freedom for non-additivity. Biometrics 5, 232-242.

J. W. Tukey 1950. The sampling theory of power spectrum estimates. (pp. 47-67) in Symposium on Applications of Autocorrelation Analysis to Physical Problems (NAVEXOS-P-735) Office of Naval Research.

J. W. Tukey 1955. Answer to query 113. Biometrics 11, 111-113.

J. W. Tukey 1960. A survey of sampling from contaminated distributions. Paper 39 (pp. 448-485) in Contributions to Probability and Statistics (edited by I. Olkin et al) Stanford University Press.

J. W. Tukey 1962. The future of data analysis. Annals Math. Statist. 33, 1-67.

J. W. Tukey 1965. The technical tools of statistics. The American Statistician (to appear, probably in April).

J. W. Tukey and D. H. McLaughlin 1963. Less vulnerable confidence and significance procedures for location based on a single sample (Trimming/Winsorization 3). Sankhya 25, 331-352.

G. C. Ward and J. D. Dick 1952. Non-additivity in randomized block designs and balanced incomplete block designs. New Zealand J. Science and Techn. B. 33, 430-435.

E. T. Whittaker and G. Robinson 1924 (4th, 1946). The Calculus of Observations (397 pp.) London, Blackie and Son.

M. B. Wilk and R. Gnanadesikan 1961. Graphical analysis of multiple response experimental data using ordered distances. Proc. Nat. Acad. Sci. USA 47, 1209-1212.

M. B. Wilk, R. Gnanadesikan and M. J. Huyett 1962. Probability plots for the gamma distribution. Technometrics 4, 1-20.

M. B. Wilk and R. Gnanadesikan 1964a. Graphical methods for internal comparisons in multiresponse experiments. Annals Math. Statist. 35, 613-631.

M. B. Wilk and R. Gnanadesikan 1964b. A probability plotting procedure for internal comparisons in a general analysis of variance. Unpublished invited talk presented at Royal Statist. Soc. meetings, Cardiff, Wales, in September 1964.

M. B. Wilk and R. Gnanadesikan and E. Lauh 1964. Scale parameter estimation from the order statistics of unequal gamma components. Unpublished memo., Contributed paper presented at IMS meetings at Amherst, Mass., in August 1964.

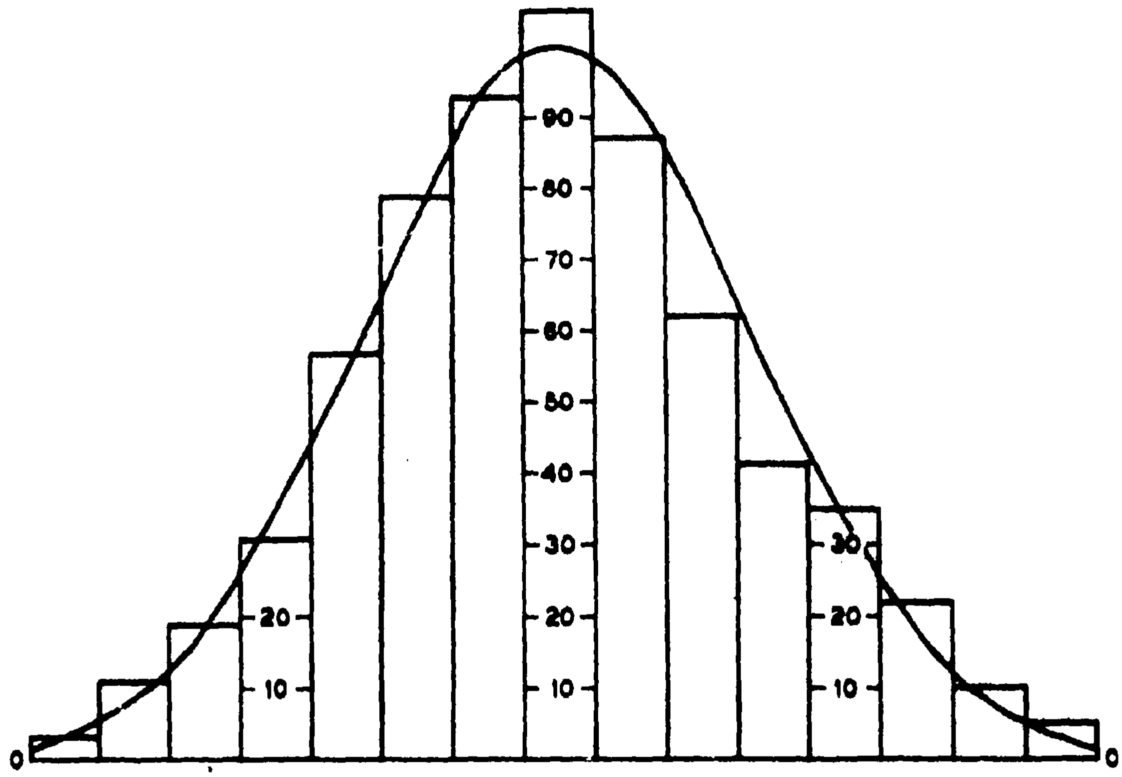


FIG. 1
HISTOGRAM

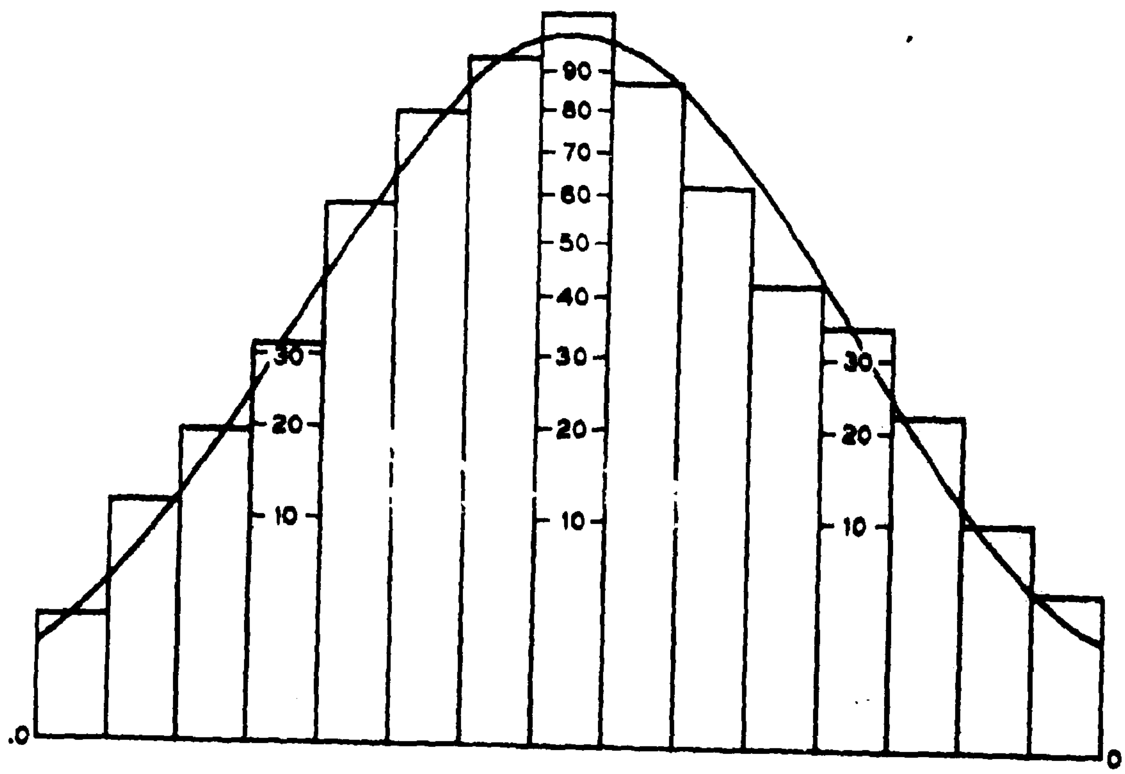


FIG. 2
ROOTOGRAM

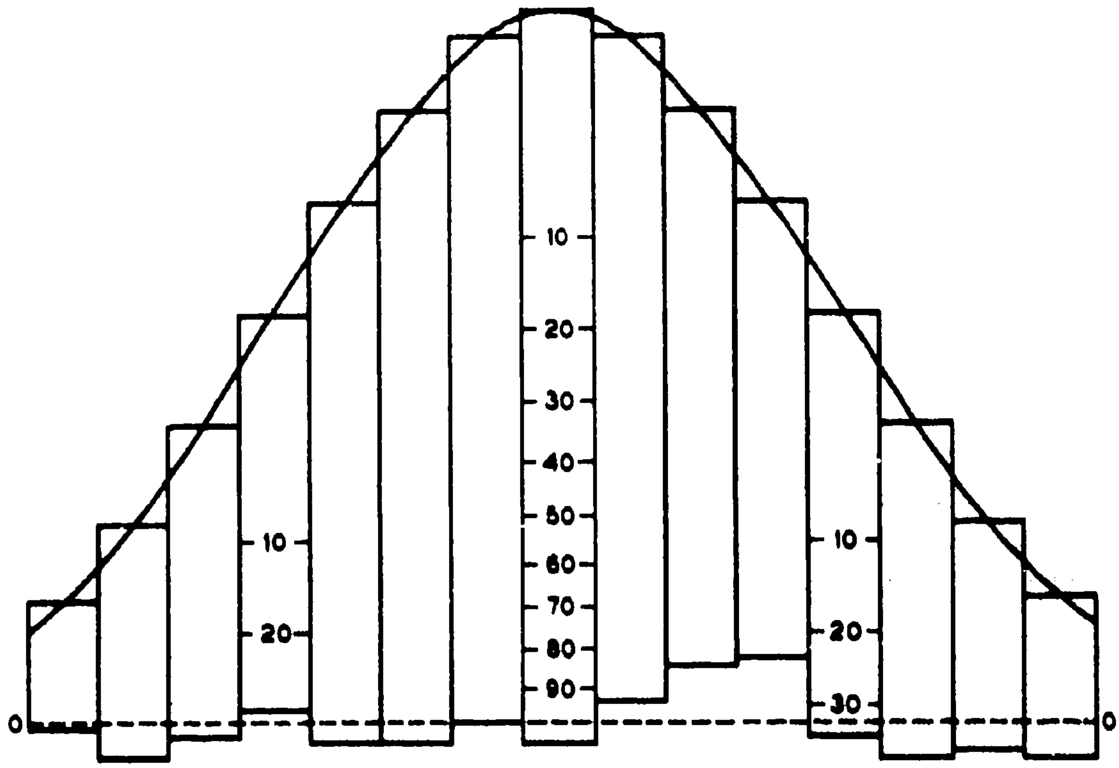


FIG. 3
HANGING ROOTOGRAM

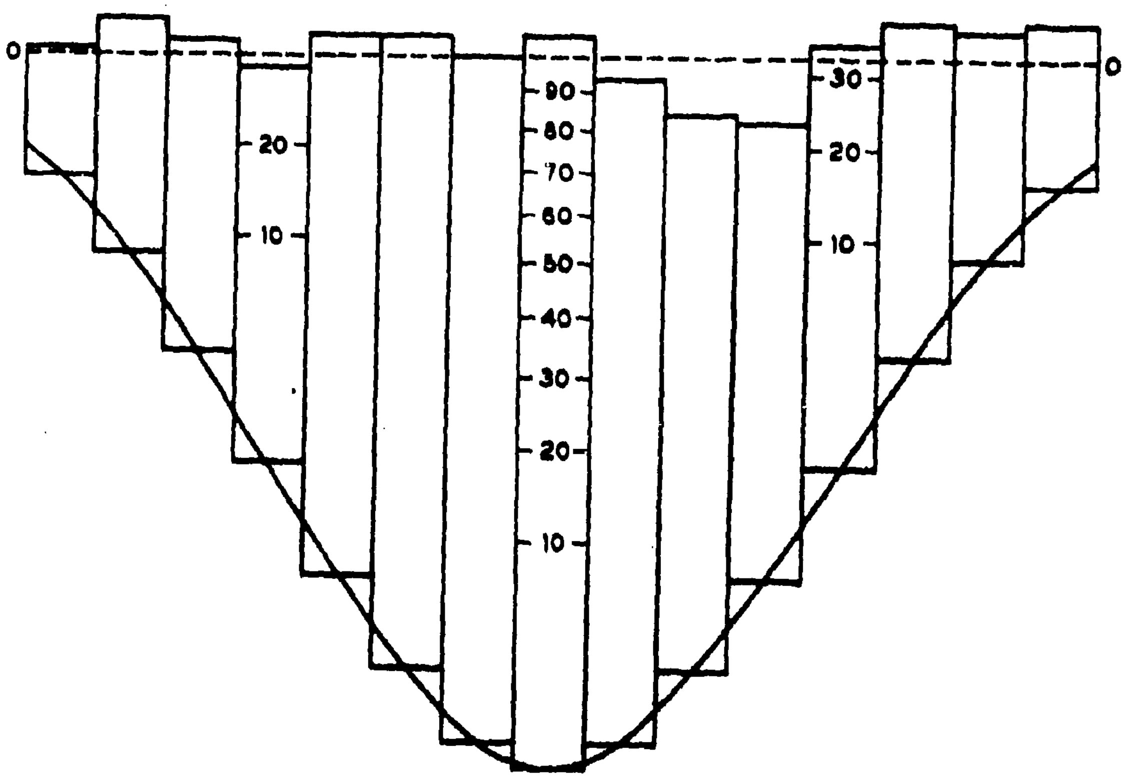


FIG. 4
SUSPENDED ROOTOGRAM

MONTE CARLO TECHNIQUES TO EVALUATE EXPERIMENTAL DESIGN ANALYSIS

M. M. Everett, D. L. Colbert, and L. W. Green, Jr.
Pratt-Whitney Aircraft
Florida Research and Development Center
West Palm Beach, Florida

1. INTRODUCTION. Monte Carlo simulation, using today's high speed computers, has opened new fields in analysis of systems previously unavailable to engineers and mathematicians. Especially valuable are the approaches they offer to study the accuracy and precision of some of the empirical relationships used in analysis of variance and experimental design.

This report gives the results of two such simulation programs made at Pratt & Whitney Aircraft's Florida Research & Development Center. It is not the purpose of this paper to present the findings of these studies as absolute truisms. They are, however, provided as the results of case histories and do offer a method for further exploration of analytical solutions in the field of analysis of variance and experimental design.

The two simulations presented here are Rejection Criteria for Approximate Student's "t" Test, and Bias in the Analysis of Variance Components from an Unbalanced Design.

2. DISCUSSION.

A. REJECTION CRITERIA FOR APPROXIMATE STUDENT'S "T" TEST

One of the simplest designed experiments is that designed to test, or compare, the first moments of two lots or populations. Of interest here is the case of the comparison of two means ($\overline{X}$ and $\overline{Y}$) calculated from the samples drawn from those populations when the population variances, σ_x^2 σ_y^2, are not equal.

This paper reports on an investigation of three different commonly used methods to determine critical values for this situation. The purpose of the investigation was to compare the relative merits of each, assuring that the true level of significance was at least as great as the pre-selected level of significance, and to obtain an unbiased estimator having a minimum variance.

To compare, by a "t" test, two means from independent samples, where it is suspected or known that the variances are unequal, the test would be:

$$\text{(a)} \qquad t_{cal} = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{s_x^2/N_x + s_y^2/N_y}} .$$

When the hypothesis to be tested is that $\mu_x = \mu_y$, this reduces to

$$\text{(b)} \qquad t_{cal} = \frac{\bar{X} - \bar{Y}}{\sqrt{s_x^2/N_x + s_y^2/N_y}} .$$

In the case where $\sigma_x^2 \neq \sigma_y^2$, t_{cal} does not follow the student's "t" distribution with $N_x + N_y - 2$ degrees of freedom. Therefore some critical criterion, such as a modified t-distribution, must be used to judge significance.

1. Methods Used to Determine Critical Values.

Method 1 - Cochran and Cox Approximation

$$t'_{\alpha}(1) = \frac{\frac{s_x^2}{N_x} t_{\alpha x} + \frac{s_y^2}{N_y} t_{\alpha y}}{\frac{s_x^2}{N_x} + \frac{s_y^2}{N_y}} .$$

This method utilizes a weighted mean of the tabular t values for the two samples.

Method 2 - Dixon and Massey Approximation

t'_α (2) = tabulated value of student's "t" associated with γ degrees of freedom where

$$\gamma = \frac{\dfrac{s_x^2}{N_x} + \dfrac{s_y^2}{N_y}}{\left(\dfrac{s_x^2}{N_x} \Big/ N_x + 1\right) + \left(\dfrac{s_y^2}{N_y} \Big/ N_y + 1\right)} - 2 .$$

This approximation assumes that the mean comparisons follow a student's "t" distribution not at $(N_x + N_y - 2)$ degrees of freedom but rather at some γ degrees of freedom.

Method 3 - Satterthwaite-Welch Approximation

t'_α (3) = tabulated value of student's "t" distribution with γ degrees of freedom where

$$\gamma = \frac{\left(\dfrac{s_x^2}{N_x} + \dfrac{s_y^2}{N_y}\right)^2}{\left(\dfrac{s_x^2}{N_x}\right)^2 \Big/ (N_x - 1) + \left(\dfrac{s_y^2}{N_y}\right)^2 \Big/ (N_y - 1)} .$$

This is the approximation for the modified degrees of freedom. It is shown in various texts in different algebraic forms.

In determining t'_α (2) and t'_α (3) it is not necessary to round the degrees of freedom to the lower value; the tables can be interpolated for an unbiased estimate. However, in tables I and II, discussed later, for t'_α (2) and t'_α (3) the lower rounded degrees of freedom were used or the percentiles shown would have been somewhat smaller.

2. Simulation Procedure

These three methods were compared by Monte Carlo simulation on the IBM 7090 and 1620 computers. For a stated set of parameters and sample size, 10,000 samples were drawn from both the X and the Y populations. These samples were randomly paired and their first moments input to equation (b). The output was 10,000 values of t'_{cal}, under the restriction that $E(\overline{X} - \overline{Y}) = 0$. Then all 10,000 values were ranked and the percentiles identified. This process was repeated for various combinations of σ_x^2, σ_y^2, N_x, and N_y.

Approximate rejection values for $t'_\alpha(1)$, $t'_\alpha(2)$, and $t'_\alpha(3)$ were calculated and averaged for α at levels of 90%, 95%, and 99%. The rejection values were then compared to the appropriate percentile level from the ranked values.

3. Conclusions and Discussion.

Tables I and II summarize the test cases that were simulated. The recorded levels for the $t'_{\alpha(i)}$ are the average estimates for the actual levels 90%, 95%, and 99%. Table I is used to demonstrate the output of the simulation process. It compares the three methods when used with equal variances. The only significant conclusion demonstrated is that the Cochran & Cox approximation is an estimator whose confidence level is at least as high as the prior selected confidence level. Table II continues to demonstrate this.

If a bias exists in Method 2 and Method 3 it does so only at certain levels of the parameters and their sample sizes. This indicates that an interaction of the variables exists. Table II compares only a few situations and is entirely too general to draw many exact conclusions. It does, however, give an insight into the comparative accuracies involved.

In additional simulation studies it has been found in all cases observed that the Satterthwaite-Welch Method was a more precise estimator than the Dixon and Massey approximations. Further studies are required to obtain an unbiased estimate with a minimum variance.

It may be noted that an exact solution due to H. Scheffé has been omitted from this study. Scheffé's method is based on the fact that if $n_1 \leq n_2$, a sample of size n_1 may be randomly selected from the larger sample size of size n_2. It is then possible to calculate Scheffé's "t" statistic by:

$$t'_{\alpha}(4) = \frac{\left(\sum_{i=1}^{n_1} \frac{X_{1i}}{n_1} - \mu_1\right) - \left(\sum_{i=1}^{n_2} \frac{X_{2i}}{n_2} - \mu_2\right)}{\sqrt{\dfrac{\sum_{i=1}^{n_1} \left[\left\{X_{1i} - \left(\frac{n_1}{n_2}\right)^{1/2} X_{2i}\right\}^2 - \left\{\sum_{i=1}^{n_1} \left[X_{1i} - \left(\frac{n_1}{n_2}\right)^{1/2} X_{2i}\right]\right\}^2 \Big/ n_1\right]}{n_1 (n_1 - 1)}}}$$

which is distributed as Student's "t" with n_1-1 degrees of freedom. It is immediately obvious that the relative information of this statistic decreases as the value of $n_2 - n_1$ becomes large, since $n_2 - n_1$ observations are randomly eliminated from the calculation of the "t" statistic under the assumption that $n_1 \leq n_2$.

Since this loss of information is especially severe for the case where one sample size is very small, this method was not considered in this study.

B. BIAS IN THE ANALYSIS OF VARIANCE COMPONENTS FROM AN UNBALANCED DESIGN.

It was suspected that a bias existed in the estimates of the components of variance when analysis of variance techniques are applied. It was further suspected that this bias was due to unequal sample sizes. If this bias could be related to sample sizes and sample size ratios then it may be possible to derive an unbiasing technique. Using the IBM 7090 computer, a Monte Carlo Simulator was written to determine if this suspected bias existed and to study the possibilities of finding a method to identify this bias.

1. Simulation Procedure

The simulator was designed to determine the distribution of estimates of process variance from a one-way ANOVA, byproduct of which were estimates of the within-process variation discussed below. To simulate the two sources of variation, two populations of normally distributed random numbers were set up. The means of these distributions were given fixed arbitrary values, while the standard deviations (and therefore, the variances) were variable. The first population was designated as the process-to-process source of variation. The second was designated as the within-process source of variation. It was decided that three ratios of standard deviations of these populations would be used. There were:

$$\frac{\sigma \text{ process-to-process}}{\sigma \text{ within-process}} \text{ of } 0.5,\ 1.0,\ 2.0 \quad .$$

These values were selected because they cover the general area of interest in estimating process-to-process variation. It was further decided that for each ratio above, a control case (balanced data) should be run, in addition to a case with mild unbalance, and a case with extreme unbalance. The balanced case had four runs with 10 data points for each run. The mildly unbalanced case had four runs with 8, 9, 10, and 12 data points each. The extremely unbalanced case had four runs with 5, 3, 10, and 15 data points in each. The analysis of variance described above was carried out for each case on an IBM 7090 computer in the following sequence:

1. Four values were selected from the population of process-to-process random numbers. One of these corresponds to each process.

2. Four sets of numbers were selected from the within-process population. Each of these sets corresponded to one of the four processes. For example, for the control case, each set would have ten random numbers; for the extremely unbalanced case, the set corresponding to the first process would have five members, the second set would have three, etc. The value, selected in step 1, for the first process is added to each member of the set of within-process numbers for the first process; the

second process number is added to each value of the second set, etc. In this manner an array of numbers is produced. Each column has a common process effect while within each column there is a within-process effect.

3. The formulas of the analysis of variance were used to estimate the process-to-process and within-process variances. These values are stored in the computer.

4. Steps 1-3 are repeated 1000 times so that the 1000 estimates of each variance are obtained.

5. The 1000 values are then ranked and the mean, standard deviation, and standard error of the mean are computed. The ranked estimates and computed statistics are then listed.

6. The plot, figure 1, was then made, showing the frequency distribution of the estimates.

For comparison purposes the plots of the control case, the mild unbalanced case, and the extremely unbalanced case are shown together in figure 1 for the ratio 1.0 to 1.0 of standard deviations.

The results of the simulation are summarized in table III. If a bias exists in either the process-to-process or the within-process variance it is not evident here. There is however, a relatively large scatter of the variance estimates.

2. Conclusions and Discussion

Based on the Monte Carlo Simulator, the following conclusions were reached:

1. The distributions of the estimates of within-process variances (discussed in the appendix) were approximately normal. These distributions exhibited a marked central tendency and a degree of symmetry.

2. The bias in the estimates of within-process variability was negligible in each case tested. The cases using unbalanced

data (unequal sample sizes within run) did not demonstrate biases significantly larger than the control case.

These first conclusions were not unexpected and were more or less byproducts of the simulation. The more important conclusions follow:

3. The distributions (figure 1) of the estimates of process-to-process variances were highly skewed to the right and truncated on the left. However, the cases using the unbalanced data were no more skewed than the control case (balanced data).

4. The bias in the estimates of process-to-process variability was negligible in each case tested, including the cases of unbalanced data.

5. Although estimates of process-to-process variance resulting from the analysis of variance technique are not optimum, no known method of improving this situation exists.

The last three conclusions presented represent the main intent of this study. It must be noted that the estimates of process-to-process variance cannot be considered "optimum" estimates. Since an optimum estimate should have minimum variance and minimum total error, the estimates based on analysis of variance techniques cannot be optimized. Any attempt to further optimize them through the use of an unbiasing technique must reduce the variance of the estimates and, at the same-time, increase the relative value of each estimate (because of skewness).

In figure 2 the sample curve for the distribution of estimates of process-to-process variance is skewed and exhibits a large amount of scatter. The distribution of an optimum estimating procedure should have minimum variance and minimum bias, approached by the second curve shown in figure 2. To optimize the present estimates (obtained from ANOVA) of process-to-process variance, the scatter of the distribution of these estimates should be reduced. To accomplish this, each estimate (s^2) should be divided by a factor K, where K is greater than one (1). The proper selection of K will minimize the variance. However, this will bias the estimate so that a correction must be made; that is, the mean estimate of the variance will be only 1/K of the true value. Thus, 1 - 1/K must be added to each estimate s^2 to unbias the estimates. However for any estimate s^2;

$$(1/K)s^2 + (1-1/K)s^2 = s^2/K + s^2 - s^2/K = s^2 .$$

Therefore, for any K selected the estimate is not improved. Any other plan to optimize these estimates will fail since to reduce the scatter a bias must be introduced and to minimize this bias the scatter must be increased.

REFERENCES

1. Anderson, R. L., and Bancroft, T. A. Statistical Theory in Research. New York: McGraw-Hill Book Company, Inc., 1952

2. Bowker, A. H. and Lieberman, G. J. Engineering Statistics. New Jersey: Prentice-Hall, Inc., 1959

3. Brownlee, K. A. Statistical Theory and Methodology in Science and Engineering. New York: John Wiley and Sons, Inc., 1960

4. Cochran, W. G. and Cox, G. M. Experimental Design. New York: John Wiley and Sons, Inc., 1957

5. Davies, Owen L. Statistical Methods in Research and Production. New York: Hafner Publishing Company, 1961

6. Dixon, W. J. and Massey, F. J., Jr. Introduction to Statistical Analysis New York: McGraw-Hill Book Company, 1957.

7. Duncan, A. J. Quality Control and Industrial Statistics. Homewood, Illinois: Richard D. Irwin, Inc., 1959

8. Hald, A. Statistical Theory with Engineering Applications. New York: John Wiley and Sons, Inc., 1962

9. Steel, R. G. D. and Torrie, J. H. Principles and Procedures of Statistics New York: McGraw-Hill Book Company, Inc., 1960

TABLE I

EQUAL VARIANCES

	N_X	N_Y	σX^2	σY^2	Simulated Actual α Level	t'_α (1)	t'_α (2)	t'_α (3)
Case I	10	10	1	1	90	90.8	89.1	89.1
	10	10	1	1	95	95.8	94.0	94.2
	10	10	1	1	99	99.5	98.7	99
Case II	5	3	1	1	90	94.6	89.1	92.6
	5	3	1	1	95	98	95.4	95
	5	3	1	1	99	99.8	97.8	98.2
Case III	15	8	1	1	90	90.7	90.6	90.4
	15	8	1	1	95	95.6	94.1	95.2
	15	8	1	1	99	99.2	99.1	99.1

All α levels expressed in percent;
Underlined values denote underestimates

TABLE II

UNEQUAL VARIANCES

	N_X	N_Y	σX^2	σY^2	Simulated Actual α Level	t'_α (1)	t'_α (2)	t'_α (3)
Case I	15	5	9	1	90.	93.6	92.1	92.0
	15	5	9	1	95.	96.4	95.3	95.3
	15	5	9	1	99.	99.3	97.9	98.6
Case II	5	15	16	1	90.	94	95	95
	5	15	16	1	95.	97	97.3	97.3
	5	15	16	1	99.	99.8	99.4	99.4
Casc III	15	8	.0625	1	90.	92	92	92
	15	8	.0625	1	95.	96.7	96.7	96.7
	15	8	.0625	1	99.	99.8	99.8	99.8
Case IV	12	8	2.25	1	90	91.8	88.1	89.4
	12	8	2.25	1	95	96.0	94.3	94.4
	12	8	2.25	1	99	99.3	98.6	98.6
Case V	5	3	4	1	90	91.4	86.1	86.9
	5	3	4	1	95	95.3	92.0	92.7
	5	3	4	1	99	99.6	97.6	98.1
Case VI	30	18	2.25	1	90	97	93	93
	30	18	2.25	1	95	99.7	97	97
	30	18	2.25	1	99	99.9	98.9	98.9

All α levels expressed in percent;
Underlined values denote underestimates

TABLE III

No. of Process	Inputs: Process-Process Variance	Inputs: Within-Process Variance	Numbers of Observations within each Process	Estimates: Process-Process Variance	Estimates: Within-Process Variance	Variance of the Variance Estimate About its Mean: Process-Process	Variance of the Variance Estimate About its Mean: Within-Process
4	1.0	0.25	8, 9, 10, 12	0.9876	0.2524	0.7381	0.0037
4	1.0	0.25	10, 10, 10, 10	1.0219	0.2543	0.6824	0.0037
4	1.0	0.25	5, 3, 10, 15	0.9588	0.2496	0.8566	0.0044
4	1.0	1.0	8, 9, 10, 12	0.9812	1.0069	0.7564	0.0558
4	1.0	1.0	10, 10, 10, 10	0.9418	0.9999	0.7086	0.0538
4	1.0	1.0	5, 3, 10, 15	1.0297	1.0163	1.0658	0.0657
4	0.25	1.0	8, 9, 10, 12	0.2569	1.0037	0.0854	0.0609
4	0.25	1.0	10, 10, 10, 10	0.2353	1.0057	0.0732	0.0568
4	0.25	1.0	5, 3, 10, 15	0.2624	0.9903	0.1127	0.0657

Process-to-Process Components of Variance
1000 Observations

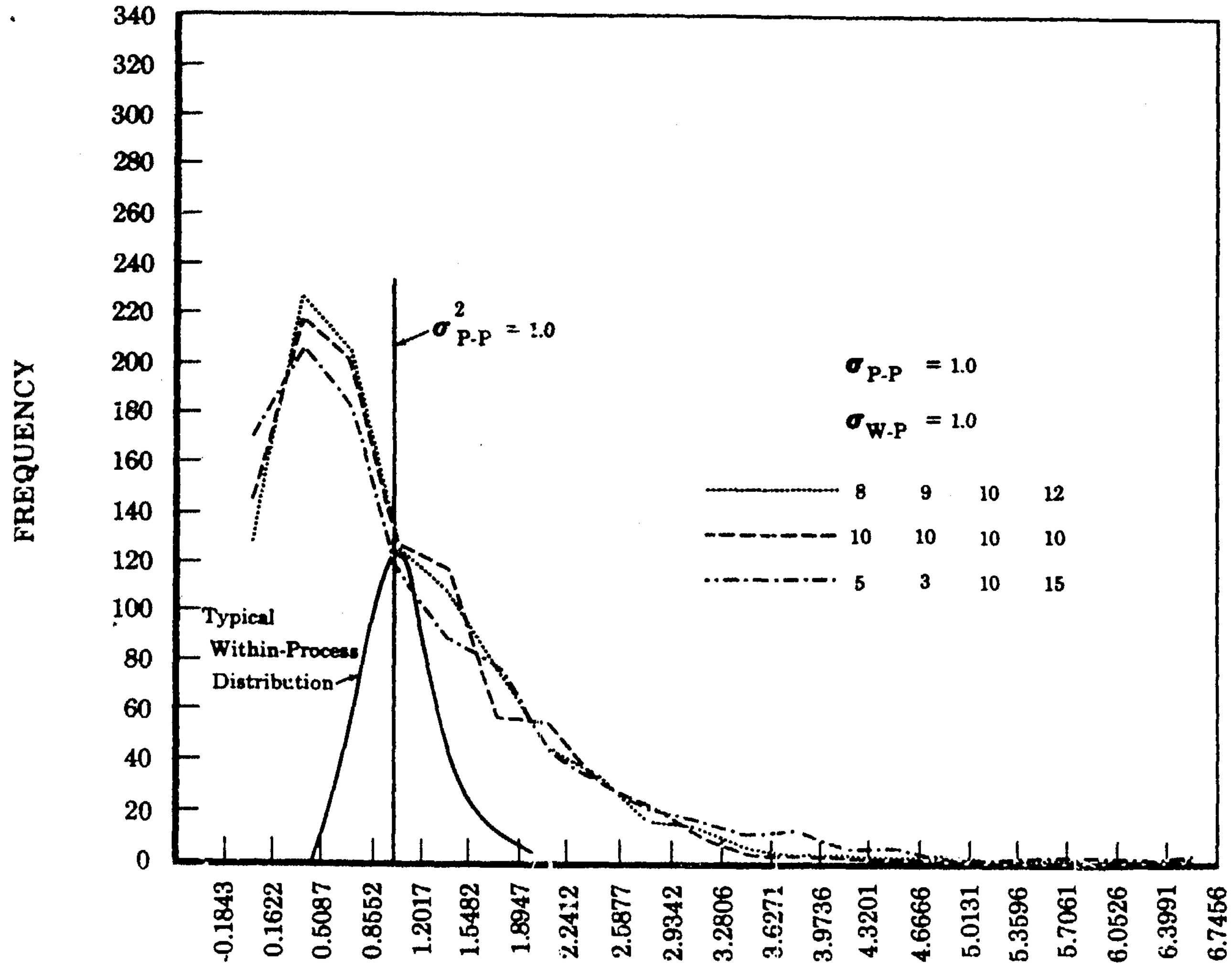
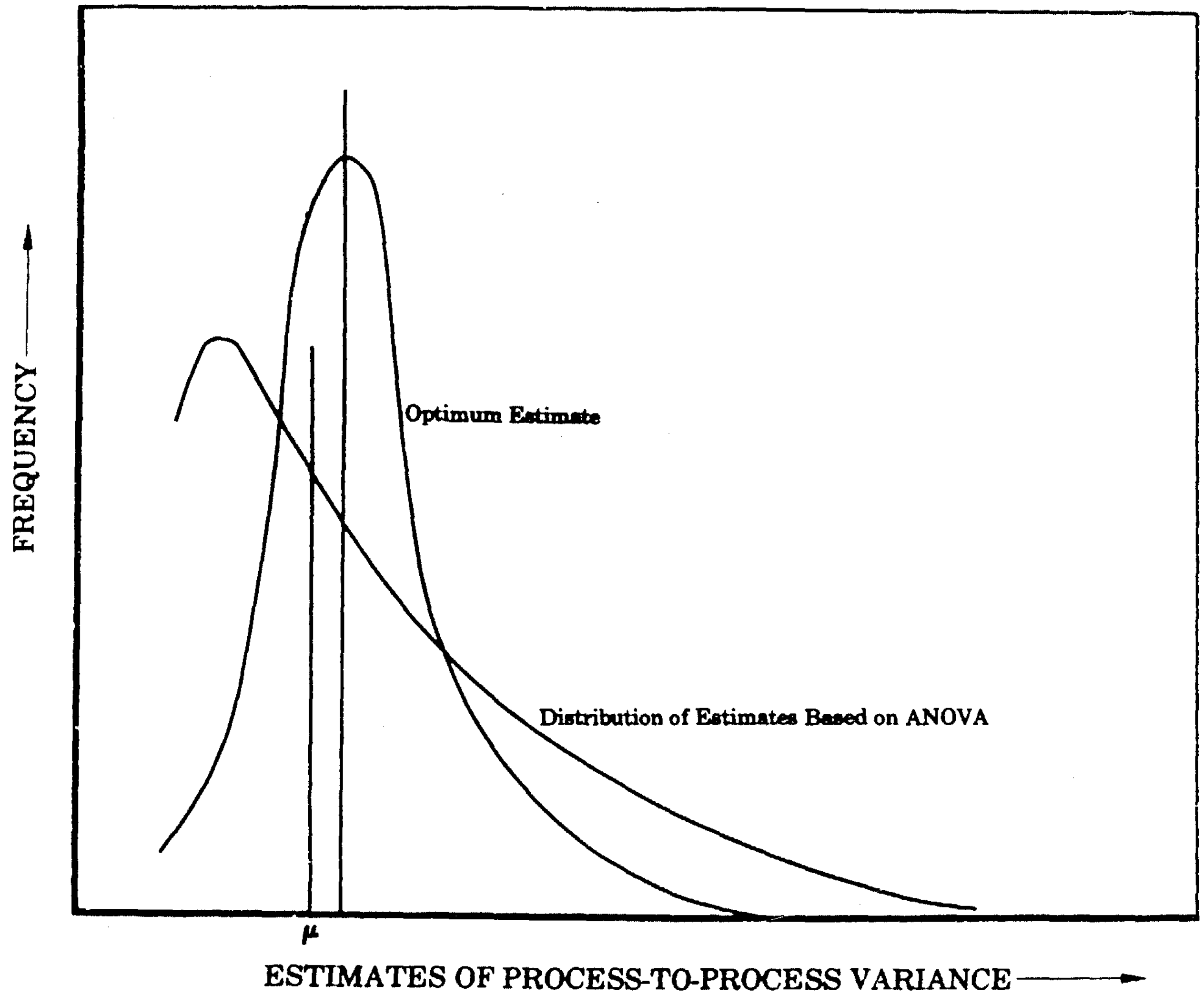


FIGURE I

Frequency Distribution, Process-to-Process Variance



643010
FD9800

FIGURE II

LIST OF ATTENDEES

Addelman, Dr. Sidney	Research Triangle Institute
Ammann, Mr. W. H.	US Army Aviation Materiel Command
Atkinson, Mr. John C.	CRDL, Edgewood Arsenal
Ayres, Mr. James N.	US Naval Ordnance Lab
Bartee, Mr. Edwin M.	Univ of Alabama
Bartko, Dr. John J.	Natl Inst of Mental Health
Bartlett, Mr. Richard P., Jr.	US Dept of Agriculture
Beall, Mr. John R.	US Army Medical R&D Command
Bechhofer, Prof Robert	Cornell Univ
Bercaw, Maj W. W.	US Army Strategy & Tactics Analysis Gp.
Berndt, Mr. Gerald D.	Hq SAC, Offutt AFB
Betts, Maj Genl. Austin W.	Deputy Chief of Res. & Devel., D/A
Box, Prof George E. P.	Univ of Wisconsin
Boyle, Mr. Douglas G.	Dugway Proving Ground
Bright, Mr. Jerry W.	CRDL, Edgewood Arsenal
Brinkmann, Mr. George L.	FDA, Washington, D. C.
Brown, Mr. Ralph E.	Frankford Arsenal
Brown, Mr. Wm. A.	Dugway Proving Ground
Bruno, Mr. O. P.	Army Ballistic Res Lab
Bryson, Dr. Marion R.	Duke Univ
Bulfinch, Mr. Alonzo	Picatinny Arsenal
Burdick, Dr. Donald S.	Duke Univ
Burke, Col James L.	Ft. Huachuca
Bustead, Ronald	US Army Natick Lab
Byas, Mr. W. E.	Picatinny Arsenal
Cameron, Dr. Jos. M.	Stat Engrg Lab, NBS
Carbonaro, Philip A. G.	US Army Materials Res. Agency
Carter, Mr. Frederick L., Jr.	Ft. Detrick
Christianson, Mr. C. J.	RAC
Ciuchta, Mr. H. P.	CRDL, Edgewood Arsenal
Coffman, Rebecca J.	Eng. R&D Lab.
Cogdell, LCdr. John J.	COMOPTEVFOR
Cohen, Prof. A. C., Jr.	Univ of Georgia
Coleman, Mr. Roger D.	Johns Hopkins Univ
Cook, Mr. Charles M., Jr.	Opera. Eval. Group
Coon, Helen J.	Ballistic Res Lab
Cox, Mrs. Claire B.	NIH
Cox, Dr. Edwin L.	USDA
Crancer, Mr. Alfred, Jr.	AF Office of Scientific Res.

Cox, Mr. Paul C.	White Sands Missile Range
Craw, Dr. Alexander R.	Fort Detrick
Curtis, Mr. John J.	Fort Detrick
Cutchis, Mrs. Angeliki D.	Johns Hopkins Univ
Danish, Mr. Michael B.	BRL, Aberdeen
David, Prof H. A.	Univ of N. C.
DeArmon, Mr. Ira A., Jr.	ORG, Edgewood Arsenal
DeCicco, Mr. Henry	USA MUCOM
Demchak, Mr. Peter	Fort Detrick
Dihm, Mr. Henry, Jr.	Redstone Arsenal
Dimling, Lt. John A.	US Army Strategy & Tactics Analysis Gp.
Dobrindt, Mr. Gerald T.	Analytical Lab, Aberdeen
Dressel, Dr. F. G.	ARO-D, Durham
Duncan, Dr. David B.	Johns Hopkins Univ
Dunn, Mr. Paul F.	Booz-Allen
Eisenhart, Dr. Churchill	NBS
Eissner, Mr. Robert M.	BRL, Aberdeen
Emero, Mr. Roland F.	Raytheon Company
Ellerson, Mrs. Elizabeth C.	US Naval Propellant Plant
Ellner, Mr. Henry	Directorate for Quality Assurance, EA
Endelman, Miss Anna	Food & Drug Admin, Bur of Reg. Compl.
Engel, Mr. Klaus H. C.	AERO Space Div
Enis, Mr. Peter	George Washington Univ
Ewart, Mr. Wade H.	US Army Missile Command
Fichter, Mr. Lewis S.	Picatinny Arsenal
Fiddleman, Capt. Paul B.	USCRDL, Edgewood
Foster, Dr. Walter D.	Fort Detrick
Freedle, Dr. Roy O.	Amer. Inst. for Res.
Frishman, Mr. Fred	ARO, OCRD, DA
Galvin, Mr. Cyril J., Jr.	Coastal Engrg Res Center
Galbraith, Dr. A. S.	ARO-D, Durham
Gardner, Dr. Roberta A.	Div of Biol. Stds., NIH
Gehan, Dr. Ed	Natl. Cancer Inst.
Geisser, Dr. Seymour	NIAMD
Glick, Mr. Charles E.	Atlantic Res Corp.
Gliser, Leon J.	Columbia Univ
Goldstein, Col. J. D.	USA R&D Command
Graesel, David B.	Natl Cash Register Co.
Granville, Mr. William, Jr.	Frankford Arsenal

Green, Mr. Lee, Jr.	Pratt-Whitney Aircraft
Green, Mr. Lyle D.	Inst of Sc. & Tech, U of Mich.
Green, Mr. Theron D.	Ft. Detrick
Greenhouse, Dr. Samuel W.	Natl. Inst. of Mental Health
Greenwood, Dr. Joseph A.	Naval Medical Center
Grimes, Mr. James Paskell	Naval Res Lab
Grubbs, Dr. Frank E.	Ballistic Res Labs, Aberdeen
Haberman, Mr. Sol	Johns Hopkins Univ
Haines, Dr. Bertram W.	Ft. Detrick
Hampton, Mr. L. D.	U. S. Naval Ordnance Lab.
Hanson, Dr. Fred S.	White Sands Missile Range
Harris, Dr. Bernard	Math. Res. Center, Madison, Wis
Hartley, Prof. H. O.	Inst. of Stat., Texas A&M Univ
Hassel, Mr. L. D.	Picatinny Arsenal
Hauser, Mr. Arthur L.	FDA, Washington, D. C.
Heaney, Mrs. Marian	Dept of Defense, Ft. Meade
Hecht, Mr. Edward C.	Picatinny Arsenal
Helvig, Mr. T. N.	Honeywell Inc.
Hershner, Dr. Ivan R.	ARO, OCRD, DA
Hess, Mr. Martin	Koppers Company
Hixson, Mr. Eugene E.	GSFC-NASA, Greenbelt
Holmes, Lt Col. Frederick B.	USAR, R&D Unit, Sacramento, Cal.
Homeyer, Mr. Paul G.	CEIR
Hook, Mr. Jonn R.	Hq, DASA
Hoppes, Mr. Harrison N.	RAC, McLean, Va.
Horner, Dr. Theodore W.	Booz-Allen
Howes, Mr. David R.	US Army Strategy & Tactics Analysis Gp
Inselmann, Dr. Edmund H.	Frankford Arsenal
Isaac, Mr. Gerhard J.	US Army Med Res & Nutritions Lab
Itkin, Mr. Arthur	Merck Sharp & Dohme Res Lab
Jacobus, Dr. David P.	WRAIR
Jebe, Prof. Emil H.	Univ of Michigan
Jenkins, Mr. Andrew H.	Redstone Arsenal
Jessup, Dr. Gordon L.	Ft. Detrick
John, Mr. Frank J.	Watervliet Arsenal
Johnson, Mr. Cecil D.	USAPRO
Johnson, Mr. Jerome R.	BRL, Aberdeen
Johnson, Mr. Robert S.	FDA, Wash. D. C.
Jones, Mrs. Marian W.	Ft. Detrick
Jones, Mr. Richard	Johns Hopkins Univ

Karp, Mr. A. E.	US Army Strategy & Tactics Analysis Gp
Katz, Dr. Darryl	Douglas Aircraft
Kaufman, Mr. J. V. Richard	USA MUCOM
Kempthorne, Prof Oscar	Iowa State Univ
Kendall, Dr. M. G.	CEIR
Kiefer, Prof Jack C.	Cornell Univ
Killion, Dr. Lawrence E.	Ft. Huachuca, Ariz
Kinsinger, Mrs. Pauline B.	TRECOM, Ft. Eustis, Va.
Kirby, Dr. Wm. H., Jr.	BRL, Aberdeen
Kirby, Mr. Michael	RAC, McLean, Va.
Knetz, Dr. Wallace J.	Amer. Inst for Res.
Kniss, Mr. James R.	Ballistic Res Labs, Aberdeen
Kokinakia, Mr. Wm.	TBL, BRL, Aberdeen
Kramer, Dr. Clyde Y.	VPI
Krimmer, Mr. Manfred W.	Ammunition Proc & Supply Agency
Kroll, Mr. Wm. F.	Johns Hopkins Univ
Kruse, Mr. Richard H.	Ft. Detrick
Ku, Mr. Hsien H.	Stat Engrg Lab, Natl Bur of Stds
Kupperman, Dr. Morton	Natl. Security Agency, Ft. Meade
Kurkjian, Dr. Badrig	Harry Diamond Labs
Kutger, Dr. Gerald	US Army Strategy & Tactics Analysis Gp
Lane, Mr. Joseph R.	US Army Res Office, Durham
Lawler, Mr. John M.	US Army Natick Labs
LeClerg, Dr. E. L.	Biometrical Svcs, ARS
Lee, Dr. C. Bruce	Science & Tech Div, Library of Congress
Lehman, Dr. Alfred	WRAIR
Lerche, Kenneth D.	Ft. Meade
Lieberman, Prof Gerald J.	Stanford Univ
Lieblein, Mr. Julius	US Post Office Dept
Lieberman, Mr. Herbert J.	Bur of Supplies & Accts
Little, Mr. Robert E.	Oklahoma Stat Univ
Loe, Miss H. V.	Bu of Ships, Navy Dept
Long, Mr. Melvin H.	Frankford Arsenal
Lowery, Mr. Earl D.	US Army Strategy & Tactics Analysis Gp.
Lucas, Prof H. L.	NC State College
Lundegard, Dr. Robert	Office of Naval Res
Lundy, Hazel L.	Springfield Armory
Madden, Mr. Dale A.	Atlantic Res. Corp.
Mall, Mr. Adolph W.	US Army Strategy & Tactics Analysis Gp.

Malligo, Mr. John E.	Ft. Detrick
Maloney, Dr. Clifford J.	NIH
Mandelson, Mr. Joseph	Dir. of Quality Assurance, Edgewood
Mann, Dr. Henry B.	Math Res Center, USA, Univ of Wis
Manthei, Mr. James H.	CRDL, MED RES, Edgewood
Martin, Mr. Francis F.	Booz Allen
Masaitis, Dr. C.	BRL-Aberdeen
Matthews, Mr. Gerald A.	Mississippi State Univ
Matthis, Mr. Carlton L.	Cornell Aeronautical Lab
Mauss, Mrs. Besse Day	Wash., D. C.
McCormick, Mr. Garth	Research Analysis Corp
McGroddy, Miss Patricia	STAG
McIntosh, Mr. Albert L.	Ft. Huachuca, Ariz
McIntosh, Dr. Wm. B.	Ft. Huachuca, Ariz
Menken, Mrs. Jane A.	Natl Inst of Mental Health
Miller, Mrs. Christine	NIH
Mood, Dr. A. M.	US Office of Education
Moshman, Mr. Jack	CEIR Corp
Moss, Mr. David M.	Booz-Allen
Moss, Mr. H. Donald	West. Elec. Corp
Myers, Mr. Raymond H.	VPI
Mylander, Mr. W. Charles	Research Analysis Corp
Nassimbene, Mr. Raymond	US Bureau of Budget
Natrella, Mrs. Mary G.	Stat. Engrg Lab, NBS
O'Connor, Mr. Desmond	GIMRADA, Ft. Belvoir
Oklin, Prof Ingram	Stanford Univ
Olon, Mr. Frederick A.	Naval Propellant Plant
Orleans, Beatrice S.	BuShips
Pabst, Dr. Wm. R., Jr.	Bureau of Naval Weapons
Panos, Mr. Robert	CEIR
Parrish, Dr. Gene B.	ARO-Durham
Parvin, Mr. David W., Jr.	Mississippi State Univ
Parzen, Dr. Emanuel	Stanford Univ
Pepper, Mr. Leonard	USAE Waterways Experiment Station
Perry, Virginia W.	Ft. Lee, Va.
Persweig, Mr. Michael	Picatinny Arsenal
Pettigrew, Mr. Hugh	Natl. Cancer Institute
Pfeiffer, Mr. Otto H.	US Army Tank-Automotive Center
Piepoli, Mr. Carl R.	Ft. Detrick
Podolsky, Mr. Benjamin	Frankford Arsenal

Pollock, Mr. Abraham	OCRD, DA
Quarles, Dr. Gilford G.	Office, Chief of Engrg
Ravenis, Mr. Jos. V. J., II	Johns Hopkins Univ
Rhian, Mr. Morris A.	ORG, Edgewood
Richardson, Mr. B. A.	Canadian Army Opera Res Establishment, Ottawa, Canada
Rider, Dr. Paul R.	Aerospace Res Lab, Wright-Patterson AFB
Riggs, Mr. Charles W.	Ft. Detrick
Riley, Mr. Donald C.	Amer Stat. Assoc
Roberts, Dr. Charles	NIH
Roberts, Mr. Sean C.	Mississippi State Univ
Roetzel, Mr. Thomas G.	Ft. Detrick
Rhode, Dr. Charles A.	Johns Hopkins Univ
Rosenblatt, Dr. H. M.	Bureau of the Census
Rosenblatt, Dr. Joan R.	Natl Bureau of Standards
Ross, Mr. Alan	Johns Hopkins Univ
Roth, Dr. Raymond E.	St. Bonaventure Univ
Rotkin, Mr. Israel	Harry Diamond Labs
Rust, Mr. Philip G.	Winnstead Plantation
Sathe, Dr. Y.	Natl. Cancer Inst.
Schilling, Col. Charles H.	US Military Academy, West Point, N. Y.
Schenker, George	Army Weapons Command
Schmidt, Dr. Th. W.	US Army Res, Durham
Schalten, Roger W.	Boeing Co.
Selig, Mr. Seymour M.	Office of Naval Research
Selman, Mr. Jerome H. N.	MUCOM
Simon, Maj Gen Leslie E. (Ret'd)	Winter Park, Fla.
Sirota, Mr. Milton	Defense Supply Agency
Smith, Dr. Eugene F.	Waterways Experiment Sta.
Snyder, Mr. Mitchell	Univ of Chicago
Soland, Dr. Richard	Research Analysis Corp.
Soni, Mr. Atmaram H.	Oklahoma State Univ
Sorenson, Mr. Richard C.	US Army Personnel Res Office
Speckman, Miss Janace A.	Natl Bureau of Standards
Starr, Dr. Selig	Army Res Office, OCRD, DA
Stearman, Dr. Robert L.	CEIR
Steedman, Mr. Joseph E.	USA CDCOA Aberdeen
Stephanides, Mrs. Agath S.	US Army Strategy & Tactics Analysis Gp.
Sutherland, William H.	Res. Analysis Corp.
Tallis, Mr.	Johns Hopkins Univ

Tank, Capt. Douglas B.	Walter Reed Army Inst. of Res.
Tingey, Lt. Henry B.	Ballistic Res Labs
Trethewey, John D.	DTC
Tukey, Prof. John W.	Princeton Univ
Vick, Mr. James A.	CRDL, Edgewood Arsenal
VonGuerard, Dr. Herman W.	FMC Corp
Wadley, Dr. F. M.	Ft. Detrick
Walner, Arthur H.	US Navy Applied Science Lab
Wallach, Mr. Harold	Public Housing Adm.
Wampler, Roy H.	Natl. Bureau of Standards
Watson, Prof. G. S.	Johns Hopkins Univ
Watson, Mr. H. H.	Dept of Natl Defense, Ontario, Canada
Weingaiten, Dr. Harry	Dept of Commerce
Weinstein, Mr. Joseph	Ft. Monmouth
Weintraub, Gertrude	Picatinny Arsenal
Wenger, Mr. Warren	Ballistic Res Labs, Aberdeen
Westrope, Mr. John	US Army Strategy & Tactics Analysis Gp.
Weyland, Maj Carl E.	Hq, OAR
Wicham, Mr. Robert E.	Picatinny Arsenal
Wigler, Kenneth	Naval Command Systems Support Activity
Wisenfeld, Mr. L.	Picatinny Arsenal
Williams, Henry K.	CRDL, Edgewood
Williams, Mr. Jacob A.	US Army Strategy and Tactics Analysis Gp.
Williams, John N.	United Aircraft Corp.
Willke, Dr. Thomas A.	Natl Bur of Standards
Wolman, Dr. Wm.	NASA-Goddard Space Center
Woodal, Mrs. Rosalie C.	HDL
Youden, Dr. W. J.	Natl Bureau of Standards
Young, Mr. Harold W.	Ft. Detrick
Zelen, Dr. Marvin	Natl. Cancer Inst.
Zimmerman, Mr. J. M.	Rocketdyne
Zundel, Brown	CEIR
Zweifel, Mr. Jim	Natl. Cancer Inst.