Abraham Wald's Work on Aircraft Survivability

MARC MANGEL and FRANCISCO J. SAMANIEGO*

While he was a member of the Statistical Research Group (SRG), Abraham Wald worked on the problem of estimating the vulnerability of aircraft, using data obtained from survivors. This work was published as a series of SRG memoranda and was used in World War II and in the wars in Korea and Vietnam. The memoranda were recently reissued by the Center for Naval Analyses. This article is a condensation and exposition of Wald's work, in which his ideas and methods are described. In the final section, his main results are reexamined in the light of classical statistical theory and more recent work.

KEY WORDS: Survivability; Missing data; Approximate methods; Maximum likelihood.

1. INTRODUCTION

December 7, 1981, was the 40th anniversary of the attack on Pearl Harbor, the subsequent entry of the United States into World War II, and also the birth of the Statistical Research Group (SRG) and the Antisubmarine Warfare Operations Research Group (ASWORG, later renamed the Operations Evaluation Group (OEG) and now part of the Center for Naval Analyses). The early histories of SRG and ASWORG/OEG were described recently by their original leaders, W.A. Wallis (1980) and P.M. Morse (1977), respectively. While in the SRG, Abraham Wald developed techniques for estimating the survivability of aircraft encountering enemy ground fire. Wald's methods were used in World War II and by the Navy and Air Force during the wars in Korea and Vietnam. Although this work was declassified many years ago, it has not appeared in the open literature. At the end of his historical paper, Wallis (1980) mentions that the Wald work will soon appear in print. The papers Wald wrote describing the methods were recently reprinted by the Center for Naval Analyses (Wald 1980); there are eight memoranda, totaling over 100 pages.

The primary goal of this article is to present an expository survey of Wald's work. Wald's work is interesting from several perspectives. It is of historical interest, since the questions Wald addressed were most urgent at the time but are substantively different from questions of interest to the defense establishment today. Second, Wald's work is interesting in the light of more recent developments (e.g., isotonic regression and numerical methods in missing data problems). It is interesting in a third way, too, for it gives us another example of a great mind in action.

In writing this exposition, we have tried to stay as close to Wald's work as possible. We have followed the logical order of the arguments in the order in which he wrote the memoranda. The work is quite complicated, and many of the details are quite technical. For ease of exposition, we have eliminated as many details as possible while attempting to retain cohesiveness and clarity. The reader interested in full details can obtain copies of the original memoranda from the Center for Naval Analyses.

In the next section, the operational and statistical problems are formulated, some sample data are given, and an overview of the SRG memoranda is given. Section 3 is a survey of Wald's work, beginning with the derivation of his basic equation. Various bounds and approximations for the survivability are then derived. The section concludes with a discussion of the effects of sampling errors. In the last section, we reexamine Wald's work in light of classical statistical theory as well as more recent work. This reexamination leads us to the general conclusion that Wald's treatment of these problems was definitive.

2. THE PROBLEMS AND AN OVERVIEW OF WALD'S WORK

2.1 The Operational and Statistical Problems

The *operational* problem can be stated as follows. Aircraft returning from missions have hits by enemy weapons distributed over various parts of the plane (e.g., wings, tail, fuselage, etc.). The operational commander must decide (a) what tactics would improve survivability, and (b) how to reinforce various parts of the aircraft to improve survivability. Reinforcement means, of course, that the aircraft is heavier, and this impairs its mission. The operational commander does not know the distribution of hits on an aircraft that did not return. This is the basic difficulty in making a decision.

The statistical treatment of the problems that Wald studied is complicated by the fact that data on downed aircraft are unobservable. If these missing data were available, survival probabilities could be estimated by the methods of isotonic regression. Without such data, Wald

* Marc Mangel is Associate Professor, Department of Mathematics, and Adjunct Associate Professor, Department of Agricultural Economics, University of California, Davis, CA 95616. Francisco J. Samaniego is Professor, Division of Statistics, University of California, Davis, CA 95616. The authors thank Phil DePoy, Director of the Operations Evaluation Group, for giving them the opportunity to write this article. They also thank the referees, associate editors, and Donald Rubin for extremely helpful and thoughtful comments on the original version of the article.

© Journal of the American Statistical Association
June 1984, Volume 79, Number 386
Applications Section

set to work on the problem of estimating the probability that an aircraft that has sustained a fixed number of hits will survive an additional hit. He also attempted to estimate the survival probability of an aircraft sustaining a hit to one of various portions of the body, with different failure rates (e.g., a hit to the nose is more lethal than a hit to the middle of the fuselage). Wald's problems were compounded by a lack of modern computing equipment, a present-day recourse when one is faced with hard problems that resist analytical solution.

2.2 A Hypothetical Set of Data

Throughout the memoranda, Wald used data to illustrate his methods. Although Wald used different data values to illustrate the analysis, we have redone the calculations for one set of data. This helps one see the usefulness of the more complicated analyses.

The set of data is divided into two subsets. The first subset pertains only to hits on the aircraft, ignoring location of the hit. Assume that 400 aircraft were sent on a mission and that the numbers of aircraft returning with i hits anywhere, A_i, are $A_0 = 320$, $A_1 = 32$, $A_2 = 20$, $A_3 = 4$, $A_4 = 2$, and $A_5 = 2$. The second subset assumes that the location of the hits is known. Subdivide the aircraft into 4 main parts: engines (part 1), fuselage (part 2), fuel system (part 3), everything else (part 4), and let $\gamma(i)$ be the fraction of the area of the aircraft occupied by part i. The total number of hits to all returning aircraft in this case is $\sum_{i=1}^{5} iA_i = 102$. Assume that the hits are distributed according to the following observations:

Part number	$\gamma(i)$	Number of hits (N_i) observed on part
1	.269	19
2	.346	39
3	.154	18
4	.231	26

In anticipation of what follows, let $\delta(i)$ be the fraction of hits observed on part i. Then $\delta(1) = .186$, $\delta(2) = .382$, $\delta(3) = .176$, $\delta(4) = .255$.

These are the kinds of data that the operational commander would obtain and pass on to the statistician working for him. We suggest that the reader now reread the operational problems described in Section 2.1, consider the data again, and then decide how one might attack the problem.

2.3 An Outline of Wald's SRG Memoranda

The basic observational variables are the number N of aircraft participating in the combat, the number A_i of aircraft returning with i hits, and $a_i = A_i/N$. From these data, one wants to find P_i, the probability that an aircraft is downed by the ith hit, and p_i, the conditional probability that an aircraft is downed by the ith hit, given that it received at least $i - 1$ hits and was not downed.

Wald then introduced distributions of hits over the aircraft and found analogous quantities for each subregion of the aircraft. Figure 1 is a flowchart of Wald's work on this problem.

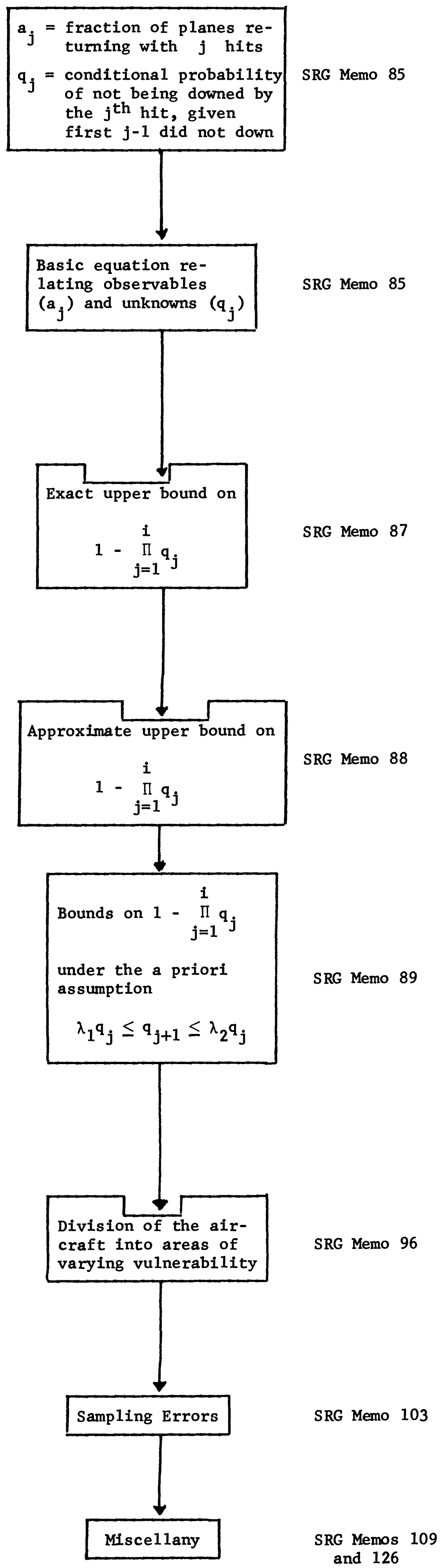

Figure 1. Schematic Outline of Wald's Memoranda.

3. SURVEY OF WALD'S MEMORANDA

This section is a survey of the memoranda. Until Section 3.6, it is assumed that sampling errors are negligible.

3.1 Wald's Basic Equation

In this section, we derive the basic equation satisfied by the probabilities P_i (or $q_i \equiv 1 - p_i$). Let $a_i \equiv A_i/N$ be the fraction of aircraft returning with i hits. Wald assumed that $a_i = 0$ if $i > n$, for some n. Thus, the fraction of aircraft lost is $L = 1 - \sum_{i=0}^{n} a_i$. Wald also assumed that an unhit aircraft always returns and that there is a value m such that the probability of receiving more than m hits is zero. He argued that $m = n + 1$.

Let x_i be the fraction of aircraft downed by the ith hit. (Thus $x_0 \equiv 0$.) Then $\sum_{i=0}^{n} x_i = L$. The x_i's then satisfy the recursion relationship

$$x_i = p_i\left(1 - \sum_{j=0}^{i-1} a_j - \sum_{j=0}^{i-1} x_j\right), \quad i = 1, \ldots, n. \qquad (3.1)$$

The term in brackets in (3.1) is the proportion of aircraft receiving at least i hits. If c_i is defined by $c_i = 1 - \sum_{j=0}^{i-1} a_j$, then (3.1) becomes

$$x_i + p_i \sum_{j=0}^{i-1} x_j = p_i c_i, \quad i = 1, 2, \ldots, n. \qquad (3.2)$$

For some of the analysis, Wald found (3.2) more useful than (3.1). The goal now is to somehow relate the observables (a_j) to the probabilities. In SRG 85, Wald derives the following equation, which is basic to all of his work.

$$\sum_{j=1}^{n} \frac{a_j}{q_1 \cdots q_j} = 1 - a_0. \qquad (3.3)$$

Equation (3.3) relates the observables a_j, the fractions of aircraft returning with j hits, and the unknowns q_j, the conditional probability of not being downed by the jth hit given that the first $j - 1$ hits did not down the aircraft. It is the fundamental equation of the analysis. In the next section, we compare Wald's work with other approaches to this problem. For this reason, it helps to review Wald's derivation of (3.3).

Let b_i be the hypothetical proportion of aircraft hit i times if dummy bullets were used. Then $b_i \geq a_i$; set $y_i = b_i - a_i$. In addition, $y_i = P_i b_i = P_i(a_i + y_i)$. Thus $y_i = (P_i/Q_i)\, a_i$, where as before, $P_i = 1 - q_1 q_2 \cdots q_i$ and $Q_i = q_1 \cdots q_i$. Hence we obtain $y_i = (a_i/q_1 \cdots q_i) - a_i$. Summing up to n and noting that $\sum_{i=1}^{n} y_i = L$ gives (3.3).

Equation (3.3) is easily solved with the simplifying assumption of constant $q_j \equiv q$. For example, for the data, (3.3) becomes the fifth-order equation

$$\frac{.08}{q} + \frac{.05}{q^2} + \frac{.01}{q^3} + \frac{.005}{q^4} + \frac{.005}{q^5} = .20, \qquad (3.4)$$

which yields $q = .851$. Hence p_i, the probability of the ith hit downing the aircraft given that the first $i - 1$ hits did not down it, is $p_i = .149$ (for all i).

Once we know p_i, we can compute x_i, the ratio of the number of aircraft downed by the ith to the total number of aircraft participating, recursively from Equations (3.1) or (3.2). We find that $x_1 = .02980$, $x_2 = .01344$, $x_3 = .00399$, $x_4 = .00190$, and $x_5 = .00087$.

These results are easily obtained, but are based on the assumption of $q_1 = q_2 = \cdots = q_n$. This severely limits their usefulness. The rest of Wald's memoranda studies ways of relaxing this assumption.

3.2 A Least Upper Bound for the Probability of i Hits Downing an Aircraft

Wald's next step was to find a bound on $P_i = 1 - \prod_{j=1}^{i} q_i$, which is the probability of an aircraft being downed by i hits. The bound he found is sharp and its attainment corresponds to the worst case in terms of survivability.

The problem of interest may be written as follows:

$$\text{minimize} \quad \prod_{j=1}^{i} q_j$$

$$\text{subject to} \quad \sum_{j=1}^{n} \frac{a_j}{q_1 \cdots q_j} = 1 - a_0. \qquad (3.5)$$

Equation (3.5) is a nonlinear optimization problem (Avriel 1976). Wald obtained an iterative solution as follows. First he showed that if a set $\{q_1^*, \ldots, q_n^*\}$ solves (3.5), then $q_i^* = q_{i+1}^* = \cdots = q_n^*$; that is, that the q_j are all equal for $j \geq i$.

Applying this result when $i = 1$ means that q_1 is minimized if it satisfies

$$\sum_{j=1}^{n} \frac{a_j}{q_1^j} = 1 - a_0. \qquad (3.6)$$

Assume now that q_1 is known by solving the algebraic equation (3.6). Next one needs to find the value of q_2 that minimizes $q_1 q_2$. Using the result given above, problem (3.5) becomes

$$\text{minimize} \quad q_1 q_2$$

$$\text{subject to} \quad \frac{1}{q_1} \sum_{j=1}^{n} \frac{a_j}{q_2^{j-1}} = 1 - a_0. \qquad (3.7)$$

Straightforward solution via the Lagrange multiplier method gives

$$q_1 = \frac{1}{1 - a_0} \sum_{j=2}^{n} \frac{(j - 1)a_j}{q_2^{j-1}}$$

and

$$\sum_{j=2}^{n-1} \frac{(j - 1)a_{j+1}}{q_2^j} = a_1. \qquad (3.8)$$

Elementary arguments show that these equations have exactly one root in (q_1, q_2).

Wald then generalized this argument to determine the minimum of $\prod_{j=1}^{i} q_j$. He followed the same kind of reasoning, starting with the assumption that $q_j = q_2$, $i \geq j \geq 2$; then one wants to minimize $q_1 q_2^{i-1}$. The Lagrange multiplier method is used again; only the details change.

It is clear that even with present-day computing abilities this approach quickly becomes complicated and time-consuming. In 1943, the task of exact computations was hopeless for any problems of operational interest; thus Wald considered various approximation schemes.

3.3 Approximate Bounds on P_i

Wald's next step was to obtain approximate upper and lower bounds on P_i. Let P_i^* be the maximum value of P_i and let $Q_i^* = 1 - P_i^*$. The first step is to find the lower bound z_i of Q_i^*, that is, to find a bound on the minimum of Q_i. Wald used an interesting kind of hypothetical reasoning: Let y_j be the fraction of the returning aircraft that would be downed if they were to receive $i - j$ additional hits. Then one obtains

$$P_i = \sum_{j=0}^{i-1} y_j + \sum_{j=1}^{i} x_j, \quad i = 1, 2, \ldots, n. \tag{3.9}$$

After some algebraic manipulations, Wald obtained the bounds

$$1 - \frac{\sum_{j=1}^{i} x_j}{\left(1 - \sum_{j=0}^{i-1} a_j\right)} < Q_i < 1 - \sum_{j=1}^{i} x_j. \tag{3.10}$$

Equation (3.10) provides a lower bound on Q_i, once an upper bound on $\sum_{j=1}^{i} x_j$ is known. Wald showed that the maximum value of $X_i \equiv \sum_{j=1}^{i} x_j$ occurs when $p_1 = p_2 = \cdots = p_n = p$. In such a case, the solution of (3.6) gives $q_1 = 1 - p$, and then the x_i are obtained from (3.1). We will let z_i be the lower bound on Q_i obtained in this manner.

Next Wald turned to the problem of estimating an upper bound on the value of Q_i. He showed that such an upper bound is given by

$$t_i = \min[\tilde{u}_1^{i}, \tilde{u}_2^{i-1}, \ldots, \tilde{u}_{i-1}^{2}, \tilde{u}_i], \tag{3.11}$$

where $\tilde{u}_r$ is the positive root of the equation

$$\sum_{j=r}^{n} \frac{a_j}{u^{j-r+1}} = 1 - \sum_{j=0}^{r-1} a_0. \tag{3.12}$$

He obtained (3.11) and (3.12) by a sequence of manipulations on equations analogous to (3.5) and (3.6).

Let us now apply these results to the data. First we will find the lower bound z_i. The first step is to find q_0, the solution of (3.3) when $q_1 = q_2 = \cdots = q_n$. In this case, we have found q_0 as the solution of (3.4); that is, $q_0 = .851$. We have also found the x_i and thus obtain the upper bounds $X_i = \sum_{j=1}^{i} x_j$. For the data, we obtain $X_1 = .02980$, $X_2 = .04324$, $X_3 = .04723$, $X_4 = .04913$, $X_5 = .05000$. According to (3.10), our lower bound is $z_i = 1 - (X_i/(1 - \sum_{j=0}^{i-1} a_j))$. Hence we obtain $z_1 = .85100$, $z_2 = .63967$, $z_3 = .32529$, and $z_4 = .18117$. It is not necessary to calculate z_5, since Q_5 can be obtained directly. In this case, $z_5 = .090909$.

Now consider the upper bounds t_i. Let us write out some of the Equations (3.12), to see what they look like. For $r = 1$, we obtain (3.4), so that $\tilde{u}_1 = .851$. For $r = 2$, (3.12) becomes

$$\frac{a_2}{u} + \frac{a_3}{u^2} + \frac{a_4}{u^3} + \frac{a_5}{u^4} = 1 - a_0 - a_1, \tag{3.13}$$

which has solution $\tilde{u}_2 = .722$. In a similar way, one finds $\tilde{u}_3 = .531$, $\tilde{u}_4 = .333$. The t_i are given by (3.11); namely

$$\begin{aligned} t_1 &= .851 \\ t_2 &= \min(\tilde{u}_1^2, \tilde{u}_2) = .722 \\ t_3 &= \min(\tilde{u}_1^3, \tilde{u}_2^2, \tilde{u}_3) = .521 \\ t_4 &= \min(\tilde{u}_1^4, \tilde{u}_2^3, \tilde{u}_3^2, \tilde{u}_4) = .282. \end{aligned} \tag{3.14}$$

Note that t_5 is not calculated since the exact value of Q_5 can be found.

In Table 1, we compare the exact result obtained by the method of the previous section with lower bound (z_i), upper bound (t_i), and the value obtained assuming all hits are equally lethal (q_0^{i}).

3.4 Bounds on P_i Under Additional Assumptions

The results of the previous section are, from a computational viewpoint, less cumbersome than the exact results. They are still complicated to use, however, so Wald studied the bounds on survival probability under additional assumptions. These assumptions are that

$$\lambda_1 q_j \leq q_{j+1} \leq \lambda_2 q_j, \quad j = 1, 2, \ldots, n - 1 \tag{3.15}$$

for fixed known λ_1 and λ_2, and that

$$\sum_{j=1}^{n} a_j \lambda_1^{-j(j-1)/2} < 1 - a_0. \tag{3.16}$$

Note that (3.16) need not be true if λ_1 is too small; but if λ_1 is close enough to 1, then (3.16) will be true. The basic Equations (3.3) and (3.16) imply that $q_1 < 1$.

Wald first calculated the values of $q_1, \ldots, q_n$, which make Q_i ($i < n$) a minimum. Denote these by $q_1^*, \ldots, q_n^*$. By using a straightforward proof by contradiction, he proved the following: (a) for $j = i, i + 1, \ldots, n - 1$, $q_{j+1}^* = \lambda_2 q_j^*$; and (b) if j is the smallest integer such

Table 1. Exact and Approximate Values of Q_i

	Value			
i	Exact Value	Lower Bound	Upper Bound	Equal Lethality of Hits
1	.851	.851	.851	.851
2	.721	.640	.722	.724
3	.517	.325	.521	.616
4	.282	.181	.282	.525

that $q_{k+1}^* = \lambda_2 q_k^*$ for all $k \geq j$, then $q_r^* = \lambda_1 q_{r-1}^*$ for $r = 2, 3, \ldots, j - 1$. These results can be viewed as analogs of the results in Section 3.3.

Let E_{ir}, $r = 1, \ldots, i - 1$, be the minimum value of Q_i under the restriction that $q_{j+1} = \lambda_1 q_j$, $j = 1, \ldots, r - 1$, and $q_{j+1} = \lambda_2 q_j$ for $j = r + 1, \ldots, n - 1$. The above results show that $Q_i = \min\{E_{i1}, E_{i2}, \ldots, E_{i,i-1}\}$. The results in Sections 3.2 and 3.3 show how the E_{ir} can be calculated. In particular, Wald showed that if g_r is the positive root (in q) of the equation (for $r = 0, 1, 2, \ldots, i - 1$)

$$\sum_{j=1}^{r+1} a_j \lambda_1^{-j(j-1)/2} q^{-j} + \sum_{j=1}^{n-r-1} \{a_{r+1+j} \lambda_1^{-r(r+1)/2-rj}\} \times \{\lambda_2^{-j(j+1)/2} q^{-(r+1+j)}\} = 1 - a_0 \quad (3.17)$$

then an approximation to E_{ir} is

$$E_{ir} \simeq \lambda_1^{r(r+1)/2+r(i-r-1)} \lambda_2^{(i-r)(i-r-1)/2} g_r^i. \quad (3.18)$$

Similar arguments show that if $q_1^*, \ldots, q_n^*$ are values of q_j minimizing $Q_n = \prod_{j=1}^n q_j$, then $q_{j+1}^* = \lambda_1 q_j^*$, $j = 1, \ldots, n - 1$. This means that if q is the root of the equation

$$\sum_{j=1}^{n} a_j \lambda_1^{-j(j-1)/2} q^{-j} = 1 - a_0, \quad (3.19)$$

then the minimum value of Q_n is $\lambda_1^{n(n-1)/2} q^n$.

Wald proceeded in the same fashion to show that the maximum of Q_n is $\lambda_2^{n(n-1)/2} q^n$, where q is a solution of the (3.19) with λ_1 replaced by λ_2.

There is a quantity analogous to E_{ir}. Namely, if D_{ir} is the maximum of Q_i under the restriction that $q_{j+1} = \lambda_1 q_j$ for $j = r + 1, \ldots, n - 1$ and $q_{j+1} = \lambda_2 q_j$ for $j = 1, \ldots, r - 1$, then Wald showed that the maximum of Q_i is $\max\{D_{i1}, \ldots, D_{i,i-1}\}$. He showed that a good approximation to D_{ir} is obtained from (3.17) and (3.18) with the λ_1 and λ_2 interchanged.

We apply these results to the data with $\lambda_1 = .85$, $\lambda_2 = .95$. It is easy to check that (3.16) is satisfied.

To find the lower limit of Q_i, the four equations (for $r = 0, 1, 2, 3$) (3.17) must be solved. For example, for $r = 0$ this equation is

$$\frac{a_1}{q} + \frac{a_2}{\lambda_2 q^2} + \frac{a_3}{\lambda_2^3 q^3} + \frac{a_4}{\lambda_2^6 q^4} + \frac{a_5}{\lambda_2^{10} q^5} = 1 - a_0. \quad (3.20)$$

The roots of (3.17) for the values $r = 0, 1, 2, 3$ are $g_0 = .887$, $g_1 = .938$, $g_2 = .964$, and $g_3 = .979$. Next, the E_{ir} are found approximately from (3.18), and then Q_i is the minimum of the E_{ir}. Table 2 shows the results of such calculations. The lower limit of Q_5 is found by using (3.19). In this case, the root of (3.19) is $q = .986$ and the lower limit of $Q_5 = \lambda_1^{10} q^5$ is .183.

To find the maximum value of Q_i, the same procedure is followed. Since the details are the same, only the final results will be given. Table 3 shows both bounds.

Table 2. Estimating the Minimum of the Survival Probability

i	r	g_r	E_{ir} Approximately	min Q_i Approximately
1	0	.887	.887	.887
2	0	.887	.747	
	1	.938	.747	.747
3	0	.887	.598	
	1	.938	.567	
	2	.964	.550	.550
4	0	.887	.455	
	1	.938	.408	
	2	.964	.364	
	3	.979	.347	.347

3.5 Analysis of Vulnerability Areas of the Aircraft

Wald considered next the problem of determining the vulnerability of different parts of the aircraft. The idea here is that the location of the hits on returning aircraft provides useful information on the vulnerability of various parts of the aircraft. Wald began with the premise that one knows the conditional probability $\gamma_i(i_1, \ldots, i_k)$ that area m will receive i_m hits given a total of $i = \sum_{m=1}^k i_m$ hits. He argued that $\gamma_i(i_1, \ldots, i_k)$ can be experimentally determined by firing dummy bullets at real aircraft. The quantity of interest here is $Q_i(i_1, \ldots, i_k)$, the probability that an aircraft is not downed given i_m hits to area m, with $\sum_{m=1}^k i_m = i$. Wald first formulated the problem in a very general setting, where it is essentially intractable.

To make any progress, he needed to introduce an assumption of independence. Thus, he assumed that if $q(i)$ is the probability that one hit on area i will not down the aircraft and if $\gamma(i)$ is the conditional probability that area i is hit given that one hit occurred, then

$$Q_i(i_1, \ldots, i_k) = \prod_{m=1}^{k} [q(m)]^{i_m}, \quad (3.21)$$

$$\gamma_i(i_1, \ldots, i_k) = \frac{i!}{\prod_{m=1}^{k} i_m!} \prod_{m=1}^{k} [\gamma(m)]^{i_m}. \quad (3.22)$$

In (3.21) and (3.22), it is understood that $\sum_{m=1}^k i_m = i$. Let $\delta(i)$ be the probability that area i is hit, given that the aircraft received exactly one hit that did not down it. Then

Table 3. Lower and Upper Bounds on Q_i

i	Lower Bound on Q_i	Upper Bound on Q_i
1	.887	.986
2	.747	.826
3	.550	.631
4	.347	.463
5	.183	.329

by its definition

$$\delta(i) = \frac{\gamma(i)q(i)}{\sum_{i=1}^{k} \gamma(i)q(i)}. \tag{3.23}$$

In (3.23), recognize the summation as the probability q that a single shot did not down the aircraft. Under the assumption of independence, q will satisfy (3.3) with $q_j \equiv q$ and may be replaced by the solution to that equation. Equation (3.23) is rewritten as

$$q(i) = \frac{\delta(i)q}{\gamma(i)}, \tag{3.24}$$

where $\gamma(i)$ is assumed to be known from auxiliary tests or equated with the proportion of surface area associated with part i, and $\delta(i)$ may be estimated from the data as

$$\delta(i) = \frac{\sum_{j_k} \cdots \sum_{j_1} j_i a(j_1, \ldots, j_k)}{\sum_{j_k} \cdots \sum_{j_1} (j_1 + \cdots + j_k) a(j_1, \ldots, j_k)}. \tag{3.25}$$

The interpretation of $\delta(i)$ is that it is the ratio of the total number of hits in area i of the returning aircraft to the total number of hits on the returning aircraft. Thus, $\delta(i)$ is empirically determined and $q(i)$ is computed by applying (3.23) to the data. Such analyses have actually been performed on real data, with success.

We apply this approach to the data. We have already seen that the positive root of (3.3) with equal q_j is $q_0 = .851$. Thus q_0 is the overall probability of surviving a hit. The probability of surviving a hit to part i is given by (3.24). The q in (3.4) is q_0; $\gamma(i)$ (the fraction of area occupied by part i) and $\delta(i)$ (the fraction of hits to part i) were given along with the data. The results of the calculations are shown in Table 4. For these data, the most vulnerable portion of the aircraft is the engine area.

3.6 Effects of Sampling Errors

Wald considered sampling errors in the special case of equal (but unknown) q_j, and he derived confidence limits for q.

In the absence of sampling errors, the x_i are recursively defined by (3.1) with equal p_i. When there are sampling errors, (3.1) is replaced by

$$x_i = \bar{p}_i\left(1 - \sum_{j=0}^{i-1} a_j - \sum_{j=1}^{i-1} x_j\right), \tag{3.26}$$

Table 4. Probability of Surviving a Single Hit to a Given Part

Part	Probability of Surviving a Single Hit
Entire Aircraft	.851
Engine	.588
Fuselage	.940
Fuel System	.973
Others	.939

where $\bar{p}_i$ has the distribution of the success ratio in a sequence of $N_i = N(1 - \sum_{j=0}^{i-1} a_j - \sum_{j=1}^{i-1} x_j)$ independent trials. Still assuming that $x_i = 0$ for $i > n$ (which is not really true for the case with sampling errors), the basic equation (3.3) becomes

$$\sum_{j=1}^{n} \frac{a_j}{\bar{q}_i \cdots \bar{q}_j} = 1 - a_0. \tag{3.27}$$

Here $\bar{q}_j = 1 - \bar{p}_j$ is an estimate for q; but the $\bar{q}_j$'s are unknown.

Wald derived confidence bounds in the following manner. Consider a hypothetical experiment in which b_i is the fraction of aircraft that would be hit exactly i times if dummy bullets were used. The distribution of Na_i is the same as the distribution of the number of successes in a sequence of Nb_i independent trials, each trial having a probability of success q^i. This gives

$$E(a_i/q^i) = b_i, \quad \operatorname{var}(a_i/q^i) = \frac{b_i(1 - q^i)}{Nq^i}. \tag{3.28}$$

Summing (3.28) gives

$$E\left(\sum_{i=1}^{n} a_i/q^i\right) = \sum_{i=1}^{n} b_i = 1 - a_0,$$

$$\operatorname{var}\left(\sum_{i=1}^{n} \frac{a_i}{q^i}\right) = \sum_{i=1}^{n} \frac{b_i(1 - q^i)}{Nq^i}. \tag{3.29}$$

For moderate to large N, appeal to the central limit theorem and conclude that if

$$\int_{-\lambda_\alpha}^{\lambda_\alpha} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \alpha,$$

then an α confidence interval for q is

$$1 - a_0 - \lambda_\alpha\left(\sum_{i=1}^{n} \frac{b_i(1 - q^i)}{Nq^i}\right)^{1/2} \le \sum_{i=1}^{n} \frac{a_i}{q^i} \le 1 - a_0 + \lambda_\alpha\left(\sum_{i=1}^{n} \frac{b_i(1 - q^i)}{Nq^i}\right)^{1/2}. \tag{3.30}$$

The only trouble with (3.30) is that the b_i are not known. Again appealing to limit theorems, Wald replaced b_i by a_i/q^i (this replacement is accurate to $O(1/\sqrt{n})$). Hence we obtain a confidence interval of the form

$$1 - a_0 - \lambda_\alpha\left(\sum_{i=1}^{n} \frac{a_i(1 - q^i)}{Nq^{2i}}\right)^{1/2} \le \sum_{i=1}^{n} \frac{a_i}{q^i} \le 1 - a_0 + \lambda_\alpha\left(\sum_{i=1}^{n} \frac{a_i(1 - q^i)}{Nq^{2i}}\right)^{1/2}. \tag{3.31}$$

A final simplification is achieved by another appeal to a limit theorem. If q_0 is the root of (3.3) with equal q_i, then as $N \to \infty$, $q \to q_0$, so Wald replaced q^{2i} by ${q_0}^{2i}$ in (3.31), and the resulting confidence limit is now very simple.

These results can be summarized in the following elegant fashion. If a_i are subject to sampling error and q is the true parameter, then $\sum_{i=1}^{n} a_i/q^i$ is normally distributed with mean $1 - a_0$ and variance given by (3.29).

To show how this works, we will derive the 95% and 99% confidence intervals for the data. The first step is to find the positive solution, q_0, of (3.3) with equal q_j. In this case, $q_0 = .851$. The second step is to find the approximate variance of $\sum_{i=1}^{n} a_i/q^i$. This variance is

$$\sigma^2 = \sum_{i=1}^{n} a_i(1 - q_0{}^i)/Nq_0{}^{2i}, \tag{3.32}$$

and in this case we find $\sigma = .01373$. According to (3.31), the confidence limits are found by solving

$$\sum_{i=1}^{n} \frac{a_i}{q^i} = 1 - a_0 \pm \lambda_\alpha \sigma, \tag{3.33}$$

where $\lambda_\alpha = 1.960, 2.576$ for the 95% and 99% limits, respectively. For the 95% confidence limit on q_0, the solution of (3.33) gives [.797, .921] and for the 99% confidence limit, [.782, .947].

3.7 Miscellany

SRG Memoranda 109 and 126 deal, very briefly, with these topics: (a) factors that are nonconstant in combat, (b) nonprobabilistic interpretation of the results, (c) the situation when $\gamma(i)$ are unknown, and (d) vulnerability to different kinds of guns. The most interesting of these topics is the last one, in which Wald generalizes the previous work to include different kinds of weapons. Namely, instead of working with $q(i)$, the probability that an aircraft survives a hit to part i, he works with $q(i, j)$, the probability that an aircraft survives a hit to part i by weapon type j. The generalization is conceptually straightforward, although the details are complicated.

4. DISCUSSION

In this section, we propose to reexamine Wald's work on aircraft survivability, relating his results to classical statistical theory as well as to more recent statistical thought. We believe that such a development makes Wald's recommendations more easily understood. It also allows us to support the general conclusion that Wald's treatment of this problem was definitive, since, through this reexamination, we are able to identify the optimal character of Wald's estimators and to explain why treatment of more general problems is impossible with the data Wald had available to him.

Let us consider the first data set. Wald does not explicitly discuss a model for the data he seeks to fit. It is clear, however, that the appropriate model is multinomial. It is also clear that there are missing data. It is useful to picture the data as embedded in the following scheme.

$$\begin{array}{cccccc} X_{01} & X_{11} & X_{21} & X_{31} & X_{41} & X_{51} \\ & X_{12} & X_{22} & X_{32} & X_{42} & X_{52} \end{array} \tag{4.1}$$

where X_{i1} = the number of aircraft returning with i hits, and X_{i2} = the number of aircraft downed with i hits. Data Set 1 amounts to X_{i1}, $i = 0, \ldots, 5$, while X_{i2}, $i = 1, \ldots, 5$ are unobservable. The multinomial distribution based on 400 observations classified into 11 cells represents the full model for the collection $\{X_{ij}\}$. Let the parameters of the full model be denoted by $\{p_{ij}\}$. Wald prefers to use the parameterization:

(1) $p_{01}, \ldots, p_{51}$ (for which $a_0, \ldots, a_5$ are the corresponding sample proportions in Wald's notation)
(2) $Q_1, \ldots, Q_5$, where

$$Q_i = \frac{p_{i1}}{p_{i1} + p_{i2}}. \tag{4.2}$$

Whatever the parameterization, the critical fact vis-a-vis the estimation problem of interest is that the full model is determined by 10 parameters while the available data have only six degrees of freedom. Put another way, the 10-parameter model for the available data is not identifiable; indeed, the likelihood depends on $\{p_{12}, \ldots, p_{52}\}$ only through the value of $\sum_{i=1}^{5} p_{i2}$. The nonidentifiability of the model for X_{i1}, $i = 0, \ldots, 5$ explains the role of the assumption

$$Q_i = q^i \quad \text{for all } i. \tag{4.3}$$

This restriction renders the estimation problem well defined. The necessity of identifiability also dictates the assumption (for the purpose of analyzing the data set) that the probability of sustaining more than five hits is zero.

We now turn to the derivation of the maximum likelihood estimators for the parameters of the multinomial distribution with missing data under the restriction (4.3). Initially, we write the likelihood as

$$\mathcal{L} \propto \left(\prod_{i=0}^{5} p_{i1}{}^{x_{i1}}\right)\left(1 - \sum_{i=0}^{5} p_{i1}\right)^{400 - \sum_{i=0}^{5} x_{i1}}.$$

The likelihood equations

$$\left\{\frac{\partial}{\partial p_{i1}} \mathcal{L} = 0\right\}_{i=0}^{5}$$

are equivalent to

$$\hat{p}_{i1} = \frac{x_{i1}}{N} \qquad i = 0, \ldots, 5.$$

Now, the parametric analog of Wald's fundamental equation (3.3) is

$$\sum_{j=1}^{n} \frac{p_{j1}}{\prod_{i=1}^{j} q_i} = 1 - p_{01}. \tag{4.4}$$

The latter equation can be shown to be algebraically equivalent to

$$\sum_{j=1}^{n} (p_{j1} + p_{j2}) = 1 - p_{01}, \tag{4.5}$$

which simply specifies that all cell probabilities sum to one. Under restriction (4.3), Equation (4.4) becomes

$$\sum_{j=1}^{n} \frac{p_{j1}}{q^j} = 1 - p_{01}, \tag{4.6}$$

specifying q implicitly as a function of $\{p_{i1}, i = 0, \ldots, n\}$. Now, let $\hat{q}$ be the solution of (3.3), which, for the first data set, can be written as

$$\sum_{j=1}^{5} \frac{\hat{p}_{j1}}{\hat{q}^j} = 1 - \hat{p}_{01}. \tag{4.7}$$

From the invariance property of the MLE's, it is clear that $\hat{q}$ is the MLE of the parameter q.

The regularity of the multinomial model implies the asymptotic optimality of Wald's estimators of the parameters $\{p_{i1}\}$ and p. Wald's confidence interval for the survival probability q can be obtained via MLE theory and thus, its optimality in large samples can be asserted. Since interesting larger models cannot be treated with the data available, Wald's estimation results are, with a sufficiently large sample size, the best possible. For larger models, Wald appropriately turns to the development of bounds on survival probabilities.

Two important areas of statistical analysis having some bearing on Wald's work have been developed since Wald's time. The first is the area of isotonic regression, a subject treated in depth in the recent book by Barlow et al. (1972). The second is the treatment of problems with missing data via the EM algorithm (see Dempster, Laird, and Rubin 1977). Isotonic regression would appear to be an appropriate methodology in Wald's problem, since aircraft vulnerability undoubtedly increases with the number of hits sustained; that is, it is reasonable to expect that $p_1 \leq p_2 \leq \cdots \leq p_n$. In spite of its intuitive appeal, the isotonic version of Wald's problem suffers from nonidentifiability, since ordering of parameters does not reduce the dimension of the parameter space. Thus, given Wald's data, estimation via the methods of isotonic regression proves impossible without additional assumptions. If complete data were available, the unrestricted MLE's for the q_i's are given by

$$\prod_{j=1}^{i} \hat{q}_j = \frac{x_{i1}}{x_{i1} + x_{i2}} \qquad i = 1, \ldots, 5. \tag{4.8}$$

The problem of "isotonizing" these estimates is formally equivalent to the problem of estimating ordered binomial parameters treated by Barlow et al. (1972, p. 102).

The EM algorithm does not help for similar reasons. When the model is not identifiable, a starting value $\mathbf{p}^{(0)}$ for the parameter produces expected $\mathbf{X}$ values, which in turn produce $\mathbf{p}^{(1)} = \mathbf{p}^{(0)}$. In the reduced model, subject to (4.3), one can treat maximum likelihood estimation analytically, and there is no need to employ the EM algorithm.

Let us now examine Wald's estimators for the survival probabilities of various aircraft sections. The portion of the data set classifying hits by part can be viewed as embedded in the array

$$\begin{matrix} Y_{11} & Y_{21} & Y_{31} & Y_{41} & N_1 \\ Y_{12} & Y_{22} & Y_{32} & Y_{42} & N_2 \end{matrix} \tag{4.9}$$

where Y_{i1} = # of hits to part i on returning aircraft; Y_{i2} = # of hits to part i on downed aircraft; $N_1 = \sum_{i=1}^{4} Y_{i1}$; $N_2 = \sum_{i=1}^{4} Y_{i2}$. The data consist of Y_{i1}, $i = 1, \ldots, 4$ and N_1, while Y_{i2}, $i = 1, \ldots, 4$ and N_2 are unobservable. Define the following events:

$$A_i = \{\text{the } i\text{th section is hit}\}$$

$$A = \{\text{the aircraft is hit}\}$$

$$B = \{\text{the aircraft is not downed}\}.$$

Wald's parameters may be identified as

$$q = P(B \mid A),\ q(i) = P(B \mid A_i)$$

$$\delta(i) = P(A_i \mid A \cap B),\ \gamma(i) = P(A_i \mid A). \tag{4.10}$$

With complete data as pictured in (4.9), the MLE's of $q(i)$ are simply

$$\hat{q}(i) = \frac{Y_{i1}}{Y_{i1} + Y_{i2}} \qquad i = 1, \ldots, 4. \tag{4.11}$$

With the incomplete data available to Wald, one must make use of the structural relationship (3.23) (which is immediate from the definitions in (4.10)) and the assumption that $\gamma(i)$, $i = 1, \ldots, 4$ are known. Wald explicitly remarks on the impossibility of estimating $\gamma(i)$ and $q(i)$ simultaneously from his data. However, MLE's for $\{\delta(i)\}$ and q may be obtained from the data, and the estimates

$$\hat{q}(i) = \frac{\hat{\delta}(i)}{\gamma(i)} \cdot \hat{q} \qquad i = 1, \ldots, 4 \tag{4.12}$$

are maximum likelihood estimates by invariance, provided these estimates lie in the unit interval. Wald does not deal with estimation problems in which one or more of the estimates $\hat{q}(i)$ exceed one. In such cases, the MLE of the vector $(q(1), \ldots, q(4))$ lies on the boundary of the parameter space, and its identification is tedious but straightforward.

In our discussion of Wald's formulation and solution of a variety of problems dealing with aircraft survivability, we have mentioned a number of assumptions he imposed to obtain closed-form solutions or efficient bounds. These assumptions deserve scrutiny. Among the assumptions one encounters are (a) constant vulnerability, that is, $q_i \equiv q$, which is an independence assumption; (b) known bounds on rate of growth of vulnerability, that is, $\lambda_1 q_j \leq q_{j+1} \leq \lambda_2 q_j$; and (c) independence of survival among and within areas of different vulnerability. The main cause for concern regarding these assumptions is that the data available do not provide a means for investigating their validity. Consider assumption (a), for example. With complete data (corresponding to $\{x_{ij}\}$ in (4.1)) one could investigate statistically, via a likelihood ratio test or otherwise, the validity of the assumption $q_i \equiv q$. With the type of data available to Wald, such an option is not open because of the lack of identifiability of larger models. Wald cautioned his readers that the solution he provides should be used only "if it is known *a priori* that $q_1 = q_2 = \cdots = q_n$." How and whether such a priori knowledge could be garnered is open to debate. Wald does provide an option for those who are more conservative. The lower bounds for Q_i may be considered conservative estimates of survival probabilities, although they might often be too small to be useful. The dilemma one encounters with the foregoing three assumptions mentioned is similar to that faced in competing risks methodology, where considerable recent work has focused on identifiability and bounds for survival probabilities (see Tsiatis 1975 and Peterson 1976).

Viewing Wald's work on aircraft survivability in light of the state of the art at the time it was done, it seems to us to be a remarkable piece of work. While the field of statistics has grown considerably since the early 1940's, Wald's work on this problem is difficult to improve upon. Much of the work appears to be ad hoc—there are few allusions to modeling and no reference to classical statistical approaches or results. By the sheer power of his intuition, Wald was led to subtle structural relationships (e.g., Equations (3.3) and (3.24)), and was able to deal with both structural and inferential questions in a definitive way.

[*Received May 1981. Revised March 1983.*]

REFERENCES

AVRIEL, M. (1976), *Nonlinear Programming: Analysis and Methods*, Englewood Cliffs, N.J.: Prentice Hall.

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M., and BRUNK, H.D. (1972), *Statistical Inference Under Order Restrictions*, New York: John Wiley.

DEMPSTER, A.P., LAIRD, N.M., and RUBIN, D.B. (1977), "Maximum Likelihood From Incomplete Data Via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

MORSE, P.M. (1977), *In at the Beginnings: A Physicist's Life*, Cambridge, Mass.: MIT Press.

PETERSON, A.V. (1976), "Bounds for a Joint Distribution Function With Fixed Sub-Distribution Functions: Application to Competing Risks," *Proceedings of the National Academy of Sciences*, 73, 11–13.

TSIATIS, A. (1975), "A Nonidentifiability Aspect of the Problem of Competing Risks," *Proceedings of the National Academy of Sciences*, 72, 20–22.

WALD, A. (1973), *Sequential Analysis*, New York: Dover.

——— (1980), "A Method of Estimating Plane Vulnerability Based on Damage of Survivors," CRC 432, July 1980. (Copies can be obtained from the Document Center, Center for Naval Analyses, 2000 N. Beauregard St., Alexandria, VA 22311.)

WALLIS, W.A. (1980), "The Statistical Research Group, 1942–1945" (with discussion), *Journal of the American Statistical Association*, 75, 320–335.

Comment

JAMES O. BERGER*

The authors are to be congratulated on a fine paper. They have distilled the key ideas in Wald's work on aircraft survivability, and have successfully related the ideas to standard statistical methods. The bulk of this discussion will be concerned with this relationship of the work to standard statistical methods, particularly the use of statistical models to describe the situation. Some attention will also be given to decision-theoretic issues.

1. STATISTICAL MODELING

As indicated in the paper, the primary quantities studied can be considered

$$P_{i1} = P\,(i \text{ hits and survival}) = Q_i \cdot \lambda_i,$$

where

$$Q_i = P\,(\text{survival} \mid i \text{ hits}),$$

$$\lambda_i = P\,(i \text{ hits}),$$

and

$$P_0{}^* = P\,(\text{not surviving}) = 1 - \sum_{i=0}^{\infty} P_{i1}.$$

If the observations can be assumed to be independent, and out of a total of n missions the data are

X_{i1} = the number of aircraft that receive i hits and survive,

* James Berger is Professor of Statistics, Department of Statistics, Purdue University, West Lafayette, IN 47907. Research was supported by the National Science Foundation Grant MCS-8101670A1.

June 1984, Volume 79, Number 386
Applications Section

one could investigate statistically, via a likelihood ratio test or otherwise, the validity of the assumption $q_i \equiv q$. With the type of data available to Wald, such an option is not open because of the lack of identifiability of larger models. Wald cautioned his readers that the solution he provides should be used only "if it is known *a priori* that $q_1 = q_2 = \cdots = q_n$." How and whether such a priori knowledge could be garnered is open to debate. Wald does provide an option for those who are more conservative. The lower bounds for Q_i may be considered conservative estimates of survival probabilities, although they might often be too small to be useful. The dilemma one encounters with the foregoing three assumptions mentioned is similar to that faced in competing risks methodology, where considerable recent work has focused on identifiability and bounds for survival probabilities (see Tsiatis 1975 and Peterson 1976).

Viewing Wald's work on aircraft survivability in light of the state of the art at the time it was done, it seems to us to be a remarkable piece of work. While the field of statistics has grown considerably since the early 1940's, Wald's work on this problem is difficult to improve upon. Much of the work appears to be ad hoc—there are few allusions to modeling and no reference to classical statistical approaches or results. By the sheer power of his intuition, Wald was led to subtle structural relationships (e.g., Equations (3.3) and (3.24)), and was able to deal with both structural and inferential questions in a definitive way.

[*Received May 1981. Revised March 1983.*]

REFERENCES

AVRIEL, M. (1976), *Nonlinear Programming: Analysis and Methods*, Englewood Cliffs, N.J.: Prentice Hall.

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M., and BRUNK, H.D. (1972), *Statistical Inference Under Order Restrictions*, New York: John Wiley.

DEMPSTER, A.P., LAIRD, N.M., and RUBIN, D.B. (1977), "Maximum Likelihood From Incomplete Data Via the EM Algorithm" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 1–38.

MORSE, P.M. (1977), *In at the Beginnings: A Physicist's Life*, Cambridge, Mass.: MIT Press.

PETERSON, A.V. (1976), "Bounds for a Joint Distribution Function With Fixed Sub-Distribution Functions: Application to Competing Risks," *Proceedings of the National Academy of Sciences*, 73, 11–13.

TSIATIS, A. (1975), "A Nonidentifiability Aspect of the Problem of Competing Risks," *Proceedings of the National Academy of Sciences*, 72, 20–22.

WALD, A. (1973), *Sequential Analysis*, New York: Dover.

——— (1980), "A Method of Estimating Plane Vulnerability Based on Damage of Survivors," CRC 432, July 1980. (Copies can be obtained from the Document Center, Center for Naval Analyses, 2000 N. Beauregard St., Alexandria, VA 22311.)

WALLIS, W.A. (1980), "The Statistical Research Group, 1942–1945" (with discussion), *Journal of the American Statistical Association*, 75, 320–335.

Comment

JAMES O. BERGER*

The authors are to be congratulated on a fine paper. They have distilled the key ideas in Wald's work on aircraft survivability, and have successfully related the ideas to standard statistical methods. The bulk of this discussion will be concerned with this relationship of the work to standard statistical methods, particularly the use of statistical models to describe the situation. Some attention will also be given to decision-theoretic issues.

1. STATISTICAL MODELING

As indicated in the paper, the primary quantities studied can be considered

$$P_{i1} = P\,(i \text{ hits and survival}) = Q_i \cdot \lambda_i,$$

where

$$Q_i = P\,(\text{survival} \mid i \text{ hits}),$$

$$\lambda_i = P\,(i \text{ hits}),$$

and

$$P_0^* = P\,(\text{not surviving}) = 1 - \sum_{i=0}^{\infty} P_{i1}.$$

If the observations can be assumed to be independent, and out of a total of n missions the data are

X_{i1} = the number of aircraft that receive i hits and survive,

* James Berger is Professor of Statistics, Department of Statistics, Purdue University, West Lafayette, IN 47907. Research was supported by the National Science Foundation Grant MCS-8101670A1.

June 1984, Volume 79, Number 386
Applications Section

$$X_0^* = n - \sum_{i=0}^{\infty} X_{i1} = \text{the number that do not survive},$$

then the likelihood function for $\mathbf{P} = (P_0^*, P_{11}, P_{21}, \ldots)$ is proportional to

$$L(\mathbf{P}) = \left(\prod_{i=0}^{\infty} P_{i1}{}^{X_{i1}}\right) (P_0^*)^{X_0^*}$$

$$= \left[\prod_{i=0}^{\infty} (Q_i \cdot \lambda_i)^{X_{i1}}\right]\left[1 - \sum_{i=0}^{\infty} Q_i \cdot \lambda_i\right]^{X_0^*}. \quad (1)$$

In this framework, which is more or less that given in Section 3 of Mangel and Samaniego, Wald's *model* can be described by the following assumptions:

(i) $Q_i = q^i$ (i.e., iid survival of each hit);
(ii) $P_{i1} = 0$ for $i \geq 6$ (or, more generally, for i for which $X_{i1} = 0$).

We will return to the crucial assumption (i) later, but for now will accept it. Assumption (ii) leaves an obvious uncomfortable feeling, but probably makes no great difference for the type of data expected. A third assumption, actually a lack of an assumption, is also a possible cause for concern: Wald effectively leaves the λ_i (the probability of i hits) completely unrestricted, whereas it would seem more natural to restrict the parameter space to consist only of decreasing λ_i. (Actually, the λ_i are never even mentioned in Wald's work, an omission of some concern, as we shall see.)

As mentioned in the paper, Wald's analysis effectively corresponds to a maximum likelihood analysis using (1) and assumptions (i) and (ii). The results of this analysis for the given data are $\hat{q} = .851$ and $\hat{\lambda}_i = X_{i1}/[400\,(.851)^i]$. The values of the $\hat{\lambda}_i$ for the data are given in Table 1, and indeed they are not decreasing ($\hat{\lambda}_5 > \hat{\lambda}_4$). The possible difference here seems minor but, as a theoretical point, it seems desirable to ensure monotonicity of the λ_i in the analysis. (Perhaps the most straightforward way of incorporating monotonicity is simply to put the (noninformative) uniform prior distribution on

$$\Lambda = [(\lambda_0, \ldots, \lambda_5)\colon \textstyle\sum \lambda_i = 1, \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_5\},$$

a uniform prior on q (in [0,1]), and calculate the posterior means, providing the numerical integration problem is feasible.)

The most significant question that can be raised concerning Wald's analysis is that of overparameterization. The parameters are $(q, \lambda_0, \ldots, \lambda_5)$, seven parameters for the seven data values $(X_0^*, X_{01}, \ldots, X_{51})$. Wald attempts a model robustness study by finding lower and upper bounds for the P_{i1} (actually, for the Q_i), but these bounds are too disparate to be of much use (more on this in Section 3). The best way to investigate model robustness is usually just to try other possible models. What follows is a minimally parameterized model, which is actually the model we produced when challenged in the paper at the end of the Section 1.2 to analyze the data before reading further. (For fear of overparameterization, it is often helpful to start out by trying very small models.)

Consider the following assumptions:

(i) $Q_i = q^i$;

(ii) $\lambda_i = (1 - \lambda_0)\, \gamma^i e^{-\gamma}/[(1 - e^{-\gamma})\, i!]$ for $i \geq 1$.

Note that this is a three-parameter model, the parameters being $0 \leq q \leq 1$, $0 < \lambda_0 < 1$, and $\gamma > 0$. Our thoughts in choosing this model were (a) independence of effect of hits is a reasonable starting point, and (b) the number of hits might be approximately Poisson, except that some planes may never come under effective fire (for a variety of reasons), so that extra mass at zero hits is to be anticipated. Thus λ_0 was left unrestricted, while the remaining λ_i were given the truncated Poisson distribution. Of course, these assumptions can also be criticized, but they seemed to be a plausible starting point. Note that these assumptions bypass the need to make Wald's assumption (ii), and also will automatically result in decreasing λ_i (except possibly for λ_0, which seemed so likely to be large that monotonization would probably be unnecessary).

Using the fact that

$$\sum_{i=1}^{\infty} q^i \gamma^i/i! = e^{q\gamma} - 1,$$

the likelihood function (1) can be written (under our assumptions and after some algebra) as

$$L(q, \lambda_0, \gamma) = \lambda_0{}^{X_{01}}(1 - \lambda_0)^{n - X_{01}}\, (e^{\gamma} - 1)^{(X_{01} - n)} \times (q\gamma)^{\Sigma i X_{i1}}\, (e^{\gamma} - e^{q\gamma})^{(n - \Sigma X_{i1})}.$$

A routine maximum likelihood analysis for the given data yields $\hat{\lambda}_0 = .8$, $\hat{q} = .85$, and $\hat{\gamma} = 1.38$. How well this model fits the data can be seen in Table 1, which presents the estimated λ_i under this model, namely $\lambda_0^* = .8$ and

$$\lambda_i^* = (1 - \lambda_0^*)\, \hat{\gamma}^i e^{-\hat{\gamma}}/[(1 - e^{-\hat{\gamma}})\, i!], \quad i \geq 1,$$

along with the expected observations,

$$\hat{X}_{i1} = n \cdot \hat{P}_{i1} = n \cdot \hat{q}^i \cdot \lambda_i^*,$$

and the actual observations, X_{i1}. For comparison purposes, the unmodeled estimates $\hat{\lambda}_i$ for the λ_i are also given.

The low-parameter model seems to fit the data extremely well. Of course, one would expect to be able to

Table 1. Model Fit

i	λ_i^*	$\hat{\lambda}_i$	$\hat{X}_{i1}$	X_{i1}
0	.8000	.8000	320.0	320
1	.0928	.0940	31.5	32
2	.0640	.0690	18.5	20
3	.0295	.0162	7.2	4
4	.0102	.0095	2.1	2
5	.0028	.0112	0.5	2

fit seven decreasing data points well with some three-parameter model, but not necessarily this well and not necessarily with a model incorporating separate and very specialized structures for the Q_i and the λ_i. In any case, the main feature of interest here is that the answers obtained with this plausible three-parameter model are virtually identical to those of Wald's analysis (especially the $\hat{q}$), so that one can feel somewhat confident about the model robustness of the answers.

Before moving on, it is worthwhile commenting that, instead of the maximum likelihood analysis, a noninformative prior Bayesian analysis could have been performed, using (say) a constant (generalized) prior on the set

$$\Omega = \{(q, \lambda_0, \gamma): \quad 0 \leq q \leq 1, \quad \lambda_0 \geq \lambda_1, \quad \gamma > 0\}.$$

The advantages of this would be (a) the constraint $\lambda_0 \geq \lambda_1$ is automatically built in; (b) one does not have to worry about having found only local maxima of the likelihood function; and (c) with essentially no extra effort, the posterior variances can be found, yielding good small-sample variance estimates (an attractive alternative to the classical need to resort to large-sample theory).

2. ANALYSIS OF VULNERABILITY AREAS

It is in this aspect of the problem that statistical modeling can reap greater rewards than Wald's approach. Wald needed to assume that the effects of hits on a given area of the aircraft were independent (an assumption that seemed to work reasonably well for the entire aircraft), but this is unlikely to be true for certain vulnerable areas of the aircraft. One obvious example is the important engine area: A multi-engine aircraft might well be able to fly with one engine out, so that the effect of the first hit to the engine area would be inconsequential, while a second hit (to a different engine) could be fatal. It is not hard to think up appropriate models for this situation, and no identifiability problems arise as long as one also makes some effort to model the probability of i hits to a given area (combining, say, the ideas discussed earlier about modeling λ_i with Wald's ideas concerning the probability that a single hit strikes a given area).

3. LOWER BOUNDS ON SURVIVABILITY

A large portion of Wald's analysis is concerned with obtaining lower bounds, Q_i^*, on Q_i, the probability of surviving i hits. One possible use of this would be to allow the aircraft commander to abort a mission if the risk of subsequent hits is too high, but common sense would argue that the relevant factor in such a decision is not how many hits have been sustained (which may even be hard to determine during combat), but rather the amount of actual damage (say, fuel lost or engines destroyed) that can be determined. Data allowing analysis of such occurrences would be hard to come by, and any such analysis would almost certainly involve detailed knowledge about the workings of the aircraft.

A second possible use of the Q_i^* would be in bounding the overall probability of mission survival, presumably for logistic purposes. Clearly

$$\begin{aligned} \Psi &= P\text{ (survival)} \\ &= \sum_{i=0}^{\infty} Q_i \cdot \lambda_i \\ &\geq \sum_{i=0}^{\infty} Q_i^* \cdot \lambda_i. \end{aligned}$$

The difficulty with this use of the Q_i^* is that Wald determined Q_i^* as $Q_i^* = \min_{\mathcal{P}} Q_i$, where $\mathcal{P}$ is the set of probability structures such that $P_0^*, P_{01}, \ldots, P_{51}$ are equal to the sample proportions. Besides the lack of attention to the effect of sampling error on the analysis, there is the more basic problem that each Q_i is minimized separately over $\mathcal{P}$, and each minimum is attained at a *different* probability structure. Thus

$$\min_{\mathcal{P}} \Psi > \sum_{i=0}^{\infty} Q_i^* \cdot \lambda_i,$$

so that one can get a better lower bound by simply minimizing Ψ directly over $\mathcal{P}$. Of course, this will be computationally more difficult, which could well explain Wald's use of the Q_i^*, but today the additional computation would pose no serious problem.

As a final point, the use of lower bounds at all is probably unwise. Providing one can arrive at model-robust estimates of survivability, use of the estimates discussed in the previous paragraph will generally prove more valuable than use of lower bounds.

4. CONCLUSIONS

All nitpicking aside, the authors seem correct in their conclusion that the answers Wald obtained could not be greatly improved upon today. It can be argued, however, that the methodology employed by Wald was much more difficult and far less flexible than standard methodology involving statistical modeling. Of course, Wald was working under computational limitations (although use of simple statistical models and maximum likelihood methods would not necessarily have been harder computationally), and could perhaps have been writing for a special (nonstatistical) audience. Whatever the reasons for his approach, we can admire his ingenuity while being thankful for the availability of more powerful methods today.

Abraham Wald's Work on Aircraft Survivability: Rejoinder
Author(s): Marc Mangel and Francisco J. Samaniego
Source: *Journal of the American Statistical Association*, Vol. 79, No. 386 (Jun., 1984), pp. 270-271
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: http://www.jstor.org/stable/2288259
Accessed: 13-05-2017 20:37 UTC

Rejoinder

MARC MANGEL and FRANCISCO J. SAMANIEGO

1. INTRODUCTION

In this rejoinder we reply to the published remarks of Berger, respond to questions and comments that were raised at the American Statistical Association annual meeting in Toronto in August, 1983, and comment briefly on our recently completed Monte Carlo study on the robustness of Wald's methods.

2. REMARKS ON BERGER'S DISCUSSION

We thank Berger for his thoughtful and thought-provoking commentary on Wald's paper and ours. We are in general agreement with Berger on the main issues he has raised: (a) careful modeling can produce an excellent fit of Wald's data, and the related statistical computations are not that imposing; (b) some of Wald's assumptions are more troublesome than others; and (c) the lower bounds produced by Wald are mathematically interesting but of limited use in decision making. In spite of the consonance of our views with Berger's, there are one or two points on which we differ.

In our Section 3, we described Wald's first data set as an incomplete sample from a multinomial distribution. Berger criticized Wald's assumption that the probability of receiving more than five hits is zero. Actually, the assumption is inconsequential in a multinomial model, since every cell probability associated with an empty cell would be estimated as zero. Thus, Wald's estimator of the parameter q surfaces as the MLE with or without Wald's assumption.

Berger's three-parameter model for Wald's first data set is intriguing. We also tinkered with the Poisson model a bit, but found the fit unacceptable. Berger's idea and rationale for separating the events {0 hits} and {at least one hit} are appealing; it is the kind of idea that seems obvious as soon as it is mentioned, but it is to Berger's credit that he thought of it. Berger mistakenly claims that his model yields decreasing probabilities for 1, 2, 3, . . . hits. Actually, the positive Poisson model with parameter γ has mode $M = \max([\gamma], 1)$, where $[\cdot]$ is the greatest integer function. Thus, these probabilities increase up to M and decrease thereafter. With Wald's data, γ is estimated to be 1.38, so that $\hat{\lambda}_1 > \hat{\lambda}_2 > \hat{\lambda}_3 > \hat{\lambda}_4 > \hat{\lambda}_5$ in this particular application. However, Berger's model does not guarantee this monotonicity. Furthermore, although the Bayesian approach that Berger proposes in order to ensure the inequality $\hat{\lambda}_0 > \hat{\lambda}_1$ can be expanded to cover $\lambda_i > \gamma_{i+1}$ for all i, one should not underestimate the difficulties involved in implementing such an approach in a reasonable manner.

Having pointed out the lack of guaranteed monotonicity of the λ_i's, we hasten to add that, in our view, Berger's model nonetheless has substantial merit. Consider the proposition that $\lambda_1 > \lambda_2$, that is, that an aircraft is more likely to receive one hit than it is to receive two hits. It seems to us that this proposition is not an inviolable imperative. Indeed, the expected number of hits depends quite crucially on the density of fire. Suppose all 400 planes in Wald's first problem were sent on a mission in which intense fire was anticipated. It might well be true that virtually no aircraft would receive only one hit. In fact, it might be that aircraft would be more likely to receive 10 or 12 hits than only one. Berger's model will accommodate such situations, and it should be useful in problems in which the number of hits (to aircraft receiving at least one hit) is expected to have a unimodal distribution. It is interesting that data analysis with the three-parameter model yields the same estimate of q that Wald obtained, which imparts a certain model robustness to Wald's results. One could also interpret this coincidence as speaking to the model robustness of the approach Berger has taken. We are in agreement with the limitations of Wald's results, as discussed by Berger in his Sections 2 and 3.

Motivated in part by Berger's comments on robustness, we conducted our own study on the robustness of Wald's methods. Although the complete details are presented elsewhere (Mangel and Samaniego 1984), we wish to describe our results briefly. We studied two questions: (a) If the assumption that $q_j \equiv q$ for all j is violated, how badly does one do in estimating the p_{i2} using Wald's method? and (b) In the case of unequal q_j, what are the behavior and proper interpretation of Wald's estimator $\hat{q}$? To answer these questions, we carried out a Monte Carlo study in which data in (4.1) were repeatedly generated using a multinomial experiment with parameters $\{p_{ij}\}$ chosen so that the q_j were unequal but had the average $\bar{q} = .851$, as in Wald's data. Our base case involved equal q_j. We measured departure from the true probabilities p_{i2} via a χ^2-like statistic. We found that Wald's model worked very well in a fairly generous neighborhood of the central value $q = .851$, and that the fit was a monotonic function of the dispersion in the set $\{q_1, \ldots, q_5\}$. We also discovered that Wald's estimator $\hat{q}$ is an excellent estimator of the average $\bar{q}$, regardless of the dispersion.

3. COMMENTS AND QUESTIONS RAISED IN TORONTO

A discussant took exception to Wald's derivations and proposed the following alternative analysis. Retaining the

© Journal of the American Statistical Association
June 1984, Volume 79, Number 386
Applications Section

notation of Section 4 of our article, let

$$p_{j1} = P\{\text{receive exactly } j \text{ hits and survive}\}$$

$$p_{j2} = P\{\text{receive exactly } j \text{ hits and go down}\}.$$

It follows that

$$1 - p_{01} = \sum_{j=1}^{\infty} (p_{j1} + p_{j2}). \qquad \text{(R.1)}$$

The following modeling assumption was then introduced (apparently after Wald):

$$p_{j2}/p_{j1} = (1 - q)/q, \quad j = 1, 2, \ldots . \qquad \text{(R.2)}$$

Using (R.2) in (R.1) yields

$$1 - p_{01} = \sum_{j=1}^{\infty} p_{j1}\left(1 + \frac{p_{j2}}{p_{j1}}\right)$$

$$= \frac{1}{q} \sum_{j=1}^{\infty} p_{j1}. \qquad \text{(R.3)}$$

Thus

$$q = \frac{1}{1 - p_{01}} \sum_{j=1}^{\infty} p_{j1}, \qquad \text{(R.4)}$$

leading to the estimator

$$\hat{q} = \frac{1}{1 - a_0} \sum_{j=1}^{\infty} a_j \qquad \text{(R.5)}$$

for q. For Wald's data, one obtains $\hat{q} = .75$, which differs from the estimate of .851 obtained by Wald. Further discussion failed to shed any light on the comparative merits of the two estimators.

The confusion during the discussion at Toronto was due in part to blind acceptance of the faulty premise that the two estimators were estimating the same parameter. The proper resolution of this apparent anomaly is that these estimators are not competing against each other, but instead are valid estimators of parameters in different models. Modeling assumption (R.2) is equivalent to

$$p_{j1}/(p_{j1} + p_{j2}) = q, \quad j = 1, 2, \ldots, n, \qquad \text{(R.6)}$$

which differs from the modeling assumption

$$p_{j1}/(p_{j1} + p_{j2}) = q^j, \quad j = 1, 2, \ldots, n \qquad \text{(R.7)}$$

made by Wald. Indeed, if $f_1, \ldots, f_n$ are continuous, increasing functions mapping (0, 1) onto itself, then the modeling assumption

$$p_{j1}/(p_{j1} + p_{j2}) = f_j(q), \quad j = 1, 2, \ldots, n \qquad \text{(R.8)}$$

for the multinomial data in (4.1) gives rise to a unique MLE that can be obtained as the solution of the equation

$$\sum_{j=1}^{n} \frac{a_j}{f_j(q)} = 1 - a_0. \qquad \text{(R.9)}$$

Each such model has a parameter q, but the estimator of q in one model has no meaning as an estimator of q in another model.

It remains to comment on the modeling assumptions (R.6) and (R.7). Equations (R.6) constitute the assumption that the chance of surviving another hit, given survival thus far, is always the same. On the other hand, equations (R.7) assert that the conditional probability of surviving another hit, given survival thus far, depends on the number of hits sustained thus far. Wald's general model, with

$$\frac{p_{j1}}{p_{j1} + p_{j2}} = \prod_{i=1}^{j} q_j, \quad j = 1, \ldots, n, \qquad \text{(R.10)}$$

stipulates that these conditional probabilities are decreasing. Wald's assumption (R.7) asserts that these probabilities decrease geometrically. It is thus clear that the choice we have discussed is between two models rather than between two estimators. Applications undoubtedly exist in which either one of these models is more appropriate than the other.

A number of people have asked whether Wald's work has actually been used. We do not know whether it was used during World War II, although it was produced early enough in the war to have been available. We do know that during the Vietnam War, analysts at the Operations Evaluation Group of the Center for Naval Analyses used Wald's techniques to study the survivability of the A-4 aircraft. Their analysis led to structural modifications that improved the A-4's survivability. Wald's methods were also used by analysts at Wright Patterson Air Force Base in studying ways of improving the B-52's survivability. Cunningham and Hynd (1946) also provided perspective on the use of statistical analysis during World War II.

One tactical use of this kind of work is the development of rules for exiting from combat. The most important case is the one in which different survival probabilities are estimated (that is, where the q_i are not constant). For example, consider the result presented in Table 1 of our article. The change in the exact value of the probability of surviving i hits as i increases from 1 to 2 is .130, from 2 to 3 is .204, and from 3 to 4 is .235. When confronted with such data, aviators could develop rules of thumb such as, "Stay in combat with up to three hits, but leave after the fourth." Similarly, having an estimate for the survival probabilities would provide the mission planner with one more piece of information that could be used to determine the number of aircraft to send into a particular combat mission.

One factor that Wald did not take into account, but that is quite important, is the crew of the aircraft. Studies done during World War II showed that the crew was an important consideration in determining survivability. For example, crews that had already survived three missions had a much higher probability of continued survival (Morse 1977 discusses this point in more detail).

REFERENCES

CUNNINGHAM, L.B.C., and HYND, W.R.B. (1946), "Random Processes of Air Warfare," *Journal of the Royal Statistical Society*, Ser. B, 8, 62–65.

MANGEL, M., and SAMANIEGO, F.J. (1984), "On the Robustness of Wald's Estimator of Aircraft Survivability," unpublished manuscript.

MORSE, P.M. (1977), *In at the Beginnings: A Physicist's Life*, Cambridge, Mass.: MIT Press.

CRC 432 / July 1980

A REPRINT OF "A METHOD OF ESTIMATING PLANE VULNERABILITY BASED ON DAMAGE OF SURVIVORS" BY ABRAHAM WALD

Abraham Wald

CENTER FOR NAVAL ANALYSES

CENTER FOR NAVAL ANALYSES
2000 North Beauregard Street, Alexandria, Virginia 22311 Telephone: (703) 998-3500

MEMORANDUM FOR DISTRIBUTION LIST

Subj: Center for Naval Analyses Research Contribution 432

Encl: (1) CRC 432, "A Reprint of 'A Method of Estimating Plane Vulnerability Based on Damage of Survivors' by Abraham Wald," July 1980

1. Enclosure (1) is forwarded as a matter of possible interest.

2. This Research Contribution contains a series of memoranda written by Abraham Wald of the Statistical Research Group at Columbia University during World War II. Unfortunately, this work was never published externally, although some copies of his original memoranda have been available and his methodology has been employed in the analysis of data from both the Korean and Vietnam Wars. It is published by CNA not only as a matter of historical interest but also because the methodology is still relevant.

3. Research Contributions are distributed for their potential value in other studies and analyses. They do not necessarily represent the opinion of the Department of the Navy.

PHIL E. DePOY
Director
Operations Evaluation Group

Subj; Center for Naval Analyses Research Contribution 432

DISTRIBUTION LIST

Department of the Navy
SNDL

A2A	CNR
B3	National Defense University (NDU)
B3	Armed Forces Staff College (AFSC)

SNDL Part II:

E3A	NRL
FF38	USNA, Nimitz Library
FF44	NAVWARCOL
FKA1A	COMNAVAIRSYSCOMHQ
FKR3C	NAVAIRTESTCEN
FT73	NAVPGSCOL

OpNav: Op-09BH, Op-96, Op-05, Op-50W

Other:
Defense Technical Information Center (DTIC) (12 Copies)
Department of the Air Force (SAMI)
Air University
Air Proving Ground Center, Eglin AFB
Institute for Defense Analysis (IDA)
Mathematics Research Center
The Mitre Corporation
The Rand Corporation
University of Rochester (10)
Navy Tactical Support Activity

CRC 432 / July 1980

A REPRINT OF "A METHOD OF ESTIMATING PLANE VULNERABILITY BASED ON DAMAGE OF SURVIVORS" BY ABRAHAM WALD

Abraham Wald

CNA Operations Evaluation Group

CENTER FOR NAVAL ANALYSES

2000 North Beauregard Street, Alexandria, Virginia 22311

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

FOREWORD

This CNA research contribution is composed of a series of memoranda prepared in 1943 by Abraham Wald of the Statistical Research Group (SRG), Columbia University, for the Applied Mathematics Panel (AMP), National Defense Research Committee. The memoranda present methods of estimating the vulnerability of various parts of an aircraft on the basis of damage observed on returning planes. Unfortunately, this work was never published externally, although some copies of the original memoranda have been available and the methodology has been employed in the analysis of data in both the Korean and Vietnam wars. It is published by CNA not only as a matter of historical interest but also because the methodology is still relevant.

The eight memoranda in the series were published separately, but actually represent parts I through VIII of a larger attempt to address plane vulnerability. The parts are kept separate here, and their original AMP and SRG memorandum numbers are given.

Only very minor format changes have been made to accommodate CNA style and to smooth the transition from one part (memorandum) to another. The substance and original wording, however, have been retained.

Copies of the memoranda were acquired through the National Archives in Washington, D.C.

THIS PAGE INTENTIONALLY LEFT BLANK

PART I

AN EQUATION SATISFIED BY THE PROBABILITIES THAT A PLANE WILL BE DOWNED BY i HITS[1]

INTRODUCTION

Denote by P_i (i = 1,2,..., ad inf.) the probability that a plane will be downed by i hits. Denote by p_i the conditional probability that a plane will be downed by the i-th hit knowing that the first i - 1 hits did not down the plane. Let $Q_i = 1 - P_i$ and $q_i = 1 - p_i$ (i = 1,2,..., ad inf.). It is clear that

$$Q_i = q_1 q_2 \cdots q_i \tag{1}$$

and

$$P_i = 1 - q_1 q_2 \cdots q_i \quad . \tag{2}$$

Suppose that p_i and P_i (i = 1,2,...) are unknown and our information consists only of the following data concerning planes participating in combat:

- The total number N of planes participating in combat.
- For any integer i (i = 0,1,2,...) the number A_i of planes that received exactly i hits but have not been downed, i.e., have returned from combat.

Denote the ratio $\frac{A_i}{N}$ by a_i (i = 0,1,2,...) and let L be the proportion of planes lost. Then we have

$$\sum_{i=0}^{\infty} a_i = 1 - L. \tag{3}$$

---

[1]This part of "A Method of Estimating Plane Vulnerability Based on Damage of Survivors" was published as SRG memo 85 and AMP memo 76.1.

The purpose of this memorandum is to draw inferences concerning the unknown probabilities p_i and P_i on the basis of the known quantities a_o, a_1, a_2,..., etc.

To simplify the discussion, we shall neglect sampling errors, i.e., we shall assume that N is infinity. Furthermore, we shall assume that

$$0 < p_i < 1 \quad (i = 1,2,..., \text{ad inf.}). \tag{4}$$

From equation 4 it follows that

$$0 < P_i < 1 \quad (i = 1,2,..., \text{ad inf.}). \tag{5}$$

We shall assume that there exists a non-negative integer n such that $a_n > 0$ but $a_i = 0$ for $i > n$.

We shall also assume that there exists a positive integer m such that the probability is zero that the number of hits received by a plane is greater than or equal to m. Let m' be the smallest integer with the property that the probability is zero that the number of hits received by a plane is greater than or equal to m'. Then the probability that the plane receives exactly m' - 1 hits is positive. We shall prove that m' = n + 1. Since $a_n > 0$, it is clear that m' must be greater than n. To show that m' cannot be greater than n + 1, let y be the proportion of planes that received exactly m' - 1 hits. Then y > 0 and $y(1 - p_{m'-1}) = a_{m'-1}$. Since y > 0 and $1 - p_{m'-1} > 0$, we have $a_{m'-1} > 0$. Since $a_i = 0$ for i > n, we see that $m' - 1 \leq n$, i.e., $m' \leq n + 1$. Hence, m' = n + 1 must hold.

Denote by x_i (i = 1,2,...) the ratio of the number of planes downed by the i-th hit to the total number of planes participating in combat. Since m' = n + 1, we obviously have $x_i = 0$ for i > n. It is clear that

$$\sum_{i=1}^{n} x_i = L = 1 - a_o - a_1 - \dots - a_n \quad . \tag{6}$$

CALCULATION OF x_i IN TERMS OF $a_o, a_1, \ldots, a_n, p_1, \ldots, p_n$

Since the proportion of planes that received at least one hit is equal to $1 - a_o$, we have

$$x_1 = p_1(1 - a_o) \quad . \tag{7}$$

The proportion of planes that received at least two hits and the first hit did not down the plane is obviously equal to $1 - a_o - a_1 - x_1$. Hence,

$$x_2 = p_2(1 - a_o - a_1 - x_1). \tag{8}$$

In general, we obtain

$$x_i = p_i(1 - a_o - a_1 - \ldots - a_{i-1} - x_1 - x_2 - \ldots - x_{i-1}) \quad (i = 2,3,\ldots,n) \tag{9}$$

Putting

$$c_i = 1 - a_o - a_1 - \ldots - a_{i-1} \; , \tag{10}$$

equation 9 can be written

$$x_i + p_i(x_1 + \ldots + x_{i-1}) = p_i c_i \quad (i = 2,3,\ldots,n). \tag{11}$$

Substituting $i - 1$ for i, we obtain from equation 11

$$x_{i-1} + p_{i-1}(x_1 + \ldots + x_{i-2}) - p_{i-1}c_{i-1} \quad (i = 3,4,\ldots,n). \tag{12}$$

Dividing by p_{i-1}, we obtain

$$\frac{x_{i-1}}{p_{i-1}} + (x_1 + \ldots + x_{i-2}) = c_{i-1} \quad (i = 3,4,\ldots,n). \tag{13}$$

Adding $x_{i-1}\left(1 - \frac{1}{p_{i-1}}\right) = \frac{-q_{i-1}}{p_{i-1}} x_{i-1}$ to both sides of equation 13, we obtain

$$x_1 + \ldots + x_{i-1} = c_{i-1} - \frac{q_{i-1}}{p_{i-1}} x_{i-1} \qquad (i = 3,4,\ldots,n+1). \tag{14}$$

From equations 11 and 14, we obtain

$$x_i + p_i\left(c_{i-1} - \frac{q_{i-1}}{p_{i-1}} x_{i-1}\right) = p_i c_i \ . \tag{15}$$

Hence,

$$x_i = p_i(c_i - c_{i-1}) + \frac{p_i q_{i-1}}{p_{i-1}} x_{i-1} \ (i = 3,4,\ldots,n). \tag{16}$$

Let

$$d_i = p_i(c_i - c_{i-1}) = -p_i a_{i-1} \qquad (i = 3,4,\ldots,n) \tag{17}$$

and

$$t_i = \frac{p_i q_{i-1}}{p_{i-1}} \qquad (i = 3,4,\ldots,n). \tag{18}$$

Then equation 16 can be written as

$$x_i = d_i + t_i x_{i-1} \qquad (i = 3,4,\ldots,n). \tag{19}$$

Denote $p_1(1 - a_o)$ by d_1, $-p_2 a_1$ by d_2, and $\frac{p_2 q_1}{p_1}$ by t_2; then we have

$$x_1 = d_1 \text{ and } x_2 = t_2 x_1 + d_2 \ . \tag{20}$$

From equations 19 and 20, we obtain

$$\begin{cases} x_1 = d_1 \\ x_i = \sum_{j=1}^{i-1} d_j t_{j+1} t_{j+2} \cdots t_i + d_i \end{cases} \qquad (i = 2,3,\ldots,n). \qquad (21)$$

EQUATION SATISFIED BY $q_1,\ldots,q_n$

To derive an equation satisfied by $q_1,\ldots,q_n$, we shall express $\sum_{i=1}^{n} x_i$ in terms of the quantities t_i and d_i $(i = 1,\ldots,n)$.

Substituting i for i - 1 in equation 14, we obtain

$$x_i = \sum_{j=1}^{i} x_j = c_i - \frac{q_i}{p_i} x_i = c_i - \frac{q_i}{p_i}\left[\sum_{j=1}^{i-1} (d_j t_{j+1} \cdots t_i) + d_i\right] . \qquad (22)$$

Hence, in particular

$$X_n = \sum_{j=1}^{n} x_j = c_n - \frac{q_n}{p_n}\left[\sum_{j=1}^{n-1} (d_j t_{j+1} \cdots t_n) + d_n\right] = L . \qquad (23)$$

Since $c_n - L = a_n$, and since $t_{j+1} \cdots t_n = \frac{p_n}{p_j} q_j \cdots q_{n-1}$, we obtain from equation 23

$$a_n - \left(\sum_{j=1}^{n-1} \frac{d_j}{p_j} q_j \cdots q_n\right) + q_n a_{n-1} = 0 . \qquad (24)$$

Dividing by $q_1 \cdots q_n$ and substituting $-p_j a_{j-1}$ for d_j, we obtain

$$\frac{a_n}{q_1 \cdots q_n} + \frac{a_{n-1}}{q_1 \cdots q_{n-1}} - \sum_{j=1}^{n-1} \frac{d_j}{p_j q_1 \cdots q_{j-1}}$$

$$= \frac{a_n}{q_1 \cdots q_n} + \frac{a_{n-1}}{q_1 \cdots q_{n-1}} \tag{25}$$

$$+ \sum_{j=2}^{n-1} \frac{a_{j-1}}{q_1 \cdots q_{j-1}} - \frac{d_1}{p_1}$$

$$= \sum_{j=1}^{n} \frac{a_j}{q_1 \cdots q_j} - (1 - a_o) = 0$$

or

$$\sum_{j=1}^{n} \frac{a_j}{q_1 \cdots q_j} = 1 - a_o \ . \tag{26}$$

If it is known a priori that $q_1 = \ldots = q_n$, then our problem is completely solved. The common value of $q_1, \ldots, q_n$ is the root (between 0 and 1) of the equation

$$\sum_{j=1}^{n} \frac{a_j}{q^j} = 1 - a_o \ .$$

It is easy to see that there exists exactly one root between zero and one. We can certainly assume that $q_1 \geq q_2 \geq \cdots \geq q_n$. We shall investigate the implications of these inequalities and equation 26 later.

ALTERNATIVE DERIVATION OF EQUATION 26

Let b_i be the hypothetical proportion of planes that would have been hit exactly i times if dummy bullets would have been used. Clearly $b_i \geq a_i$. Denote $b_i - a_i$ by y_i $(i = 0, 1, 2, \ldots, n)$. Of course, $b_o = a_o$, i.e., $y_o = 0$. We have $\sum_{j=0}^{n} b_i = 1$. Clearly

$$y_i = P_i b_i = P_i(a_i + y_1) \qquad (i = 1,2,\ldots,n). \tag{27}$$

Hence,

$$y_i = \frac{P_i}{Q_i} a_i = \frac{1 - q_1 \cdots q_i}{q_1 \cdots q_i} a_i = \frac{a_i}{q_1 \cdots q_i} - a_i . \tag{28}$$

Since $\sum_{i=1}^{n} y_i = L$, we obtain from equation 28

$$\sum_{i=1}^{n} \frac{a_i}{q_1 \cdots q_i} = L + \sum_{i=1}^{n} a_i = 1 - a_o . \tag{29}$$

This equation is the same as equation 26. This is a simpler derivation than the derivation of equation 26 given before. However, equations 21 and 22 (on which the derivation of equation 26 was based) will be needed later for other purposes.

As mentioned before, equation 29 leads to a solution of our problem if it is known that $q_1 = \ldots = q_n$. In the next memorandum (part II) we shall investigate the implications of equation 29 under the condition that $q_1 \geq q_2 \geq \ldots \geq q_n$.

NUMERICAL EXAMPLES

N is the number of planes participating in combat. $A_o, A_1, A_2, \ldots, A_n$ are the number returning with no hits, one hit, two hits, ..., n hits, respectively. Then

$$a_i = \frac{A_i}{N} \qquad (i = 0,1,2,\ldots,n)$$

i.e., a_i is the proportion of planes returning with i hits. The computations below were performed under the following two assumptions:

- The bombing mission is representative so that there is no sampling error.
- The probability that a plane will be shot down does not depend on the number of previous non-destructive hits.

Example 1: Let $N = 400$
and $A_o = 320$ then $a_o = .80$
$A_1 = 32$ $a_1 = .08$
$A_2 = 20$ $a_2 = .05$
$A_3 = 4$ $a_3 = .01$
$A_4 = 2$ $a_4 = .005$
$A_5 = 2$ $a_5 = .005$

We assume $q_1 = q_2 = \ldots = q_5 = q_i$, where q_i is the probability of a plane surviving the i-th hit, knowing that the first i - 1 hits did not down the plane.

Then equation 26,

$$\sum_{j=1}^{n} \frac{a_j}{q_1 \ldots q_j} = 1 - a_o ,$$

reduces to

$$\sum_{j=1}^{n} \frac{a_j}{q^j} = 1 - a_o \quad .$$

Substituting values of a_i

$$\frac{.08}{q} + \frac{.05}{q^2} + \frac{.01}{q^3} + \frac{.005}{q^4} + \frac{.005}{q^5} = .20$$

or

$$.200q^5 - .080q^4 - .050q^3 - .010q^2 - .005q - .005 = 0.$$

The Birge-Vieta method of finding roots described in Marchant Method No. 225 is used to solve this equation (table 1). We find $q = q_i = .851$, $p_i = .149$ where p_i is the probability of a plane being downed by the i-th hit, knowing that the first i - 1 hits did not down the plane.

x_i equals the ratio of the number of planes downed by the i-th hit to the total number of planes participating in combat. Using equation 9

$$x_i = p_i(1 - a_o - a_1 - \dots - a_{i-1} - x_1 - x_2 - \dots - x_{i-1})$$

$$(i = 2,3,\dots,n)$$

for n = 5, we obtain

$$x_1 = p_1(1 - a_o) = .030$$
$$x_2 = p_2(1 - a_o - a_1 - x_1) = .013$$
$$x_3 = p_3(1 - a_o - a_1 - a_2 - x_1 - x_2) = .004$$
$$x_4 = p_4(1 - a_o - a_1 - a_2 - a_3 - x_1 - x_2 - x_3) = .002$$
$$x_5 = p_5(1 - a_o - a_1 - a_2 - a_3 - a_4 - x_1 - x_2 - x_3 - x_4) = .001$$

Example 2: Let $a_o = .3$, $a_1 = .2$, $a_2 = .1$, $a_3 = .1$, $a_4 = .05$, and $a_5 = .05$. Then the following results are obtained: $q = .87$, $p = 1 - q = .13$, $x_1 = .09$, $x_2 = .05$, $x_3 = .03$, $x_4 = .02$, and $x_5 = .01$.

The value of q in the second example is nearly equal to the value in the first example in spite of the fact that the values a_i $(i = 0,1,\dots,5)$ differ considerably. The difference in the values a_i in these two examples is mainly due to the fact that the probability that a plane will receive a hit is much smaller in the first example than in the second example. The probability that a plane will receive a hit has, of course, no relation to the probability that a plane will be downed if it receives a hit.

TABLE 1

1. Assume $q = 1 = y_1$

.200	-.080	-.050	-.010	-.005	-.005
	+.200	+.120	+.070	+.060	+.055
.200	+.120	+.070	+.060	+.055	+.050 = A_o
	+.200	+.320	+.390	+.450	
.200	+.320	+.390	+.450	+.505 = A_1	

$$y_2 = y_1 - \frac{A_o}{A_1} = 1 - .0990 = .9010$$

2. Assume $q = .9010 = y_2$

.2000	-.0800	-.0500	-.0100	-.0050	-.0050
	+.1802	+.0903	+.0363	+.0237	+.0168
.2000	+.1002	+.0403	+.0263	+.0187	+.0118 = B_o
	+.1802	+.2526	+.2639	+.2615	
.2000	+.2804	+.2929	+.2902	+.2802 = B_1	

$$y_3 = y_2 - \frac{B_o}{B_1} = .9010 - .042113 = .858887$$

3. Assume $q = .858887 = y_3$

.200000	-.080000	-.050000	-.010000	-.005000	-.005000
	+.171777	+.078826	+.024758	+.012675	+.006592
.200000	+.091777	+.028826	+.014758	+.007675	+.001592 = C_o
	+.171777	+.226363	+.219179	+.200925	
.200000	+.263554	+.255189	+.233937	+.208600 = C_1	

$$y_4 = y_3 - \frac{C_o}{C_1} = .858887 - .007632 = .851255$$

4. Assume $q = .851255 = y_4$

.2000000	+.080000	-.050000	-.010000	-.005000	-.005000
	+.170251	+.076827	+.022837	+.010928	+.005046
.2000000	+.090251	+.026827	+.012837	+.005928	+.000046 = D_o
	+.170251	+.221754	+.211606	+.191058	
.2000000	+.260502	+.248581	+.224443	+.196986 = D_1	

$$y_5 = y_4 - \frac{D_o}{D_1} = .851255 - .000234 = .851021$$

PART II

MAXIMUM VALUE OF THE PROBABILITY THAT A PLANE WILL BE DOWNED BY A GIVEN NUMBER OF HITS[1]

The symbols defined and the results obtained in part I will be used here without further explanation. The purpose of this memorandum is to derive the least upper bound of $X_i = \sum_{j=1}^{i} x_j$ and that of P_i $(i = 1,\ldots,n)$ under the restriction that $q_1 \geq q_2 \geq ,\ldots, \geq q_n$.

First, we shall show that X_i is a strictly increasing function of p_j for $j \leq i$. Let us replace p_j by $p_j + \Delta$ $(\Delta > 0)$ and let us study the effect of this change on $x_1,\ldots,x_i$. Denote the changes in $x_1,\ldots,x_i$ by $\Delta_1,\ldots,\Delta_i$, respectively. Clearly, $\Delta_1 = \ldots = \Delta_{j-1} = 0$. It follows easily from equation 9 that

$\Delta_j > 0$ and

$$\Delta_{j+1} = -p_{j+1}\,\Delta_j \quad .$$

Hence,

$$\Delta_j + \Delta_{j+1} = (1 - p_{j+1})\,\Delta_j > 0.$$

Similarly, we obtain from equation 9

$$\Delta_{j+2} = -p_{j+2}(\Delta_j + \Delta_{j+1}) = -p_{j+2}(1 - p_{j+1})\,\Delta_j \quad .$$

Hence,

$$\Delta_j + \Delta_{j+1} + \Delta_{j+2} = (1 - p_{j+2})\,(1 - p_{j+1})\,\Delta_j > 0.$$

In general

$$\Delta_j + \Delta_{j+1} + \ldots + \Delta_{j+k} = (1 - p_{j+1}) \ldots (1 - p_{j+k})\,\Delta_j > 0$$

$$(k = 1,\ldots,i-j)$$

Hence, we have proved that X_i is a strictly increasing function of p_j $(j = 1,\ldots,i)$.

[1]This part of "A Method of Estimating Plane Vulnerability Based on Damage of Survivors" was published as SRG memo 87 and AMP memo 76.2.

On the basis of the inequalities $p_i \geq p_{i-1}$, we shall derive the least upper bound of X_i. For the purpose of this derivation we shall admit 0 and 1 as possible values of p_i $(i = 1,\ldots,n)$, thus making the domain of all possible points $(p_1,\ldots,p_n)$ to be a closed and bounded subset of the n-dimensional Cartesian space. Since X_i is a continuous function of the probabilities p_1, p_2, etc. (X is a polynomial in $p_1,\ldots,p_i$), the maximum of X_i exists and coincides, of course, with the least upper bound. Hence, our problem is to determine the maximum of X_i.

First, we show that the value of X_i is below the maximum if $p_n > p_i$. Assume that $p_n > p_i$ and let k be the smallest positive integer for which $p_k > p_i$. Obviously $k > i$. Let $p'_j = p_j(1 + \varepsilon)$ for $j = 1,\ldots,k-1$, and $p'_j = p_j(1 - \eta)$ for $j = k,k+1,\ldots,n$, where $\varepsilon > 0$ and η is a function $\eta(\varepsilon)$ of ε determined so that $\sum_{j=1}^{n} x'_j = L$ (x'_j is the proportion of planes that would have been brought down with the j-th hit if $p'_1,\ldots,p'_n$ were the true probabilities). Since X_r $(r = 1,\ldots,n)$ is a strictly monotonic function of $p_1,\ldots,p_r$, it is clear that for sufficiently small such a function $\eta(\varepsilon)$ exists. It is also clear that for sufficiently small ε the condition $p'_1 \leq p'_2 \leq \ldots \leq p'_n$ is fulfilled. Since $p'_j > p_j$ $(j = 1,\ldots,i)$, we see that $X'_1 > X_i$ (X_i does not depend on p'_r for $r > i$). Hence, we have proved that if $p_1,\ldots,p_n$ is a point at which X_i becomes a maximum, we must have $p_i = p_{i+1} = \ldots = p_n$.

Now we shall show that if X_i is a maximum then $p_1 = p_2 = \ldots = p_i$. For this purpose assume that $p_i > p_1$ and we shall derive a contradiction. Let j be the greatest integer for which $p_j = p_1$. Since $p_i > p_1$, we must have $j < i$. Let $p'_r = p_r(1 + \varepsilon)$ for $r = 1,\ldots,j$ and $p'_r = p_r(1 - \eta)$ for $r = j+1,\ldots,i$, where $\varepsilon > 0$ and η is determined so that $\sum_{k=1}^{i} x'_k = \sum_{k=1}^{i} x_k$. Then for the probabilities $p'_1,\ldots,p'_i$, $p_{i+1},\ldots,p_n$ the proportion of lost planes is not changed, i.e., it is equal to L. Now let $p'_r = p'_i$ for $r > i$. Then the proportion L' of lost planes corresponding to $p'_1,\ldots,p'_n$ is less than L. Hence, there exists a positive Δ so that the proportion L" of lost planes corresponding to the probabilities $p''_r = p'_r\,(1 + \Delta)$ is equal to L. But, since $p''_r > p'_r$ $(r = 1,\ldots,i)$ we must have $\sum_{j=1}^{i} x''_j > \sum_{j=1}^{i} x'_j = \sum_{j=1}^{i} x_j$. Hence, we arrived at a contradiction and our statement that $p_1 = p_2 = \ldots = p_i$ is proved. Thus, we see that the maximum of X_i is reached when $p_1 = p_2 = \ldots = p_n$.

LEAST UPPER BOUND OF P_i

Now we shall calculate the least upper bound of P_i. Admitting the values 0 and 1 for p_j, the maximum of P_i exists and is equal to the least upper bound of P_i. Since $P_i = 1 - q_1 \ldots q_i$, maximizing P_i is the same as minimizing $q_1 \ldots q_i$. We know that $q_1,\ldots,q_n$ are subject to the restriction

$$\sum_{j=1}^{n} \frac{a_j}{q_1 \ldots q_j} = 1 - a_o \quad . \tag{30}$$

Let $q_1^o,\ldots,q_n^o$ be a set of values of $q_1,\ldots,q_n$ (satisfying equation 30) for which $q_1 \ldots q_j$ becomes a minimum. First, we show that $q_i^o = q_{i+1}^o = \ldots = q_n^o$. Suppose that $q_n^o < q_i^o$. Consider the set of probabilities $q'_r = q_r^o$ for $r \leq 1$ and $q'_r = q_i^o$ for $r > i$. Then

$$\sum_{j=1}^{n} \frac{a_j}{q'_1 \ldots q'_j} < 1 - a_o \quad .$$

Hence, there exists a positive factor $\lambda < 1$ so that

$$\sum_{j=1}^{n} \frac{a_j}{q_1'' \cdots q_j''} = 1 - a_o ,$$

where $q_i'' = \lambda q_i'$ $(i = 1,\ldots,n)$. Then

$$q_1''q_2'' \cdots q_1'' < q_1^o q_2^o \cdots q_i^o$$

in contradiction to our assumption that $q_1^o \ldots q_i^o$ is a minimum. Hence, we have proved that $q_i^o = \ldots = q_n^o$.

Now we show that there exists at most one value j such that $1 > q_j^o > q_i^o$. Suppose there are two integers j and k such that $1 > q_j^o \geq q_k^o > q_i^o$. Let j' be the smallest integer for which $q_{j'}^o = q_j^o$ and let k' be the largest integer for which $q_{k'}^o = q_k^o$. Let $\bar{q}_{j'} = (1 + \varepsilon)\, q_{j'}^o$, $\bar{q}_{k'} = \frac{1}{1 + \varepsilon} q_{k'}^o$ $(\varepsilon > 0)$, and $\bar{q}_r = q_r^o$ for $r \neq j', \neq k'$. Then

$$\bar{q}_1 \cdots \bar{q}_i = q_1^o \cdots q_i^o \quad \text{and} \quad \sum_{r=1}^{n} \frac{a_r}{\bar{q}_1 \cdots \bar{q}_r} < 1 - a_o .$$

Hence, there exists a positive factor $\lambda < 1$ such that

$$\sum_{r=1}^{n} \frac{a_r}{q_1^* \cdots q_r^*} = 1 - a_o ,$$

where $q_r^* = \lambda \bar{q}_r$. But $q_1^* \ldots q_i^* < \bar{q}_1 \ldots \bar{q}_i = q_1^o \ldots q_i^o$, which contradicts the assumption that $q_1^o \ldots q_i^o$ is a minimum. This proves our statement.

It follows from our results that the minimum of q_1 is the root of the equation

$$\sum_{r=1}^{n} \frac{a_r}{q^r} = 1 - a_o \quad . \tag{32}$$

Now we shall calculate the minimum of $q_1 q_2$. First, we know that $q_i = q_2$ ($i \geq 2$) if $q_1 q_2$ be a minimum. Hence, we have to minimize $q_1 q_2$ under the restriction

$$\frac{1}{q_1}\left(a_1 + \frac{a_2}{q_2} + \frac{a_3}{q_2^2} + \ldots + \frac{a_n}{q_2^{n-1}}\right) = 1 - a_o \quad . \tag{33}$$

Using the Lagrange multiplier method we obtain the equations

$$q_2 - \frac{\lambda}{q_1^2}\left(a_1 + \frac{a_2}{q_2} + \frac{a_3}{q_2^2} + \ldots + \frac{a_n}{q_2^{n-1}}\right) = 0 \tag{34}$$

(Lagrange multiplier = λ)

$$q_1 - \frac{\lambda}{q_1}\left(\frac{a_2}{q_2^2} + \frac{2a_3}{q_2^3} + \ldots + \frac{(n-1)a_n}{q_2^n}\right) = 0 \; . \tag{35}$$

Because of equation 33, we can write equation 34 as follows:

$$q_2 - \frac{\lambda}{q_1} (1 - a_o) = 0; \quad \lambda = \frac{q_1 q_2}{1 - a_o} \quad .$$

Substituting for λ in equation 35, we obtain

$$q_1 - \frac{1}{1 - a_o}\left(\frac{a_2}{q_2} + \frac{2a_3}{q_2^2} + \frac{3a_4}{q_2^3} + \ldots + \frac{(n-1)a_n}{q_2^{n-1}}\right) = 0 \tag{36}$$

or

$$q_1 = \frac{1}{1 - a_o} \left(\frac{a_2}{q_2} + \frac{2a_3}{q_2^2} + \ldots + \frac{(n - 1)a_n}{q_2^{n-1}} \right) . \qquad (37)$$

On the other hand, from equation 33 we obtain

$$q_1 = \frac{1}{1 - a_o} \left(a_1 + \frac{a_2}{q_2} + \frac{a_3}{q_2^2} + \ldots + \frac{a_n}{q_2^{n-1}} \right) . \qquad (38)$$

Equating the right-hand sides of equations 37 and 38, we obtain

$$\frac{a_3}{q_2^2} + \frac{2a_4}{q_2^3} + \frac{3a_5}{q_2^4} + \ldots + \frac{(n - 2)a_n}{q_2^{n-1}} - a_1 = 0. \qquad (39)$$

It is clear that equation 39 has exactly one positive root. The root is less than or equal to 1 if and only if

$$a_3 + 2a_4 + 3a_5 + \ldots + (n - 2)a_n \leq a_1 . \qquad (40)$$

Equations 38 and 39 have exactly one positive root in q_1 and q_2. We shall show that if the roots satisfy the inequalities $1 \geq q_1 \geq q_2$, then for these roots $q_1 q_2$ becomes a minimum. We can assume that $2 < n$, since the derivation of the minimum value of $q_1 \ldots q_n$ will be given later in this memorandum. It is clear that for any value $q_1 > \frac{a_1}{1 - a_o}$ equation 38 has exactly one positive root in q_2. Denote this root by $\phi(q_1)$. Hence, $\phi(q_1)$ is defined for all values $q_1 > \frac{a_1}{1 - a_o}$. It is easy to see that

$$\lim_{q_1 \to \frac{a_1}{1-a_0}} \phi(q_1) = +\infty$$

Hence (assuming $a_1 > 0$)

$$\lim_{q_1 \to \frac{a_1}{1-a_0}} \psi(q_1) = +\infty$$

where $\psi(q_1) = q_1 \ \phi(q_1)$.

It is clear that $\lim_{q_1 \to \infty} \phi(q_1) = 0$. Since $a_n > 0$, it follows from equation 38 that $q_1 [\phi(q_1)]^{n-1}$ has a positive lower bound when $q_1 \to \infty$. But then, since $n > 2$, $\lim_{q_1 \to \infty} q_1 \ \phi(q_1) = +\infty$. From the relations $\lim_{q_1 \to \frac{a_1}{1-a_0}} \psi(q_1) = \lim_{q_1 \to \infty} \psi(q_1) = +\infty$ it follows that the absolute minimum value of $\psi(q_1)$ is reached for some positive value q_1. Since equations 38 and 39 have exactly one positive root in q_1 and q_2, the absolute minimum value of $\psi(q_1)$ must be reached for this root. This proves our statement that if the roots of equations 38 and 39 satisfy the inequalities $1 \geq q_1 \geq q_2$, then for these roots $q_1 q_2$ becomes a minimum consistent with our restrictions on q_1 and q_2. If $1 \geq q_1 \geq q_2$ is not satisfied by the roots of equations 38 and 39, then q_1 is equal either to 1 or to q_2 and the minimum value of $q_1 q_2$ is either $\phi(1)$ or q^2, where q is the root of the equation

$$\sum_{r=1}^{n} \frac{a_r}{q^r} = 1 - a_o \quad .$$

Now we shall determine the minimum of $q_1 \ldots q_i$ $(2 < i < n)$. First, we determine the minimum M_{i1} of $q_1 \ldots q_i$ under the restriction that $q_2 = q_i$. Thus, we have to minimize $q_1 q_2^{i-1}$ under the restriction that

$$\frac{a_1}{q_1} + \frac{a_2}{q_1 q_2} + \frac{a_3}{q_1 q_2^2} + \ldots + \frac{a_n}{q_1 q_2^{n-1}} = 1 - a_o. \qquad (40a)$$

Using the Lagrange multiplier method, we obtain

$$q_2^{i-1} - \frac{\lambda}{q_1}\left(\frac{a_1}{q_1} + \ldots + \frac{a_n}{q_1 q_2^{n-1}}\right) = q_2^{i-1} - \frac{\lambda}{q_1}(1 - a_o) = 0; \qquad (41)$$

and

$$(i - 1)q_1 q_2^{i-2} - \frac{\lambda}{q_1}\left(\frac{a_2}{q_2^2} + \frac{2a_3}{q_2^3} + \ldots + \frac{(n-1)a_n}{q_2^n}\right) = 0 \ . \qquad (41a)$$

Substituting $\frac{q_1 q_2^{i-1}}{1 - a_o}$ for λ (the value of λ obtained from equation 41), we obtain

$$(i - 1)q_1 - \frac{1}{1 - a_o}\left(\frac{a_2}{q_2} + \frac{2a_3}{q_2^2} + \ldots + \frac{(n-1)a_n}{q_2^{n-1}}\right) = 0 \ . \qquad (42)$$

From equation 40a

$$(i - 1)q_1 - \frac{i - 1}{1 - a_o}\left(a_1 + \frac{a_2}{q_2} + \ldots + \frac{a_n}{q_2^{n-1}}\right) = 0 \ . \qquad (43)$$

From equations 42 and 43, we obtain

$$(i - 1)a_1 + \frac{(i - 2)a_2}{q_2} + \frac{(i - 3)a_3}{q_2^2} + \dots + \frac{(i - n)a_n}{q_2^{n-1}} = 0. \quad (44)$$

From Descartes' sign rule it follows that equation 44 has exactly one positive root.

Let $q_1 = q_1^o$ and $q_2 = q_2^o$ be the roots of the equations 43 and 44. If $1 \geq q_1^o \geq q_2^o$, then $M_{i1} = q_1^o(q_2^o)^{i-1}$. If $1 \geq q_1^o \geq q_2^o$ does not hold, then M_{i1} is either $(q')^i$ or $(q")^{i-1}$, where q' is the root of the equation

$$\sum_{j=1}^{n} \frac{a_j}{(q')^j} = 1 - a_o \quad (45)$$

and q" is the root of the equation

$$a_1 + \frac{a_2}{q"} + \frac{a_3}{(q")^2} + \dots + \frac{a_n}{(q")^{n-1}} = 1 - a_o . \quad (46)$$

Let M_{ir} $(r = 2,\dots,i-1)$ be the minimum of $q_1 \dots q_i$ under the restriction that $q_1 = \dots = q_{r-1} = 1$ and $q_{r+1} = q_i$. Then M_{ir} can be calculated in the same way as M_{i1}; we have merely to make the substitutions

$$\begin{aligned} n^* &= n - r + 1 \\ a_o^* &= a_o + a_1 + \dots + a_{r-1} \\ a_j^* &= a_{j+r-1} && (j = 1,\dots,n^*) \\ q_j^* &= q_{j+r-1} && (j = 1,\dots,n^*) \\ i^* &= i - r + 1, \end{aligned}$$

and we have to calculate the minimum of $q_1^* \dots q_{i*}^*$. Thus, we have to solve the equations corresponding to equations 43 and 44, i.e., the equations

$$(i^* - 1)q_1^* - \frac{i^* - 1}{1 - a_o^*}\left(a_1^* + \frac{a_2^*}{q_2^*} + \frac{a_3^*}{(q_2^*)^2} + \dots + \frac{a_{n^*}^*}{(q_2^*)^{n^*-1}}\right) = 0 \qquad (43^*)$$

and

$$(i^* - 1)a_1^* + \frac{(i^* - 2)a_2^*}{q_2^*} + \frac{(i^* - 3)a_3^*}{(q_2^*)^2} + \dots + \frac{(i^* - n^*)a_{n^*}^*}{(q_2^*)^{n^*-1}} = 0. \qquad (44^*)$$

Let $q_1^* = v_1$ and $q_2^* = v_2$ be the positive roots of the equations 43* and 44*. If $1 \geq v_1 \geq v_2$, then $M_{ir} = v_1 v_2^{i^*-1}$. If $1 \geq v_1 \geq v_2$ does not hold, then M_{ir} is equal to either $(v')^{i^*}$ or $(v'')^{i^*-1}$, where v' is the positive root of the equation

$$\sum_{j=1}^{n^*} \frac{a_j^*}{(v')^j} = 1 - a_o^* \qquad (45^*)$$

and v'' is the positive root of the equation

$$a_1^* + \frac{a_2^*}{v''} + \frac{a_3^*}{(v'')^2} + \dots + \frac{a_{n^*}^*}{(v'')^{n^*-1}} = 1 - a_o^* . \qquad (46^*)$$

The minimum M_i of $q_1 \dots q_i$ $(i = 2,3,\dots,n-1)$ is equal to the smallest of the $i - 1$ values $M_{i1},\dots,M_{i,i-1}$.

Now we shall determine the minimum of $q_1 \dots q_n$. We show that the minimum is reached when $q_1 = \dots = q_{n-1} = 1$. Suppose that this is not true and we shall derive a contradiction. Let j be the smallest integer for which $q_j < 1$ $(j < n)$. Let $\bar{q}_j = (1 + \varepsilon)q_j$ $(\varepsilon > 0)$, $\bar{q}_n = \frac{q_n}{1 + \varepsilon}$, and $\bar{q}_r = q_r$ for all $r \neq j, \neq n$.

Then $\bar{q}_1 \dots \bar{q}_n \dots = q_1 \dots q_n$ and

$$\sum_{r=1}^{n} \frac{a_r}{\bar{q}_1 \cdots \bar{q}_r} < 1 - a_o \quad .$$

Hence, there exists a positive $\lambda < 1$ such that

$$\sum_{r=1}^{n} \frac{a_r}{q_1^* \cdots q_r^*} = 1 - a_o ,$$

where

$$q_r^* = \lambda \bar{q}_r \quad .$$

But then $q_1^* \ldots q_n^* < \bar{q}_1 \ldots \bar{q}_n = q_1 \ldots q_n$ in contradiction to the assumption that $q_1 \ldots q_n$ is a minimum. Hence, we must have $q_1 = \ldots = q_{n-1} = 1$. Then, from equation 26 it follows that the minimum value of $q_1 \ldots q_n$ is given by

$$\frac{a_n}{1 - a_o - a_1 - \cdots - a_{n-1}} \quad .$$

If $i > 1$ but $< n$, the computation of the minimum value of $q_1 \ldots q_i$ is involved, since a large number of algebraic equations have to be solved. In the next part we shall discuss some approximation methods by means of which the amount of computational work can be considerably reduced.

PART III

APPROXIMATE DETERMINATION OF THE MAXIMUM VALUE OF THE PROBABILITY THAT A PLANE WILL BE DOWNED BY A GIVEN NUMBER OF HITS[1]

The symbols defined in parts I and II will be used here without further explanations. We have seen in part II that the exact determination of the maximum value of P_i $(i < n)$ involves a considerable amount of computational work, since a large number of algebraic equations have to be solved. The purpose of this memorandum is to derive some approximations to the maximum of P_i which can be computed much more easily than the exact values.

Let us denote the maximum of P_i by P_i^o and let $Q_i^o = 1 - P_i^o$. Thus, Q_i^o is the minimum value of Q_i. Before we derive approximate values of P_i^o (or Q_i^o) we shall discuss some simplifications that can be made in calculating the exact value P_i^o (or Q_i^o) assuming $1 < i < n$. We have seen in part II that Q_i^o is equal to the smallest of the $i - 1$ values $M_{i1}, \dots, M_{i,i-1}$. We shall make some simplifications in calculating M_{ir} $(r = 1, \dots, i-1)$.

For this purpose consider the equation

$$\frac{a_r}{u} + \frac{a_{r+1}}{uv} + \dots + \frac{a_n}{uv^{n-r}} = 1 - a_o - a_1 - \dots - a_{r-1}. \qquad (47)$$

It is clear that for any value $u > \dfrac{a_r}{1 - a_o - \dots - a_{r-1}}$, equation 47 has exactly one positive root in v. Denote this root by $\phi_r(u)$. Thus, $\phi_r(u)$ is defined for all values $u > \dfrac{a_r}{1 - a_o - \dots - a_{r-1}}$. In all that follows we shall assume that $a_i > 0$ $(i = 1, \dots, n)$. We shall prove that

$$\lim_{u \to \frac{a_r}{1-a_o-\dots-a_{r-1}}} \left(u \left[\phi_r(u) \right]^{i-r} \right) = +\infty \qquad (48)$$

[1]This part of "A Method of Estimating Plane Vulnerability Based on Damage of Survivors" was published as SRG memo 88 and AMP memo 76.3.

and

$$\lim_{u\to\infty} \left(u \left[\phi_r(u) \right]^{i-r} \right) = +\infty \quad . \tag{49}$$

It follows easily from equation 47 that if $u \to \dfrac{u_r}{1 - a_o - \ldots - a_{r-1}}$, then $\phi_r(u) \to +\infty$. Since $i > r$, we see that equation 48 must hold. It follows easily from equation 47 that $\lim_{u=+\infty} \phi_r(u) = 0$. We also see from equation 47 that if $u\to\infty$, the product $u\left[\phi_r(u)\right]^{n-r}$ must have a positive lower bound. Equation 49 follows from this and the fact that $\lim_{u\to\infty} \phi_r(u) = 0$.

We have seen in part II that equations 43* and 44* have exactly one positive root in the unknowns, q_1^* and q_2^*. Let the root in q_1^* be u_{ir}^o. Then the root in q_2^* is equal to $\phi_r(u_{ir}^o)$. From equations 48 and 49 it follows that $u\left[\phi_r(u)\right]^{i-r}$ is strictly decreasing in the interval $\dfrac{a_r}{1 - a_o - \ldots - a_{r-1}} < u < u_{ir}^o$, and is strictly increasing in the interval $u_{ir}^o < u < +\infty$.

Denote by u_r' the positive root of the equation

$$\frac{a_r}{u} + \frac{a_{r+1}}{u^2} + \ldots + \frac{a_n}{u^{n-r+1}} = 1 - a_o - \ldots - a_{r-1}. \tag{50}$$

It is clear that $u_r' < 1$ and $\phi_r(u_r') = u_r'$. The value M_{ir} is equal to the smallest of the three values

$$u_r' \left[\phi_r(u_r')\right]^{i-r}, \quad \left[\phi_r(1)\right]^{i-r}, \text{ and } u_{ir}^o \left[\phi_r(u_{ir}^o)\right]^{i-r}.$$

A simplification in the calculation of M_{ir} can be achieved by the fact that in some areas M_{ir} can be determined without calculating the value u_{ir}^o . We consider three cases.

Case A: $\quad u_r' \left[\phi_r(u_r')\right]^{i-r} < \left[\phi_r(1)\right]^{i-r}$.

In this case,

$$M_{ir} = u'_r \left[\phi_r(u'_r)\right]^{i-r} \text{ if } \frac{d}{du} u\left[\phi_r(u)\right]^{i-r} \geq 0 \text{ for } u = u'_r$$

and

$$M_{ir} = u^o_{ir} \left[\phi_r(u^o_{ir})\right]^{i-r} \text{ if } \frac{d}{du} u\left[\phi_r(u)\right]^{i-r} < 0 \text{ for } u = u'_r.$$

Case B: $\quad u'_r \left[\phi_r(u'_r)\right]^{i-r} > \left[\phi_r(1)\right]^{i-r}$.

In this case,

$$M_{ir} = \left[\phi_r(1)\right]^{i-r} \text{ if } \frac{d}{du} u\left[\phi_r(u)\right]^{i-r} \leq 0 \text{ for } u = 1$$

and

$$M_{ir} = u^o_{ir} \left[\phi_r(u^o_{ir})\right]^{i-r} \text{ if } \frac{d}{du} u\left[\phi_r(u)\right]^{i-r} > 0 \text{ for } u = 1.$$

Case C: $\quad u'_r \left[\phi(u'_r)\right]^{i-r} = \left[\phi(1)\right]^{i-r}$.

In this case,

$$M_{ir} = u^o_{ir} \left[\phi(u^o_{ir})\right]^{i-r} .$$

We can easily calculate the value of $\frac{d}{du} u\left[\phi_r(u)\right]^{i-r}$ for $u = u'_r$ and $u = 1$. In fact, we have

$$\frac{d}{du} \left[\phi_r(u)\right]^{i-r} = \left[\phi_r(u)\right]^{i-r} + (i - r)u \left[\phi_r(u)\right]^{i-r-1} \frac{d\phi_r(u)}{du} \quad (51)$$

and $\frac{d\phi_r(u)}{du} = \frac{dv}{du}$ can be obtained from equation 47 as follows.

Denote $\frac{a_r}{u} + \frac{a_{r+1}}{uv} + \ldots + \frac{a_n}{uv^{n-r}}$ by G(u,v). Then

$$\frac{d\phi_r(u)}{du} = \frac{dv}{du} = - \frac{\frac{\partial}{\partial u} G(u,v)}{\frac{\partial}{\partial v} G(u,v)}$$

$$= \frac{-\frac{1}{u}\left(\frac{a_r}{u} + \frac{a_{r+1}}{uv} + \ldots + \frac{a_n}{uv^{n-r}}\right)}{\frac{1}{u}\left(\frac{a_{r+1}}{v^2} + \frac{2a_{r+2}}{v^3} + \ldots + \frac{(n-r)a_n}{v^{n-r+1}}\right)} \tag{52}$$

$$= \frac{-(1 - a_o - a_1 - \ldots - a_{r-1})}{\left(\frac{a_{r+1}}{v^2} + \frac{2a_{r+2}}{v^3} + \ldots + \frac{(n-r)a_n}{v^{n-r+1}}\right)} \quad \left(v = \phi_r(u)\right) .$$

On the basis of equations 51 and 52, we can easily obtain the value of $\frac{d}{du}u\left[\phi_r(u)\right]^{i-r}$ for $u = u_r'$ and $u = 1$ if u_r' and $\phi_r(1)$ have been calculated. If $u = u_r'$, then $\phi_r(u) = v = u_r'$; if $u = 1$, then $v = \phi_r(1)$.

Since $\phi_r(1)$ is equal to the root of the equation in v

$$a_r + \frac{a_{r+1}}{v} + \ldots + \frac{a_n}{v^{n-r}} = 1 - a_o - a_1 - \ldots - a_{r-1} ,$$

it follows from equation 50 that

$$\phi_r(1) = u_{r+1}' . \tag{53}$$

Thus, for carrying out the investigations of cases A, B, and C for $r = 1,\ldots,i-1$, we merely have to calculate $u_1',\ldots,u_i'$.

If we want to calculate Q_i^o for all values $i < n$, then it seems best to compute first the n quantities $u_1',\ldots,u_n'$.

Since $u_r' = \phi_r(u_r')$ and $\phi_r(1) = u_{r+1}'$, we can say that M_{ir} is the smallest of the three values

$$\left(u_r'\right)^{i-r+1}, \left(u_{r+1}'\right)^{i-r}, \text{ and } u_{ir}^o \left[\phi_r(u_{ir}^o)\right]^{i-r} .$$

Since Q_i^o is equal to the minimum of the i - 1 values, $M_{i1},\ldots,M_{i,i-1}$, we see that

$$Q^o \leq t_i, \tag{54}$$

where

$$t_i = \text{Min} \left[(u_1')^i, (u_2')^{i-1},\ldots, (u_{i-1}')^2, u_i'\right] . \tag{55}$$

If n is large, it can be expected that Q_i^o will be nearly equal to t_i. Thus, t_i can be used as an approximation to Q_i^o . In order to see how good this approximation is, we shall derive a lower bound z_i for Q_i^o . If the difference $t_i - z_i$ is small, we are certain to have a satisfactory approximation to Q_i^o . If $t_i - z_i$ is large, then t_i still may be a good approximation to Q_i^o, since it may be that z_i is considerably below Q_i^o .

To obtain a lower bound z_i of Q_i^o, denote by y_j $(j = 0,1,\ldots,i-1)$ the proportion of planes (number of planes divided by the total number of planes participating in combat) that would be downed out of the returning planes with j hits if they were subject to i - j additional hits. Then

$$P_i = y_o + y_1 + \ldots + y_{i-1} + x_1 + x_2 + \ldots + x_i . \tag{56}$$

It is clear that $a_j P_i > y_j$ $(j = 0,1,\ldots,i-1)$ and consequently

$$(a_o + a_1 + \ldots + a_{i-1})\ P_i > y_o + y_1 + \ldots + y_{i-1}\ .$$

Hence,

$$\frac{y_o + y_1 + \ldots + y_{i-1}}{a_o + a_1 + \ldots + a_{i-1}} < P_i\ . \tag{57}$$

Equation 56 can be written

$$P_i = (a_o + \ldots + a_{i-1})\ \frac{y_o + y_1 + \ldots + y_{i-1}}{a_o + \ldots + a_{i-1}}$$
$$+ (1 - a_o - \ldots - a_{i-1})\ \frac{x_1 + \ldots + x_i}{1 - a_o - \ldots - a_{i-1}}\ . \tag{58}$$

Hence, P_i is a weighted average of $\frac{y_o + \ldots + y_{i-1}}{a_o + \ldots + a_{i-1}}$ and $\frac{x_i + \ldots + x_i}{1 - a_o - \ldots - a_{i-1}}$. Then, from equation 57 it follows that

$$P_i < \frac{x_1 + \ldots + x_i}{1 - a_o - a_1 - \ldots - a_{i-1}}\ . \tag{59}$$

Since $y_j > 0$, we obtain from equations 56 and 59

$$x_1 + \ldots + x_i < P_i < \frac{x_1 + \ldots + x_i}{1 - a_o - \ldots - a_{i-1}}\ . \tag{60}$$

Hence,

$$1 - \frac{x_1 + \ldots + x_i}{1 - a_o - \ldots - a_{i-1}} < Q_i < 1 - (x_1 + \ldots + x_i)\ . \tag{61}$$

In part II we have calculated the maximum value of $x_1 + \ldots + x_i$. Denote this maximum value by A_i. Then a lower bound of Q_i^o is given by

$$z_i = 1 - \frac{A_i}{1 - a_o - \ldots - a_{i-1}} < Q_i^o \quad . \tag{62}$$

NUMERICAL EXAMPLE

The same notation will be used as in the numerical examples for part I. q_i is the probability of a plane surviving the i-th hit, knowing that the first i - 1 hits did not down the plane. Then the probability that a plane will survive i hits is given by

$$Q_i = q_1 q_2 \ldots q_i \ .$$

In part I it was assumed that

$$q_1 = q_2 = \ldots = q_i = q_o \quad \text{(say)},$$

which is equivalent to the assumption that the probability that a plane will be shot down does not depend on the number of previous non-destructive hits. Under this assumption

$$Q_i = q_o^i \ .$$

The example below is based on the assumption that

$$q_1 \geq q_2 \geq \ldots \geq q_n,$$

i.e., the probability of surviving the i + 1 hit is less than or equal to the probability of surviving the i-th hit. In this case, it is not possible to find an explicit formula for Q_i, but a lower bound can be obtained. That is, a value of Q_i can be found such that the actual value of Q_i must lie above it. The greatest lower bound is denoted by Q_i^o . Hence, we have

$$Q_i^o \leq Q_i \ .$$

If

$$P_i^o = 1 - Q_i^o,$$

P_i^o is the least upper bound of P_i; that is, the probability of being downed by i bullets cannot be greater than P_i^o.

Since the computation of the exact value of Q_i^o is relatively complex, an approximate formula has been developed. This approximation is called t_i and $t_i \geq Q_i^o$. Another approximation (z_i) is available such that $z_i \leq Q_i^o$. However, z_i is not as accurate as t_i. Whenever the full computation is to be omitted, it is recommended that t_i be used.

The observed data of example 1, part I, will be used. Thus,

$$a_o = .80,\ a_1 = .08,\ a_2 = .05,\ a_3 = .01,\ a_4 = .005,\ a_5 = .005$$

The calculations are in three sections:

- The calculation of $t_i \geq Q_i^o$.
- The calculation of $z_i \leq Q_i^o$.
- The exact value of Q_i^o.

1. <u>Calculation of t_i ($t_i \geq Q_i^o$)</u>

(1) Calculate u_r', the positive root of equation 50:

$$\frac{a_r}{u} + \frac{a_{r+1}}{u^2} + \ldots + \frac{a_n}{u^{n-r+1}} = 1 - a_o - \ldots - a_{r-1} .$$

For r = 1, we obtain

$$\frac{a_1}{u} + \frac{a_2}{u^2} + \frac{a_3}{u^3} + \frac{a_4}{u^4} + \frac{a_5}{u^5} = 1 - a_o ,$$

which reduces to

$$.20u^5 - .08u^4 - .05u^3 - .01u^2 - .005u - .005 = 0$$

$$u_1' = .851 \ .$$

For r = 2,

$$\frac{a_2}{u} + \frac{a_3}{u^2} + \frac{a_4}{u^3} + \frac{a_5}{u^4} = 1 - a_o - a_1,$$

which reduces to

$$.12u^4 - .05u^3 - .01u^2 - .005u = 0$$

$$u_2' = .722 \ .$$

For r = 3,

$$\frac{a_3}{u} + \frac{a_4}{u^2} + \frac{a_5}{u^3} = 1 - a_o - a_1 - a_2 \ ,$$

which reduces to

$$.07u^3 - .01u^2 - .005u - .005 = 0$$

$$u_3' = .531 \ .$$

For r = 4,

$$\frac{a_4}{u} + \frac{a_5}{u^2} = 1 - a_o - a_1 - a_2 - a_3,$$

which reduces to

$$.06u^2 - .005u - .005 = 0$$

$$u_4' = .333 \ .$$

(2) $t_1,\ldots,t_5$ are given by equation 54:

$$t_i = \text{Min}\left[(u_1')^1,\ (u_2')^{i-1}, \ldots,\ (u_{i-1}')^2,\ (u_i')\right] .$$

We have

$$u_1' = .851,\ u_2' = .722,\ u_3' = .531,\ u_4' = .333 .$$

Hence,

$$\begin{aligned} t_1 &= \text{Min}\ [(u_1')] = u_1' \\ &= .851 \end{aligned}$$

$$\begin{aligned} t_2 &= \text{Min}\ [(u_1')^2,\ (u_2')] \\ &= \text{Min}\ [.724,\ .722] \\ &= .722 \end{aligned}$$

$$\begin{aligned} t_3 &= \text{Min}\ [(u_1')^3,\ (u_2')^2,\ (u_3')] \\ &= \text{Min}\ [.616,\ .521,\ .531] \\ &= .521 \end{aligned}$$

$$\begin{aligned} t_4 &= \text{Min}\ [(u_1')^4,\ (u_2')^3,\ (u_3')^2,\ (u_4')] \\ &= \text{Min}\ [.524,\ .376,\ .282,\ .333] \\ &= .282 \end{aligned}$$

t_5 is not calculated since the exact value of Q_5^o can be easily obtained.

2. <u>Calculation of z_i ($z_i \leq Q_i^o$)</u>

The following values must be obtained:

q_o, the root of equation 26A

$$\frac{a_1}{q} + \frac{a_2}{q^2} + \frac{a_3}{q^3} + \frac{a_4}{q^4} + \frac{a_5}{q^5} = 1 - a_o .$$

This has already been obtained as u_1' . Thus q_o = .851. The values of $x_1, \ldots, x_5$ have been calculated in part I:

$x_1 = .030,\ x_2 = .013,\ x_3 = .004,\ x_4 = .002,\ x_5 = .001.$

$A_i = x_1 + x_2 + \ldots + x_i.$

$A_1 = x_1 = .030$
$A_2 = x_1 + x_2 = .043$
$A_3 = x_1 + x_2 + x_3 = .047$
$A_4 = x_1 + x_2 + x_3 + x_4 = .049$
$A_5 = x_1 + x_2 + x_3 + x_4 + x_5 = .050.$

From equation 62 the lower bounds z_i are calculated:

$$z_i = 1 - \frac{A_i}{1 - a_o - \ldots - a_{i-1}} < Q_i^o .$$

Then

$$z_1 = 1 - \frac{A_1}{1 - a_o} = 1 - \frac{.030}{.20} = .850$$

$$z_2 = 1 - \frac{A_2}{1 - a_o - a_1} = 1 - \frac{.043}{.12} = .642$$

$$z_3 = 1 - \frac{A_3}{1 - a_o - a_1 - a_2} = 1 - \frac{.047}{.07} = .329$$

$$z_4 = 1 - \frac{A_4}{1 - a_o - a_1 - a_2 - a_3} = 1 - \frac{.049}{.06} = .183$$

z_5 is not calculated since Q_5^o can be obtained directly.

3. The Exact Value of Q_i^o

We have calculated t_i and z_i such that

$$z_i \le Q_i^o \le t_i \qquad (i = 1,2,\ldots,5) .$$

The exact value of Q_i^o is obtained as follows:

$$M_{ir} = \text{Min}\left\{(u_r')^{i-r+1},\ (u_{r+1}')^{i-r},\ u_{ir}^o\left[\phi_r(u_{ir}^o)\right]^{i-r}\right\},$$

where u_{ir}^o and $\phi_r(u_{ir}^o)$ will be defined below.

$$Q_i^o = \text{Min}\ [M_{i1},\ldots,M_{i,i-1}]$$

or combining these equations with the definition of t_i we obtain

$$Q_1^o = \text{Min}\ \{t_1\} = .851$$
$$Q_2^o = \text{Min}\ \{t_2,\ u_{21}^o[\phi_1(u_{21}^o)]\}$$
$$Q_3^o = \text{Min}\ \{t_3,\ u_{31}^o[\phi_1(u_{31}^o)]^2,\ u_{32}^o[\phi_2(u_{32}^o)]\}$$
$$Q_4^o = \text{Min}\ \{t_4,\ u_{41}^o[\phi_1(u_{41}^o)]^3,\ u_{42}^o[\phi_2(u_{42}^o)]^2,\ u_{43}^o[\phi_3(u_{43}^o)]\}$$

If $u_{ir}^o > 1$, $\left[\phi_r(u_{ir}^o)\right] > 1$, or $u_{ir}^o < \phi_r(u_{ir}^o)$, then $u_{ir}^o\left[\phi_r(u_{ir}^o)\right]^{i-r}$ is neglected in the equations above.

$$Q_5^o = \frac{a_5}{1 - a_o - a_1 - a_2 - a_3 - a_4} = \frac{.005}{.055} = .091.$$

In the equation of Q_i^o the additional quantities we have to compute are

$u_{21}^o \qquad \phi_1(u_{21}^o)$
$u_{31}^o \qquad \phi_1(u_{31}^o)$
$u_{32}^o \qquad \phi_2(u_{32}^o)$
$u_{41}^o \qquad \phi_1(u_{41}^o)$
$u_{42}^o \qquad \phi_2(u_{42}^o)$
$u_{43}^o \qquad \phi_3(u_{43}^o)$

The following equations have exactly one positive root in q_1^*, q_2^*. The root in q_1^* is u_{ir}^o; the root in q_2^* is $\phi_r(u_{ir}^o)$.

$$a_1^* + \frac{a_2^*}{q_2^*} + \frac{a_3^*}{(q_2^*)^2} + \dots + \frac{a_{n^*}^*}{(q_2^*)^{n^*-1}} = (1 - a_o^*)q_1^* ,$$

where q_2^* satisfies

$$(i^* - 1)a_1^* + \frac{(i^* - 2)a_2^*}{q_2^*} + \frac{(i^* - 3)a_3^*}{(q_2^*)^2} + \dots + \frac{(i^* - n^*)a_{n^*}^*}{(q_2^*)^{n^*-1}} = 0,$$

where

$$n^* = n - r + 1$$

$$a_o^* = a_o + a_1 + \dots + a_{r-1}$$

$$a_j^* = a_{j+r-1} \qquad (j = 1,2,\dots,n^*)$$

$$i^* = i - r + 1 .$$

The details of the computation are given in tables 2 and 3.

TABLE 2

u_{ir}^o	i	r	n*	i*	a_o^*	a_1^*	a_2^*	a_3^*	a_4^*	a_5^*
u_{21}^o	2	1	5	2	.80	.08	.05	.01	.005	.005
u_{31}^o	3	1	5	3	.80	.08	.05	.01	.005	.005
u_{32}^o	3	2	4	2	.88	.05	.01	.005	.005	
u_{41}^o	4	1	5	4	.80	.08	.05	.01	.005	.005
u_{42}^o	4	2	4	3	.88	.05	.01	.005	.005	
u_{43}^o	4	3	3	2	.93	.01	.005	.005		

where

$a_o = .80$, $a_1 = .08$, $a_2 = .05$, $a_3 = .01$, $a_4 = .005$, $a_5 = .005$

TABLE 3

Computation of	Equation	Numerical Equation	Result Obtained
$\phi_1(u^o_{21})$	$(i^*-1)a^*_1 + \frac{(i^*-2)a^*_2}{q^*_2} + \frac{(i^*-3)a^*_3}{(q^*_2)^2} + \frac{(i^*-4)a^*_4}{(q^*_2)^3} + \frac{(i^*-5)a^*_5}{(q^*_2)^4} = 0$	$.08(q^*_2)^4 - .01(q^*_2)^2 - .01(q^*_2) - .015 = 0$	.774
u^o_{21}	$a^*_1 + \frac{a^*_2}{q^*_2} + \frac{a^*_3}{(q^*_2)^2} + \frac{a^*_4}{(q^*_2)^3} + \frac{a^*_5}{(q^*_2)^4} = (1-a^*_o)q^*_1$	$.08 + \frac{.05}{.774} + \frac{.01}{(.774)^2} + \frac{.005}{(.774)^3} + \frac{.005}{(.774)^4} = .20q^*_1$	.932
$\phi_1(u^o_{31})$	$(i^*-1)a^*_1 + \frac{(i^*-2)a^*_2}{q^*_2} + \frac{(i^*-3)a^*_3}{(q^*_2)^2} + \frac{(i^*-3)a^*_4}{(q^*_2)^3} + \frac{(i^*-5)a^*_5}{(q^*{}_2)^4} = 0$	$.16(q^*_2)^4 + .05(q^*_2)^3 - .005(q^*_2) - .01 = 0$	.463
u^o_{31}	$a^*_1 + \frac{a^*_2}{q^*_2} + \frac{a^*_3}{(q^*_2)^2} + \frac{a^*_4}{(q^*_2)^3} + \frac{a^*_5}{(q^*_2)^4} = (1-a^*_o)q^*_1$	$.08 + \frac{.05}{.463} + \frac{.01}{(.463)^2} + \frac{.005}{(.463)^3} + \frac{.005}{(.463)^4} = .20q^*_1$	1.968[a]
$\phi_2(u^o_{32})$	$(i^*-1)a^*_1 + \frac{(i^*-2)a^*{}_2}{q^*_2} + \frac{(i^*-3)a^*_3}{(q^*_2)^2} + \frac{(i^*-4)a^*_4}{(q^*_2)^3} = 0$	$.05(q^*{}_2)^3 - .005(q^*_2) - .01 = 0$	.642
u^o_{32}	$a^*_1 + \frac{a^*_2}{q^*_2} + \frac{a^*_3}{(q^*_2)^2} + \frac{a^*_4}{(q^*_2)^3} = (1-a^*_o)q^*_1$	$.05 + \frac{.01}{.642} + \frac{.005}{(.642)^2} + \frac{.005}{(.642)^3} = .12q^*_1$	.805
$\phi_1(u^o_{41})$	$(i^*-1)a^*_1 + \frac{(i^*-2)a^*_2}{q^*_2} + \frac{(i^*-3)a^*_3}{(q^*_2)^2} + \frac{(i^*-4)a^*_4}{(q^*_2)^3} + \frac{(i^*-5)a^*_5}{(q^*_2)^4} = 0$	$.24(q^*_2)^4 + .10(q^*_2)^3 + .01(q^*_2)^2 - .005 = 0$	.290

TABLE 3 (Continued)

Computation of	Equation	Numerical Equation	Result Obtained
u^o_{41}	$a^*_1 + \frac{a^*_2}{q^*_2} + \frac{a^*_3}{(q^*_2)^2} + \frac{a^*_4}{(q^*_2)^3} + \frac{a^*_5}{(q^*_2)^4} = (1-a^*_o)q^*_1$	$.08 + \frac{.05}{.290} + \frac{.01}{(.290)^2} + \frac{.005}{(.290)^3} + \frac{.005}{(.290)^4} = .20q^*_1$	6.402[b]
$\phi_2(u^o_{42})$	$(i^*-1)a^*_1 + \frac{(i^*-2)a^*_2}{q^*_2} + \frac{(i^*-3)a^*_3}{(q^*_2)^2} + \frac{(i^*-4)a^*_4}{(q^*_2)^3} = 0$	$.10(q^*_2)^3 + .01(q^*_2)^2 - .005 = 0$	.338
u^o_{42}	$a^*_1 + \frac{a^*_2}{q^*_2} + \frac{a^*_3}{(q^*_2)^2} + \frac{a^*_4}{(q^*_2)^3} = (1-a^*_o)q^*_1$	$.05 + \frac{.01}{.338} + \frac{.005}{(.338)^2} + \frac{.005}{(.338)^3} = .12q^*_1$	2.108[c]
$\phi_3(u^o_{43})$	$(i^*-1)a^*_1 + \frac{(i^*-2)a^*_2}{q^*_2} + \frac{(i^*-3)a^*_3}{(q^*_2)^2} = 0$	$.01(q^*_2)^2 - .005 = 0$	.707
u^o_{43}	$a^*_1 + \frac{a^*_2}{q^*_2} + \frac{a^*_3}{(q^*_2)^2} = (1-a^*_o)q^*_1$	$.01 + \frac{.005}{.707} + \frac{.005}{(.707)^2} = .07q^*_1$	.387[d]

[a] $1.968 > 1 \therefore [u^o_{31}\ \phi_1(u^o_{31})]^2$ is not used.

[b] $6.402 > 1 \therefore [u^o_{41}\ \phi_1(u^o_{41})]^3$ is not used.

[c] $2.108 > 1 \therefore u^o_{42}[\phi_2(u^o_{42})]^2$ is not used.

[d] $.387 < \phi_3(u^o_{43}) \therefore u^o_{43}[\phi_3(u^o_{43})]$ is not used.

Substituting the values from table 3 in equation A and neglecting several terms as explained in table 3, we have

$$Q_1^o = .851$$

$$Q_2^o = \text{Min } \{.722, .721\} = .721$$

$$Q_3^o = \text{Min } \{.521, .517\} = .517$$

$$Q_4^o = .282$$

$$Q_5^o = .091$$

The results obtained are shown in table 4.

TABLE 4

i	z_i	Q_i^o	t_i	q_o^i
1	.851	.851	.851	.851
2	.642	.721	.722	.724
3	.329	.517	.521	.616
4	.183	.282	.282	.524
5	--	.091	--	.446

Thus, with the observed data, this example, if all the information available about the q_i's is that

$$q_1 \geq q_2 \geq \cdots \geq q_5 ,$$

all we can say about the Q_i is that

$$Q_1 \geq .85,\ Q_2 \geq .72,\ Q_3 \geq .52,\ Q_4 \geq .28,\ Q_5 = .09 .$$

Note that

$$z_1 = Q_1^o = t_1 = q_o .$$

This is always true.

It is interesting to compare Q_i^o with the values of Q_i obtained under the assumption that all the q_i's are equal and have the value q_o. Under this assumption,

$$Q_i = q_o^i \qquad (i = 1,2,...,5).$$

In table 4, $Q_1^o = q_o$ and Q_2^o is very close to q_o^2 . Q_3^o and q_o^3 differ by approximately .1 and the agreement between Q_i^o and q_o^i gets progressively worse. It will usually be true that q_o^i and Q_i^o are approximately equal for small values of i, but will differ widely as i increases.

PART IV

MINIMUM AND MAXIMUM VALUE OF THE PROBABILITY THAT A PLANE WILL BE DOWNED BY A GIVEN NUMBER OF HITS CALCULATED UNDER SOME FURTHER RESTRICTIONS ON THE PROBABILITIES $q_1,\ldots,q_n$[1]

In parts I, II, and III we merely assumed that $q_1 \geq q_2 \geq \ldots \geq q_n$ In many cases we may have some further a priori knowledge concerning the values $q_1,\ldots,q_n$. We shall consider here the case when it is known a priori that $\lambda_1 q_j \leq q_{j+1} \leq \lambda_2 q_j$ $(j = 1,\ldots,n-1)$, where λ_1 and λ_2 $(\lambda_1 < \lambda_2 < 1)$ are known positive constants.

We shall also assume that

$$\sum_{j=1}^{n} \frac{a_j}{\lambda_1^{\frac{j(j-1)}{2}}} < 1 - a_o \ . \qquad (63)$$

Since $a_1 + a_2 + \ldots + a_n < 1 - a_o$, the inequality in equation 63 is certainly fulfilled if λ_1 is sufficiently near 1. It follows immediately from equations 63 and 26 that $q_1 < 1$.

CALCULATION OF THE MINIMUM VALUE OF $Q_i = 1 - P_i$ $(i < n)$

Let $q_1^o,\ldots,q_n^o$ be the values of $q_1,\ldots,q_n$ for which Q_i becomes a minimum. We shall prove the following.

<u>Lemma 1</u>: The relations

$$q_{j+1}^o = \lambda_2 q_j^o \qquad (j = i,\ldots,n-1) \qquad (64)$$

must hold.

Proof: Suppose that the relation in equation 64 does not hold for at least one value $j \geq i$ and we shall derive a contradiction.

---

[1]This part of "A Method of Estimating Plane Vulnerability Based on Damage of Survivors" was published as SRG memo 89 and AMP memo 76.4.

Let $q_r' = q_r^o$ for $r = 1,\ldots,i$ and $q_{j+1}' = \lambda_2 q_j'$ for $j = i,\ldots,n-1$. Then we have

$$q_1' \ldots q_i' = q_1^o \ldots q_i^o \text{ and } \sum_{j=1}^{n} \frac{a_j}{q_1' \ldots q_j'} < 1 - a_o. \quad (65)$$

Hence, there exists a positive value $\Delta < 1$ such that

$$\sum_{j=1}^{n} \frac{a_j}{q_1'' \ldots q_j''} = 1 - a_o ,$$

where $q_j'' = \Delta q_j'$ $(j = 1,\ldots,n)$. But then

$$q_1'' \ldots q_i'' < q_1' \ldots q_i' \; q_1^o \ldots q_i^o$$

in contradiction to our assumption that $q_1^o \ldots q_i^o$ is a minimum. Hence, Lemma 1 is proved.

<u>Lemma 2:</u> If j is the smallest integer such that $q_{k+1}^o = \lambda_2 q_k^o$ for all $k \geq j$, then $q_r^o = \lambda_1 q_{r-1}^o$ for $r = 2,3,\ldots,j-1$.

Proof: Assume that Lemma 2 does not hold and we shall derive a contradiction. Let u be the smallest integer greater than one such that $q_u^o > \lambda_1 q_{u-1}^o$. It follows from the definition of the integer u that if $u > 2$, then $q_{u-1}^o = \lambda_1 q_{u-2}^o$. From assumption 63 it follows that $q_1^o < 1$. Hence, if we replace q_{u-1}^o by $q_{u-1}' = (1 + \varepsilon) q_{u-1}^o$ $(\varepsilon > 0)$, then for sufficiently small ε the inequalities $\lambda_1 q_r \leq q_{r+1} \leq \lambda_2 q_r$ $(r = 1,\ldots,n-1)$ will not be disturbed. Let v be the smallest integer greater than or equal to u such that $q_{v+1}^o < \lambda_2 q_v^o$. Since by assumption j is the smallest integer such that $q_{k+1}^o = \lambda_2 q_k^o$ for all $k \geq j$, we must have $q_j < \lambda_2 q_{j-1}$. Hence, $v \leq j-1$. It is clear that replacing q_v^o by $q_v' = \frac{q_v^o}{1 + \varepsilon}$ we shall not disturb the inequalities $\lambda_1 q_r \leq q_{r+1} \leq \lambda_2 q_r$ $(r = 1,\ldots,n-1)$. Hence, if

$$q'_{u-1} = (1 + \varepsilon)\, q^o_{u-1},\ q'_v = \frac{q^o_v}{1+\varepsilon}, \text{ and } q'_r = q^o_r$$

for $r \neq u, \neq v$, then $\lambda_1 q'_k \leq q'_{k+1} \leq \lambda_2 q'_k$ $(k = 1,\ldots,n-1)$ is fulfilled. Furthermore, we have

$$q'_1 \ldots q'_i = q^o_1 \ldots q^o_i \text{ and } \sum_{j=1}^{n} \frac{a_j}{q'_1 \ldots q'_j} < 1 - a_o.$$

Hence, there exists a positive $\Delta < 1$ such that

$$\sum_{j=1}^{n} \frac{a_j}{q''_1 \ldots q''_j} = 1 - a_o$$

and $q''_j = \Delta q'_j$ $(j = 1,\ldots,n)$. But then

$$q''_1 \ldots q''_i < q'_1 \ldots q'_i = q^o_1 \ldots q^o_i$$

in contradiction to the assumption that $q^o_1 \ldots q^o_i$ is a minimum. Hence, Lemma 2 is proved.

Let E_{ir} $(r = 1,\ldots,i-1)$ be the minimum value of Q_i under the restriction that $q_{j+1} = \lambda_2 q_j$ for $j = r+1,\ldots,n-1$ and $q_{j+1} = \lambda_1 q_j$ for $j = 1,\ldots,r-1$. From Lemma 1 and 2 it follows that the minimum of Q_i is equal to the smallest of the $i - 1$ values $E_{i1},\ldots,E_{i,i-1}$. The computation of the exact value of E_{ir} can be carried out in a way similar to the computation of M_{ir} described in part II. Since these computations are involved if n is large, we shall discuss here an approximation method.

Let E^*_{ir} $(r = 1,\ldots,i-1)$ be the value of Q_i if $q_{j+1} = \lambda_2 q_j$ for $j = r+1,\ldots,n-1$ and $q_{j+1} = \lambda_1 q_j$ for $j = 1,\ldots,r$. Furthermore, let E^*_{io} be the value of Q_i if $q_{j+1} = \lambda_2 q_j$ $(j = 1,\ldots,n-1)$. Then, if n is large, the minimum of $E^*_{i,r-1}$ and E^*_{ir} will be nearly equal to E_{ir}. Hence, we obtain an approximation to the minimum of Q_i by taking the minimum of the i numbers $E^*_{io}, E^*_{i1},\ldots,E^*_{i,i-1}$. The quantity E_{ir} can be computed as follows. Let g_r be the positive root in q of the equation

$$\sum_{j=1}^{r+1} \frac{a_j}{\lambda_1^{\frac{j(j-1)}{2}} q^j} + \sum_{j=1}^{n-r-1} \frac{a_{r+1+j}}{\lambda_1^{\frac{r(r+1)}{2}+rj} \lambda_2^{\frac{j(j+1)}{2}} q^{r+1+j}} = 1 - a_o \tag{66}$$

$(r = 0,1,\ldots,i-1)$.

Then

$$E^*_{ir} = \lambda_1^{\frac{r(r+1)}{2}+r(i-r-1)} \quad \lambda_2^{\frac{(i-r)(i-r-1)}{2}} \quad g_r^i \; . \tag{67}$$

MINIMUM OF Q_n

Let $q_1^o,\ldots,q_n^o$ be values of $q_1,\ldots,q_n$ for which Q_n becomes a minimum. We shall prove that $q_{j+1}^o = \lambda_1 q_j^o$ $(j = 1,\ldots,n-1)$. Assume that there exists a value $j < n$ such that $q_{j+1}^o > \lambda_1 q_j^o$ and we shall derive a contradiction. Let u be the smallest integer such that $q_{u+1}^o > \lambda_1 q_u^o$ and let v be the largest integer such that $q_{v+1}^o > \lambda_1 q_v^o$. Let $q_u' = (1 + \varepsilon)q_u^o$ $(\varepsilon > 0)$, $q_{v+1}' = \frac{q_{v+1}^o}{1 + \varepsilon}$, and $q_j' = q_j^o$ for $j \neq u, \neq v+1$. Then for sufficiently small ε we shall have $\lambda_1 q_r' \leq q_{r+1}' \leq \lambda_2 q_r'$ $(r = 1,\ldots,n-1)$. Furthermore, we have

$$q_1 \;\ldots\; q_n' = q_1^o \;\ldots\; q_n^o \text{ and } \sum_{j=1}^{n} \frac{a_j}{q_1' \;\ldots\; q_j'} < 1 - a_o \; .$$

Hence, there exists a positive $\Delta < 1$ such that $q_j'' = q_j'$ $(j = 1,\ldots,n)$ and

$$\sum_{j=1}^{n} \frac{a_j}{q_1'' \;\ldots\; q_j''} = 1 - a_o.$$

But then $q''_1 \ldots q''_n < q^o_1 \ldots q^o_n$ in contradiction to the assumption that $q^o_1 \ldots q^o_n$ is a minimum. Hence, our statement is proved.

If q is the root of the equation

$$\sum_{j=1}^{n} \frac{a_j}{\lambda_1^{\frac{j(j-1)}{2}} q^j} = 1 - a_o ,$$

then the minimum of Q_n is equal to $\lambda_1^{\frac{n(n-1)}{2}} q^n$.

MAXIMUM OF Q_i $(i < n)$

Let $q^*_1,\ldots,q^*_n$ be values of $q_1,\ldots,q_n$ for which Q_i becomes a maximum. We shall prove the following:

<u>Lemma 3:</u> The relations

$$q^*_{j+1} = \lambda_1 q^*_j \qquad (j = i,\ldots,n-1) \qquad (68)$$

must hold.

Proof: Assume that there exists an integer $j \geq i$ such that $q^*_{j+1} > \lambda_1 q^*_j$ and we shall derive a contradiction. Let $q'_r = q^*_r$ for $r = 1,\ldots,i$ and let $q'_{j+1} = \lambda_1 q'_j$ $(j = i,\ldots,n-1)$. Then

$$q'_1 \ldots q'_i = q^*_1 \ldots q^*_i \text{ and } \sum_{j=1}^{n} \frac{a_j}{q'_1 \ldots q'_j} > 1 - a_o .$$

Hence, there exists a value $\Delta > 1$ such that

$$\sum_{j=1}^{n} \frac{a_j}{q''_1 \ldots q''_j} = 1 - a_o,$$

where $q''_j = \Delta q'_j$ $(j = 1,\ldots,n)$. But then $q''_1 \ldots q''_i > q^*_1 \ldots q^*_i$ in contradiction to the assumption that $q^*_1 \ldots q^*_i$ is a maximum. Hence, Lemma 3 is proved.

Lemma 4: If for some $j < i$ we have $q^*_{j+1} > \lambda_1 q^*_j$, then $q^*_{k+1} = \lambda_2 q^*_k$ for $k = 1,\ldots,j-1$.

Proof: Assume that $q^*_{j+1} > \lambda_1 q^*_j$ for some $j < i$ and that there exists an integer $k \leq j-1$ such that $q^*_{k+1} < \lambda_2 q^*_k$. We shall derive a contradiction from this assumption. Let u be the smallest integer such that $q^*_{u+1} < \lambda_2 q^*_u$. Furthermore, let v be the smallest integer greater than or equal to u + 1 such that $q^*_{v+1} > \lambda_1 q^*_v$. It is clear that $v \leq j$. Let $q'_u = \frac{q^*_u}{1+\varepsilon}$ $(\varepsilon > 0)$, $q'_v = (1+\varepsilon)\, q^*_v$, and $q'_r = q^*_r$ for $r \neq u, \neq v$. Then for sufficiently small ε we have

$$\lambda_1 q'_j \leq q'_{j+1} \leq \lambda_2 q'_j \qquad (j = 1,\ldots,n-1)\ .$$

Furthermore, we have

$$q'_1 \ldots q'_i = q^*_1 \ldots q^*_i \text{ and } \sum_{j=1}^{n} \frac{a_j}{q'_1 \ldots q'_j} > 1 - a_o\ .$$

Hence, there exists a value $\Delta > 1$ such that

$$\sum_{j=1}^{n} \frac{a_j}{q''_1 \ldots q''_j} = 1 - a_o\ ,$$

where $q''_j = \Delta q'_j$ $(j = i,\ldots,n)$. But then $q''_1 \ldots q''_i > q^*_1 \ldots q''_i$ in contradiction to the assumption that $q^*_1 \ldots q^*_i$ is a maximum.

Let D_{ir} $(r = 1,\ldots,i-1)$ be the maximum of Q_i under the restriction that $q_{j+1} = \lambda_1 q_j$ for $j = r+1,\ldots,n-1$ and $q_{j+1} = \lambda_2 q_j$ for $j = 1,\ldots,r-1$. From Lemma 3 and 4 it follows that the maximum of Q_i is equal to the maximum of the $i - 1$ values $D_{i1},\ldots,D_{i,i-1}$. The computation of the exact value of D_{ir} can be carried out in a way similar to the computation of M_{ir} in part II. Since these computations are involved if n is large, we shall discuss here only an approximation method.

Let D^*_{ir} $(r = 1,\ldots,i-1)$ be the value of Q_i if $q_{j+1} = \lambda_1 q_j$ for $j = r+1,\ldots,n-1$ and $q_{j+1} = \lambda_2 q_j$ for $j = 1,\ldots,r$. Furthermore, let D^*_{io} be the value of Q_i if $q_{j+1} = \lambda_1 q_j$ $(j = 1,\ldots,n-1)$. Then, if λ_1 is not much below one, the maximum of D^*_{ir} and $D^*_{i,r-1}$ $(r = 1,\ldots,i-1)$ will be nearly equal to D_{ir}. Hence, we obtain an approximation to the maximum value of Q_i by taking the largest of the i values $D^*_{io},\ldots,D^*_{i,i-1}$.

The value of D^*_{ir} can be determined as follows. Let g_r be the root in q of the equation

$$\sum_{j=1}^{r+1} \frac{a_j}{\lambda_2^{\frac{j(j-1)}{2}} q^j} + \sum_{j=1}^{n-r-1} \frac{a_{r+1+j}}{\lambda_2^{\frac{r(r+1)}{2}+jr} \lambda_1^{\frac{j(j+1)}{2}} q^{r+1+j}} = 1 - a_o.$$

Then

$$D^*_{ir} = \lambda_2^{\frac{r(r+1)}{2}+(i-r-1)} \lambda_1^{\frac{(i-r-1)\ (i-r)}{2}} g_r^i .$$

MAXIMUM OF Q_n

We shall prove that the maximum of Q_n is reached when $q_{j+1} = \lambda_2 q_j$ $(j = 1,\ldots,n-1)$. Denote by $q^*_1 \ldots q^*_n$ the values of $q_1 \ldots q_n$ for which Q_n becomes a maximum. We shall assume that there exists a value $j < n$ such that $q^*_{j+1} < \lambda_2 q^*_j$ and we shall derive a contradiction from this assumption. Let u be the smallest and v be the largest integer such that $q^*_{u+1} < \lambda_2 q^*_u$ and $q^*_{v+1} < \lambda_2 q^*_v$.

Let $q'_u = \frac{q^*_u}{1 + \varepsilon}$ $(\varepsilon > 0)$, $q'_{v+1} = (1 + \varepsilon)\, q^*_{v+1}$, and $q'_r = q^*_r$ for $r \neq u, \neq v+1$. Then for sufficiently small ε we shall have $\lambda_1 q'_r \leq q'_{r+1} \leq \lambda_2 q'_r$ $(r = 1,\ldots,n-1)$.

Furthermore, we have

$$q_1' \cdots q_n' = q_1^* \cdots q_n^* \text{ and } \sum_{j=1}^{n} \frac{q_j}{q_1' \cdots q_j'} > 1 - a_o .$$

Hence, there exists a value $\Delta > 1$ such that $q_j'' = \Delta q_j'$ $(j = 1,\ldots,n)$ and

$$\sum_{j=1}^{n} \frac{a_j}{q_1'' \cdots q_j''} = 1 - a_o .$$

But then $q_1'' \cdots q_n'' > q_1^* \cdots q_n^*$ in contradiction to the assumption that $q_1^* \cdots q_n^*$ is a maximum. Hence, our statement is proved.

The maximum of Q_n is equal to

$$\lambda_2^{\frac{n(n-1)}{2}} q^n ,$$

where q is the root of the equation

$$\sum_{j=1}^{n} \frac{a_j}{\lambda_2^{\frac{j(j-1)}{2}} q^j} = 1 - a_o .$$

NUMERICAL EXAMPLE

The same notation will be used as in the previous numerical examples. The assumption of no sampling error, which is common to all the previous examples, is retained. In part I it was assumed that the q_i, the probability of a plane surviving the i-th hit, knowing that the first i - 1 hits did not down the plane, were equal for all i ($q_1 = q_2 = \ldots = q_n = q_o$ (say)). Under this assumption, the exact value of the probability of a plane surviving i hits is given by

$$Q_i = q_o^i .$$

In part III it was assumed that $q_1 \geq q_2 \geq \ldots \geq q_n$. Since no lower limit is assumed in the decrease from q_i to q_{i+1}, only a lower bound to the Q_i could be obtained. The assumption here is that the decrease from q_i to q_{i+1} lies between definite limits. Therefore, both an upper and lower bound for the Q_i can be obtained.

We assume that

$$\lambda_1 q_i \leq q_{i+1} \leq \lambda_2 q_i \ ,$$

where $\lambda_1 < \lambda_2 < 1$ and such that the expression

$$\sum_{j=1}^{n} \frac{a_j}{\lambda_1^{\frac{j(j-1)}{2}}} < 1 - a_o \qquad \text{(A)}$$

is satisfied.

The exact solution is tedious but close approximations to the upper and lower bounds to the Q_i for $i < n$ can be obtained by the following procedure. The set of hypothetical data used is

$$\begin{array}{llll} a_o = .780 & & a_3 = .010 \\ a_1 = .070 & & a_4 = .005 \\ a_2 = .040 & & a_5 = .005 \\ \lambda_1 = .80 & & \lambda_2 = .90 \end{array}$$

Condition A is satisfied, since by substitution

$$.07 + \frac{.04}{.8} + \frac{.01}{(.8)^3} + \frac{.005}{(.8)^6} + \frac{.005}{(.8)^{10}} = .20529 \ ,$$

which is less than

$$1 - a_o = .22 \ .$$

THE LOWER LIMIT OF Q_i

The first step is to solve equation 66. This involves the solution of the following four equations for positive roots g_o, g_1, g_2, g_3.

$$\frac{a_1}{q} + \frac{a_2}{\lambda_2 q^2} + \frac{a_3}{\lambda_2^3 q^3} + \frac{a_4}{\lambda_2^6 q^4} + \frac{a_5}{\lambda_2^{10} q^5} = 1 - a_o = .22 \qquad \text{(B)}$$

$$\frac{.07}{q} + \frac{.04}{.9q^2} + \frac{.01}{.729q^3} + \frac{.005}{.531441q^4} + \frac{.005}{.348678q^5} = .22$$

$$.22q^5 - .07q^4 - .044444q^3 - .013717q^2 - .009408q - .014340 = 0$$

$g_o = .844.$

$$\frac{a_1}{q} + \frac{a_2}{\lambda_1 q^2} + \frac{a_3}{\lambda_1^2 \lambda_2 q^3} + \frac{a_4}{\lambda_1^3 \lambda_2^3 q^4} + \frac{a_5}{\lambda_1^4 \lambda_2^6 q^5} = 1 - a_o \qquad \text{(C)}$$

$$\frac{.07}{q} + \frac{.04}{.8q^2} + \frac{.01}{(.64)(.9)q^3} + \frac{.005}{(.512)(.729)q^4} + \frac{.005}{(.4096)(.531441)q^5} = .22$$

$$.22q^5 - .07q^4 - .05q^3 - .017361q^2 - .013396q - .022970 = 0$$

$g_1 = .904.$

$$\frac{a_1}{q} + \frac{a_2}{\lambda_1 q^2} + \frac{a_3}{\lambda_1^3 q^3} + \frac{a_4}{\lambda_1^5 \lambda_2 q^4} + \frac{a_5}{\lambda_1^7 \lambda_2^3 q^5} = 1 - a_o \qquad \text{(D)}$$

$$\frac{.07}{q} + \frac{.04}{.8q^2} + \frac{.01}{.512q^3} + \frac{.005}{(.32768)(.9)q^4} + \frac{.005}{(.209715)(.729)q^5} = .22$$

$$.22q^5 - .07q^4 - .05q^3 - .019531q^2 - .016954q - .032705 = 0$$

$g_2 = .941.$

$$\frac{a_1}{q} + \frac{a_2}{\lambda_1 q^2} + \frac{a_3}{\lambda_1^3 q^3} + \frac{a_4}{\lambda_1^6 q^4} + \frac{a_5}{\lambda_1^9 \lambda_2 q^5} = 1 - a_o \qquad \text{(E)}$$

$$\frac{.07}{q} + \frac{.04}{.8q^2} + \frac{.01}{.512q^3} + \frac{.005}{.262144q^4} + \frac{.005}{(.134218)(.9)q^5} = .22$$

$$.22q^5 - .07q^4 - .05q^3 - .019531q^2 - .019073q - .041392 = 0$$

$$g_3 = .964 \ .$$

Next, calculate the i numbers defined by

$$E^*_{ir} = \lambda_1^{a(i,r)} \ \lambda_2^{b(i,r)} \ g_r^i \qquad (r = 0,1,\ldots,i-1),$$

where

$$a(i,r) = \frac{r(r + 1)}{2} + r(i - r - 1)$$

$$b(i,r) = \frac{(i - r)(i - r - 1)}{2}$$

$$g_o = .844$$
$$g_1 = .904$$
$$g_2 = .941$$
$$g_3 = .964$$

The minimum of the E^*_{ir} $(r = 0,\ldots,i-1)$ will be the lower limit of Q_i. The computations are given in table 5.

TABLE 5

COMPUTATION OF LOWER LIMIT OF Q_i

Q_i	i	r	a(i,r)	b(i,r)	g_r	g_r^i	E^*_{ir}
Q_1	1	0	0	0	.844	.844	.844

Min $[E^*_{10}]$ = .844

Q_i	i	r	a(i,r)	b(i,r)	g_r	g_r^i	E^*_{ir}
Q_2	2	0	0	1	.844	.712	.641
	2	1	1	0	.904	.817	.654

Min $[E^*_{20}, E^*_{21}]$ = .641

Q_i	i	r	a(i,r)	b(i,r)	g_r	g_r^i	E^*_{ir}
Q_3	3	0	0	3	.844	.601	.438
	3	1	2	1	.904	.739	.426
	3	2	3	0	.941	.833	.427

Min $[E^*_{30}, E^*_{31}, E^*_{32}]$ = .426

Q_i	i	r	a(i,r)	b(i,r)	g_r	g_r^i	E^*_{ir}
Q_4	4	0	0	6	.844	.507	.270
	4	1	3	3	.904	.668	.249
	4	2	5	1	.941	.784	.231
	4	3	6	0	.964	.864	.226

Min $[E^*_{40}, E^*_{41}, E^*_{42}, E^*_{43}]$ = .226

The lower limit of Q_5 can be obtained directly. The lower limit of

$$Q_5 = \lambda_1^{10} q^5 ,$$

where q is the positive root of

$$\frac{a_1}{q} + \frac{a_2}{\lambda_1 q^2} + \frac{a_3}{\lambda_1^3 q^3} + \frac{a_4}{\lambda_1^6 q^4} + \frac{a_5}{\lambda_1^{10} q^5} = 1 - a_o$$

$$\frac{.07}{q} + \frac{.04}{.8q^2} + \frac{.01}{.512q^3} + \frac{.005}{.262144q^4} + \frac{.005}{.107374q^5} = .22$$

$$q = .974 .$$

The lower limit of

$$Q_5 = (.8)^{10}(.974)^5 = .094 .$$

THE UPPER LIMIT OF Q_i

The computations for the upper limit of Q_1 are entirely analogous to the computations of the lower limit. First, we solve the equations of part IV, which for this example are the following:

$$\frac{a_1}{q} + \frac{a_2}{\lambda_1 q^2} + \frac{a_3}{\lambda_1^3 q^3} + \frac{a_4}{\lambda_1^6 q^4} + \frac{a_5}{\lambda_1^{10} q^5} = 1 - a_o$$

$$\frac{.07}{q} + \frac{.04}{.8q^2} + \frac{.01}{.512q^3} + \frac{.005}{.262144q^4} + \frac{.005}{.107374q^5} = .22$$

$$.22q^5 - .07q^4 - .05q^3 - .019531q^2 - .019073q - .046566 = 0$$

$$g_o^* = .974$$

$$\frac{a_1}{q} + \frac{a_2}{\lambda_2 q^2} + \frac{a_3}{\lambda_2^2 \lambda_1 q^3} + \frac{a_4}{\lambda_2^3 \lambda_1^3 q^4} + \frac{a_5}{\lambda_2^4 \lambda_1^6 q^5} = 1 - a_o$$

$$\frac{.07}{q} + \frac{.04}{.9q^2} + \frac{.01}{(.81)(.8)q^3} + \frac{.005}{(.729)(.512)q^4} + \frac{.005}{(.6561)(.262144)q^5} = .22$$

$$.22q^5 - .07q^4 - .044444q^3 - .015432q^2 - .013396q - .029071 = 0$$

$$g_1^* = .905$$

$$\frac{a_1}{q} + \frac{a_2}{\lambda_2 q^2} + \frac{a_3}{\lambda_2^3 q^3} + \frac{a_4}{\lambda_2^5 \lambda_1 q^4} + \frac{a_5}{\lambda_2^3 \lambda_1^7 q^5} = 1 - a_o$$

$$\frac{.07}{q} + \frac{.04}{.9q^2} + \frac{.01}{.729q^3} + \frac{.005}{(.59049)(.8)q^4} + \frac{.005}{(.512)(.478297)q^5} = .22$$

$$.22q^5 - .07q^4 - .044444q^3 - .013717q^2 - .010584q - .020417 = 0$$

$$g_2^* = .869$$

$$\frac{a_1}{q} + \frac{a_2}{\lambda_2 q^2} + \frac{a_3}{\lambda_2^3 q^3} + \frac{a_4}{\lambda_2^6 q^4} + \frac{a_5}{\lambda_2^9 \lambda_1 q^5} = 1 - a_o$$

$$\frac{.07}{q} + \frac{.04}{.9q^2} + \frac{.01}{.729q^3} + \frac{.005}{.531441q^4} + \frac{.005}{(.387420)(.8)q^5} = .22$$

$$.22q^5 - .07q^4 - .044444q^3 - .013717q^2 - .009408q - .016132 = 0$$

$$g_3^* = .851$$

Next, calculate the i numbers defined by

$$D^*_{ir} = \lambda_2^{a(i,r)} \lambda_1^{b(i,r)} g_r^{*i} \qquad (r = 0,1,\ldots,i-1),$$

where

$$a(i,r) = \frac{r(r+1)}{2} + r(i - r - 1)$$

$$b(i,r) = \frac{(i - r)(i - r - 1)}{2}$$

$$g^*_o = .974$$
$$g^*_1 = .905$$
$$g^*_2 = .869$$
$$g^*_3 = .851$$

The maximum of the D^*_{ir} $(r = 0,\ldots,i-1)$ will be the upper limit of Q_i. The computations are given in table 6.

The upper limit of Q_5 can be obtained directly. The limit of

$$Q_5 = \lambda_2^{10} q^{*5},$$

where q* is the positive root of

$$\frac{a_1}{q} + \frac{a_2}{\lambda_2 q^2} + \frac{a_3}{\lambda_2^3 q^3} + \frac{a_4}{\lambda_2^6 q^4} + \frac{a_5}{\lambda_2^{10} q^5} = 1 - a_o$$

$$\frac{.07}{q} + \frac{.04}{.9q^2} + \frac{.01}{.729q^3} + \frac{.005}{.531441q^4} + \frac{.005}{.348678q^5} = .22$$

$$q^* = .844.$$

TABLE 6

COMPUTATION OF UPPER LIMIT OF Q_i

Q_i	i	r	a(i,r)	b(i,r)	g_r^*	g_r^{*i}	D_{ir}^*
Q_1	1	0	0	0	.974	.974	.974

Max $[D_{10}^*]$ = .974

Q_2	2	0	0	1	.974	.949	.759
	2	1	1	0	.905	.819	.737

Max $[D_{20}^*, D_{21}^*]$ = .759

Q_3	3	0	0	3	.974	.924	.473
	3	1	2	1	.905	.741	.480
	3	2	3	0	.869	.656	.478

Max $[D_{30}^*, D_{31}^*, D_{32}^*]$ = .480

Q_4	4	0	0	6	.974	.890	.236
	4	1	3	3	.905	.671	.250
	4	2	5	1	.869	.570	.269
	4	3	6	0	.851	.524	.279

Max $[D_{40}^*, D_{41}^*, D_{42}^*, D_{43}^*]$ = .279

The upper limit of

$$Q_5 = (.9)^{10}(.844)^5 = .149$$

Summarizing the results, the upper and lower limits of the probability of a plane surviving i hits are given by

$$.844 < Q_1 < .974$$
$$.641 < Q_2 < .759$$
$$.426 < Q_3 < .480$$
$$.226 < Q_4 < .279$$
$$.094 < Q_5 < .149$$

PART V

SUBDIVISION OF THE PLANE INTO SEVERAL EQUI-VULNERABILITY AREAS[1]

In parts I through IV we have considered the probability that a plane will be downed by a hit without any reference to the part of the plane that receives the hit. Undoubtedly, the probability of downing a plane by a hit will depend considerably upon the part that receives the hit. The purpose of this memorandum is to extend the previous results to the more general case where the probability of downing a plane by a hit depends on the part of the plane sustaining the hit. To carry out this generalization of the theory, we shall subdivide the plane into k equi-vulnerability areas $A_1,\ldots,A_k$. For any set of non-negative integers $i_1,\ldots,i_k$ let $P(i_1,\ldots,i_k)$ be the probability that a plane will be downed if the area A_1 receives i_1 hits, the area A_2 receives i_2 hits,..., and the area A_k receives i_k hits. Let $Q(i_1,\ldots,i_k) = 1 - P(i_1,\ldots,i_k)$. Then $Q(i_1,\ldots,i_k)$ is the probability that the plane will not be downed if the areas $A_1,\ldots,A_k$ receive $i_1,\ldots,i_k$ hits, respectively. We shall assume that $Q(i_1,\ldots,i_k)$ is a symmetric function of the arguments $i_1,\ldots,i_k$.

To estimate the value of $Q(i_1,\ldots,i_k)$ from the damage to returning planes, we need to know the probability distribution of hits over the k areas $A_1,\ldots,A_k$ knowing merely the total number of hits received. In other words, for any positive integer i we need to know the conditional probability $\gamma_1(i_1,\ldots,i_k)$ that the areas $A_1,\ldots,A_k$ will receive $i_1,\ldots,i_k$ hits, respectively, knowing that the total number of hits is i. Of course, $\gamma_i(i_1,\ldots,i_k)$ is defined only for values $i_1,\ldots,i_k$ for which $i_1 + \ldots + i_k = i$. To avoid confusion, it should be emphasized that the probability $\gamma_i(i_1,\ldots,i_k)$ is determined under the

[1]This part of "A Method of Estimating Plane Vulnerability Based on Damage of Survivors" was published as SRG memo 96 and AMP memo 76.5.

assumption that dummy bullets are used. It can easily be shown that it is impossible to estimate both $\gamma_i(i_1,\dots,i_k)$ and $Q(i_1,\dots,i_k)$ from the damage to returning planes only. To see this, assume that k is equal to 2 and all hits on the returning planes were located in the area A_1. This fact could be explained in two different ways. One explanation could be that $\gamma_i(i_1,i_2) = 0$ for $i_2 > 0$. The other possible explanation would be that $Q(i_1,i_2) = 0$ for $i_2 > 0$. Hence, it is impossible to estimate both $\gamma_i(i_1,i_2)$ and $Q(i_1,i_2)$. Fortunately, $\gamma_i(i_1,\dots,i_k)$ can be assumed to be known a priori (on the basis of the dispersion of the guns), or can be established experimentally by firing with dummy bullets and recording the hits scored. Thus, in what follows we shall assume that $\gamma_i(i_1,\dots,i_k)$ is known for any set of integers $i_1,\dots,i_k$.

Clearly, the probability that i hits will not down the plane is given by

$$Q_i = \sum_{i_k} \dots \sum_{i_1} \gamma_i(i_1,\dots,i_k)Q(i_1,\dots,i_k), \tag{69}$$

where the summation is to be taken over all non-negative integers $i_1,\dots,i_k$ for which $i_1 + \dots + i_k = i$.

Let $\delta_i\ (i_1,\dots,i_k)$ be the conditional probability that the areas $A_1,\dots,A_k$ received $i_1,\dots,i_k$ hits, respectively, knowing that the plane received i hits and that the plane was not downed. Then we have

$$\delta_i(i_1,\dots,i_k) = \frac{\gamma_i(i_1,\dots,i_k)Q(i_1,\dots,i_k)}{Q_i} \quad . \tag{70}$$

Of course, $\delta_i(i_1,\dots,i_k)$ is defined only for non-negative integers $i_1,\dots,i_k$ for which $i_1 + \dots + i_k = i$.

The probability $\delta_i(i_1,\ldots,i_k)$ can be determined from the distribution of hits on returning planes. In fact, let $a(i_1,\ldots,i_k)$ be the proportion of planes (out of the total number of planes participating in combat) that returned with i_1 hits on area A_1, i_2 hits on area $A_2,\ldots,$ and i_k hits on area A_k. Then we obviously have

$$\delta_i(i_1,\ldots,i_k) = \frac{a(i_1,\ldots,i_k)}{a_i} \quad . \tag{71}$$

From equations 70 and 71, we obtain

$$Q(i_1,\ldots,i_k) = \frac{Q_i a(i_1,\ldots,i_k)}{a_i \gamma_i(i_1,\ldots,i_k)} \qquad (i = i_1 + \ldots + i_k) \tag{72}$$

Since Q_i can be estimated by methods described in parts I through IV, estimates of $Q(i_1,\ldots,i_k)$ can be obtained from equation 72.

According to equation 29, the probabilities $Q_1,\ldots,Q_n$ satisfy the equation

$$\sum_{j=1}^{n} \frac{a_j}{Q_j} = 1 - a_o \quad . \tag{73}$$

We have assumed that $q_1 \geq q_2 \geq \ldots \geq q_n$. This is equivalent to stating that

$$\frac{Q_{i+1}}{Q_i} \leq \frac{Q_{j+1}}{Q_j} \qquad \text{for } j \leq i \quad . \tag{74}$$

A similar assumption can be made with respect to the probabilities $Q(i_1,\ldots,i_k)$. In fact, the conditional probability that an additional hit on the area A_r will not down the plane knowing that the areas $A_1,\ldots,A_k$ have already sustained $i_1,\ldots,i_k$ hits, respectively, is given by

$$\frac{Q(i_1,\ldots,i_{r-1},\ i_r+1,\ i_{r+1},\ldots,i_k)}{Q(i_1,\ldots,i_{r-1},\ i_r,\ i_{r+1},\ldots,i_k)} \quad . \tag{75}$$

Obviously, we can assume that if

$$j_1 \leq i_1,\ j_2 \leq i_2,\ldots,j_k \leq i_k$$

then

$$\frac{Q(i_1,\ldots,i_{r-1},i_r+1,i_{r+1},\ldots,i_k)}{Q(i_1,\ldots,i_{r-1},i_r,i_{r+1},\ldots,i_k)} \leq \frac{Q(j_1,\ldots,j_{r-1},j_r+1,j_{r+1},\ldots,j_k)}{Q(j_1,\ldots,j_{r-1},j_r,j_{r+1},\ldots,j_k)} \tag{76}$$

for $r = 1,2,\ldots,k$.

Hence, the possible values of $Q_1,\ldots,Q_n$ are restricted to those for which equation 73 is fulfilled and for which the quantities $Q(i_1,\ldots,i_k)$ computed from equation 72 are less than or equal to one and satisfy the inequalities of equation 76. It should be remarked that the inequalities of equation 76 do not follow from the inequalities of equation 74. From equation 72 and the inequality $Q(i_1,\ldots,i_k) \leq 1$, it follows that

$$Q_i \leq \frac{a_i\,\gamma_i(i_1,\ldots,i_k)}{a(i_1,\ldots,i_k)} \quad . \tag{77}$$

If the right-hand side expression in equation 77 happens to be less than one, then equation 77 imposes a restriction on Q_i. Since

$$\sum_{i_k} \cdots \sum_{i_1} \frac{a(i_1,\ldots,i_k)}{a_i} = \sum_{i_k} \cdots \sum_{i_1} \gamma_i(i_1,\ldots,i_k) = 1$$

(the summation is taken over all values $i_1,\ldots,i_k$ for which $i_1+ \ldots + i_k = i$), we must have either

$$\frac{a_i \gamma_i(i_1,\ldots,i_k)}{a(i_1,\ldots,i_k)} = 1$$

for all values $i_1,\ldots,i_k$ for which $i_1 + \ldots + i_k = i$, or

$$\frac{a_i \gamma_i(i_1,\ldots,i_k)}{a(i_1,\ldots,i_k)} < 1$$

at least for one set of values $i_1,\ldots,i_k$ satisfying the condition $i_1 + \ldots + i_k = i$. Hence, equation 77 gives an upper bound for Q_i whenever there exists a set of integers $i_1,\ldots,i_k$ such that $i_1 + \ldots + i_k = i$ and

$$\frac{a(i_1,\ldots,i_k)}{a_i} \neq \gamma_i \ (i_1,\ldots,i_k) \ .$$

It is of interest to investigate the case of independence, i.e., the case when the probability that an additional hit will not down the plane does not depend on the number and distribution of hits already received. Denote by $q(i)$ the probability that a single hit on the area A_i will not down the plane. Then under the assumption of independence we have

$$Q(i_1,\ldots,i_k) = [q(1)]^{i_1} \ [q(2)]^{i_2} \ \ldots \ [q(k)]^{i_k} \ . \qquad (78)$$

Hence, the only unknown probabilities are $q(1),\ldots,q(k)$.

Let $\gamma(i)$ be the conditional probability that the area A_i is hit knowing that the plane received exactly one hit. Obviously

$$\gamma_i(i_1,\ldots,i_k) = \frac{i!}{i_1! \ \ldots \ i_k!} \ [\gamma(1)]^{i_1} \ \ldots \ [\gamma(k)]^{i_k} \ . \qquad (79)$$

Similarly, let $\delta(i)$ be the conditional probability that the area A_i is hit knowing that the plane received exactly one hit and this hit did not down the plane. Because of the assumption of independence, we have

$$\delta_i(i_1,\ldots,i_k) = \frac{i!}{i_1! \ldots i_k!} [\delta(1)]^{i_1} \ldots [\delta(k)]^{i_k} . \qquad (80)$$

Furthermore, we have

$$\delta(i) = \frac{\gamma(i)q(i)}{\sum_{i=1}^{k} \gamma(i)q(i)} . \qquad (81)$$

Since the probability q that a single hit does not down the plane is equal to $\sum_{i=1}^{k} \gamma(i)q(i)$, we obtain from equation 81

$$q(i) = \frac{\delta(i)}{\gamma(i)} q . \qquad (82)$$

Because of the assumption of independence, we see that $\delta(i)$ is equal to the ratio of the total number of hits in the area A_i of the returning planes to the total number of hits received by the returning planes. That is

$$\delta(i) = \frac{\sum_{j_k} \cdots \sum_{j_1} j_i a(j_1,\ldots,j_k)}{\sum_{j_k} \cdots \sum_{j_1} (j_1 + \ldots + j_k)a(j_1,\ldots,j_k)} . \qquad (83)$$

Since $\gamma(i)$ is assumed to be known and since $\delta(i)$ can be computed from equation 83, we see from equation 82 that $q(i)$ can be determined as soon as the value of q is known. The value of q can be obtained by solving the equation

$$\sum_{j=1}^{n} \frac{a_j}{q^j} = 1 - a_o . \qquad (84)$$

NUMERICAL EXAMPLE

In the examples for parts I, III, and IV we have estimated the probability that a plane will be downed without reference to the part of the plane that receives the hit. However, the vulnerability of a particular part (say the motors) may be of interest and this example illustrates the methods of estimating part vulnerabilities under the following assumptions:

- The number of planes participating in combat is large so that sampling errors can be neglected.

- The probability that a hit will down the plane does not depend on the number of previous non-destructive hits. That is, $q_1 = q_2 = \cdots = q_n = q_o$.

- Given that a shot has hit the plane, the probability that it hit a particular part is assumed to be known. In this example it is put equal to the ratio of the area of this part to the total surface area of the plane.[1]

- The division of the plane into several parts is representative of all the planes of the mission. If the types of planes are radically different so that no representative division is possible, we may consider the different classes of planes separately.

Consider the following example. Of 400 planes on a bombing mission, 359 return. Of these, 240 were not hit, 68 had one hit, 29 had two hits, 12 had three hits, and 10 had four hits. Following the example in part I we have

$N = 400,$

whence

$$
\begin{aligned}
A_o &= 240 & a_o &= .600 \\
A_1 &= 68 & a_1 &= .170 \\
A_2 &= 29 & a_2 &= .072 \\
A_3 &= 12 & a_3 &= .030 \\
A_4 &= 10 & a_4 &= .025
\end{aligned}
$$

[1]By area is meant here the component of the area perpendicular to the direction of the enemy attack. If this direction varies during the combat, some proper average direction may be taken.

As before, the probability that a single hit will not down the plane is given by the root of

$$\frac{a_1}{q_o} + \frac{a_2}{q_o^2} + \frac{a_3}{q_o^3} + \frac{a_4}{q_o^4} = 1 - a_o ,$$

which reduces to

$$.4q_o^4 - .170q_o^3 - .072q_o^2 - .030q_o - .025 = 0$$

and

$$q_o = .850.$$

Suppose that we are interested in estimating the vulnerability of the engines, the fuselage, and the fuel system. Assume that the following data is representative of all the planes of the mission:

Part number	Description	Area of part	Ratio of area of part to total area ($\gamma(i)$)
1	2 engines	35 sq. ft.	$\frac{35}{130} = .269$
2	Fuselage	45 sq. ft.	$\frac{45}{130} = .346$
3	Fuel system	20 sq. ft.	$\frac{20}{130} = .154$
4	All other parts	30 sq. ft.	$\frac{30}{130} = .231$
	Total area	130 sq. ft.	

The ratio of the area of the i-th part to the total area is designated $\gamma(i)$. Given that the plane is hit, by the third assumption, $\gamma(i)$ is the probability that this hit occurred on part i. Thus

$\gamma(1) = .269$
$\gamma(2) = .346$
$\gamma(3) = .154$
$\gamma(4) = .231$

The only additional information we require is the number of hits on each part. Let the observed number of hits be 202. In general, the total number of hits (on returning planes) must be equal to

$$A_i + 2A_2 + 3A_3 + \ldots + nA_n$$

and in this example

$$A_1 + 2A_2 + 3A_3 + 4A_4 = 68 + 2(29) + 3(12) + 4(10) = 202$$

The hits on the returning planes were distributed as follows:

Part number	Number of hits observed on part	Ratio of number of hits observed on part to total number of observed hits ($\delta(i)$)
1	39	.193
2	78	.386
3	31	.154
4	54	.267
Total number of hits	202	

The ratio of the number of hits on part i to the total number of hits on surviving planes is designated $\delta(i)$. Then $q(i)$, the probability that a hit on the i-th part does not down the plane, is given by

$$q(i) = \frac{\delta(i)}{\gamma(i)} q_o \quad ,$$

whence

$$q(1) = \frac{\delta(1)}{\gamma(1)} q_o = \frac{.193}{.269} (.850) = .61$$

$$q(2) = \frac{\delta(2)}{\gamma(2)} q_o = \frac{.386}{.346} (.850) = .95$$

$$q(3) = \frac{\delta(3)}{\gamma(3)} q_o = \frac{.154}{.154} (.850) = .85$$

$$q(4) = \frac{\delta(4)}{\gamma(4)} q_o = \frac{.267}{.231} (.850) = .98$$

The results may be summarized as follows:

Part	Probability of surviving a single hit (q(i))	Probability of being downed by a single hit (1 - q(i))
Entire plane	.85	.15
Engines	.61	.39
Fuselage	.95	.05
Fuel system	.85	.15
Other parts	.98	.02

Thus, for the observed data of this hypothetical example, the engine area is the most vulnerable in the sense that a hit there is most likely to down the plane. The fuselage has a relatively low vulnerability.

PART VI

SAMPLING ERRORS[1]

In parts I through V we have assumed that the total number of planes participating in combat is so large that sampling errors can be neglected altogether. However, in practice N is not excessively large and therefore it is desirable to take sampling errors into account. We shall deal here with the case when $q_1 = q_2 \ldots = q_n = q$ (say) and we shall derive confidence limits for the unknown probability q.

If there were no sampling errors, then we would have

$$x_i = p(1 - a_o - a_1 - \ldots - a_{i-1} - x_1 - x_2 - \ldots - x_{i-1}) \qquad (i = 2,3,\ldots), \tag{85}$$

where $p = 1 - q$. However, because of sampling errors we shall have the equation

$$x_i = \bar{p}_i(1 - a_o - \ldots - a_{i-1} - x_1 - \ldots - x_{i-1}), \tag{86}$$

where $\bar{p}_i$ is distributed like the success ratio in a sequence of $N_i = N(1 - a_o - a_1 - \ldots - a_{i-1} - x_1 - \ldots - x_{i-1})$ independent trials, the probability of success in a single trial being equal to p.

Let $\bar{q}_i = 1 - \bar{p}_i$. Then, according to equation 26 we have

$$\sum_{j=1}^{n} \frac{a_j}{\bar{q}_1 \ldots \bar{q}_j} = 1 - a_o, \tag{87}$$

[1]This part of "A Method of Estimating Plane Vulnerability Based on Damage of Survivors" was published as SRG memo 103 and AMP memo 76.6.

provided that $x_i = 0$ for $i > n$. In part I we have shown that $x_i = 0$ for $i > n$ if there are no sampling errors. This is not necessarily true if sampling errors are taken into account. However, in the case of independence, i.e., when $q_i = q$ $(i = 1,2,\ldots)$, x_i is very small for $i > n$ so that $\sum_{i=n+1}^{\infty} x_i$ can be neglected. In fact, if the number of planes that received more than n hits were not negligibly small, it follows from the assumption of independence that the probability is very high that at least some of these planes would return. Since no plane returned with more than n hits, for practical purposes we may assume that $\sum_{i=n+1}^{\infty} x_i = 0$. In what follows we shall make this assumption.

Each of the quantities $\bar{q}_1,\ldots,\bar{q}_n$ can be considered as a sample estimate of the unknown probability q. However, the quantities $\bar{q}_1,\ldots,\bar{q}_n$ are unknown. It is merely known that they satisfy the relation in equation 87. Confidence limits for q may be derived on the basis of equation 87. However, we shall use another more direct approach.

To derive confidence limits for the unknown probability q we shall consider the hypothetical proportion b_i of planes that would have been hit exactly i times if dummy bullets would have been used. We shall treat the quantities $b_1,\ldots,b_k$ as fixed (but unknown) constants. This assumption does not involve any loss of generality, since the confidence limits for q obtained on the basis of this assumption remain valid also when $b_1,\ldots,b_k$ are random variables. Clearly, the probability distribution of Na_i $(i = 1,\ldots,n)$ is the same as the distribution of the number of successes in a sequence of Nb_i independent trials, the probability of success in a single trial being q^i. Hence

$$E(Na_i) = q^i Nb_i \qquad (88)$$

$$\sigma^2\,(Na_i) = Nb_i q^i(1 - q^i)\ . \qquad (89)$$

From equations 88 and 89 we obtain

$$E\left(\frac{a_i}{q^i}\right) = b_i \tag{90}$$

$$\sigma^2\left(\frac{a_i}{q^i}\right) = \frac{b_i(1 - q^i)}{Nq^i} \quad . \tag{91}$$

Since the variates $\frac{a_1}{q}, \frac{a_2}{q^2}, \ldots, \frac{a_n}{q^n}$ are independently distributed, and since a_i is nearly normally distributed if N is not small, we can assume with very good approximation that the sum

$$\sum_{i=1}^{n} \frac{a_i}{q^i} \tag{92}$$

is normally distributed. We obtain from equations 90 and 91

$$E\left(\sum_{i=1}^{n} \frac{a_i}{q^i}\right) = \sum_{i=1}^{n} b_i = 1 - a_o \tag{93}$$

$$\sigma^2\left(\sum_{i=1}^{n} \frac{a_i}{q^i}\right) = \sum_{i=1}^{n} \frac{b_i(1 - q^i)}{Nq^i} \quad . \tag{94}$$

For any positive $\alpha < 1$ let λ_α be the value for which

$$\int_{-\lambda_\alpha}^{\lambda_\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \alpha$$

The set of all values q for which the inequality

$$1 - a_o - \lambda_\alpha \sqrt{\sum_{i=1}^{n} \frac{b_i(1 - q^i)}{Nq^i}} \leq \sum_{i=1}^{n} \frac{a_i}{q^i} \leq 1 - a_o + \lambda_\alpha \sqrt{\sum_{i=1}^{n} \frac{b_i(1 - q^i)}{Nq^i}} \tag{95}$$

is fulfilled forms a confidence set for the unknown probability q with confidence coefficient α. However, formula 95 cannot be used, since it involves the unknown quantities $b, \ldots, b_n$. Since $\frac{a_i}{q^i}$ converges stochastically to b_i as $N \to \infty$, we change the standard deviation of $\sum \frac{a_i}{q^i}$ only by a quantity of order less than $\frac{1}{\sqrt{N}}$ if we replace b_i by $\frac{a_i}{q^i}$. Thus, the set of values q that satisfy the inequalities

$$1 - a_o - \lambda_\alpha \sqrt{\sum_{i=1}^{n} \frac{a_i(1 - q^i)}{Nq^{2i}}} \leq \sum_{i=1}^{n} \frac{a_i}{q^i} \leq 1 - a_o + \lambda_\alpha \sqrt{\sum_{i=1}^{n} \frac{a_i(1 - q^i)}{Nq^{2i}}} \quad (96)$$

is an approximation to a confidence set with confidence coefficient α.

Denote by q_o the root of the equation in q

$$\sum_{1}^{n} \frac{a_j}{q^j} = 1 - a_o .$$

Then q_o converges stochastically to q as $N \to \infty$. A considerable simplification can be achieved in the computation of the confidence set by substituting q_o for q in the expression of the standard deviation of $\sum \frac{a_i}{q^i}$. The error introduced by this substitution is small if N is large. Making this substitution, the inequalities defining the confidence set are given by

$$1 - a_o - \lambda_\alpha \sqrt{\sum_{i=1}^{n} \frac{a_i(1 - q_o^i)}{Nq_o^{2i}}} \leq \sum_{i=1}^{n} \frac{a_i}{q^i} \leq 1 - a_o + \lambda_\alpha \sqrt{\sum_{i=1}^{n} \frac{a_i(1 - q_o^i)}{Nq_o^{2i}}} \quad (97)$$

Hence, the confidence set is an interval. The upper end point of the confidence interval is the root of the equation

$$\sum_{i=1}^{n} \frac{a_i}{q^i} = 1 - a_o - \lambda_\alpha \sqrt{\sum_{i=1}^{n} \frac{a_i(1 - q_o^i)}{Nq_o^{2i}}} \qquad (98)$$

and the lower end point of the confidence interval is the root of the equation

$$\sum_{i=1}^{n} \frac{a_i}{q^i} = 1 - a_o + \lambda_\alpha \sqrt{\sum_{i=1}^{n} \frac{a_i(1 - q_o^i)}{Nq_o^{2i}}} \qquad (99)$$

NUMERICAL EXAMPLE

In all previous examples it was assumed that A_i (the number of planes returning with i hits) was compiled from such a large number of observations that they were not subject to sampling errors. If it is further assumed that the probability q that a hit will down a plane does not depend on the number of previous non-destructive hits, it is possible to obtain an exact solution for the probability that a hit will down a plane. Here we introduce the possibility that the $A_o,\ldots,A_n$ are subject to sampling errors but retain the assumption of independence. Under these less restrictive assumptions we cannot obtain the exact solution for q, but for any positive number $\alpha < 1$ we can construct two functions of the data, called confidence limits, such that the statement that q lies between the confidence limits will be true 100α percent of the time in the long run. The confidence limits are calculated for $\alpha = .95$ and .99.

Under the assumptions of part I, it was proved that no planes received more hits than the greatest number of hits observed on a returning plane. This is not necessarily true when the possibility of sampling error is introduced, but it is retained as an assumption, since the error involved is small.

If the a_i are subject to sampling error, and q is the true parameter,

$$\sum_{1}^{n} \frac{a_i}{q^i} \qquad \text{(A)}$$

will be approximately normally distributed with mean value $1 - a_o$.

In outlining the steps necessary to calculate the confidence limits, the following hypothetical set of data will be used. Given

$N = 500 \qquad a_i = \frac{A_i}{N}$

$A_o = 400$	$a_o = .80$
$A_1 = 40$	$a_1 = .08$
$A_2 = 25$	$a_2 = .05$
$A_3 = 5$	$a_3 = .01$
$A_4 = 3$	$a_4 = .006$
$A_5 = 2$	$a_5 = .004$
475	

The first step is to find the value q_o, for which expression A is equal to its mean value, by finding the positive root of

$$\frac{a_1}{q} + \frac{a_2}{q^2} + \frac{a_3}{q^3} + \frac{a_4}{q^4} + \frac{a_5}{q^5} = 1 - a_o \ .$$

We obtain

$$.20q^5 - .08q^4 - .05q^3 - .01q^2 - .006q - .004 = 0$$

$$q_o = .850 \ .$$

The next step is to calculate the standard deviation of expression A. This can be shown to be approximately equal to

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{a_i(1 - q_o^i)}{Nq_o^{2i}}}$$

$$= \sqrt{\frac{a_1(1-q_o^1)}{Nq_o^2} + \frac{a_2(1-q_o^2)}{Nq_o^4} + \frac{a_3(1-q_o^3)}{Nq_o^6} + \frac{a_4(1-q_o^4)}{Nq_o^8} + \frac{a_5(1-q_o^5)}{Nq_o^{10}}}$$

$$= .01226 \; .$$

Knowing that $\sum_{i=1}^{n} \frac{a_i}{q^i}$ is approximately normally distributed with mean value $1 - a_o$ and the standard deviation σ, we can determine the range in which $\sum \frac{a_i}{q^i}$ can be expected to be 100α percent of the time (say 95 and 99 percent) by determining $\lambda_{.95}$ and $\lambda_{.99}$ such that

$$\frac{1}{\sqrt{2\pi}} \int_{-\lambda_{.95}}^{\lambda_{.95}} \exp\left(-\frac{t^2}{2}\right) dt = .95$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\lambda_{.99}}^{\lambda_{.99}} \exp\left(-\frac{t^2}{2}\right) dt = .99 \; .$$

From the table or the areas of a normal curve, it is found that

$$\lambda_{.95} = 1.959964$$
$$\lambda_{.99} = 2.575829 \; .$$

We can now calculate the confidence limits for each value of α by finding the two values of q for which the equality sign of the following expression holds:

$$\left| \sum_{i=1}^{n} \frac{a_i}{q^i} - (1 - a_o) \right| \leq \lambda_\alpha \sigma \quad .$$

It follows that for each α, the confidence limits are the positive roots of the equation

$$\sum_{i=1}^{n} \frac{a_i}{q^i} = 1 - a_o \pm \lambda_\alpha \sigma$$

α	λ_α	$.0122678\lambda_\alpha$	$1 - a_o - \lambda_\alpha \sigma$	$1 - a_o + \lambda_\alpha \sigma$
.95	1.959964	.024044	.175956	.224044
.99	2.575829	.031600	.168400	.231600

For α = .95 the confidence limits of q_o are the positive roots of equation

$$\frac{a_1}{q} + \frac{a_2}{q^2} + \frac{a_3}{q^3} + \frac{a_4}{q^4} + \frac{a_5}{q^5} = .175956,$$

which reduces to

$$.175956q^5 - .08q^4 - .05q^3 - .01q^2 - .006q - .004 = 0$$

$$q = .912,$$

and equation

$$\frac{a_1}{q} + \frac{a_2}{q^2} + \frac{a_3}{q^3} + \frac{a_4}{q^4} + \frac{a_5}{q^5} = .224044,$$

which reduces to

$$.224044q^5 - .08q^4 - .05q^3 - .01q^2 - .006q - .004 = 0$$

$$q = .801.$$

Similarly, for α = .99 we have

$$.168400q^5 - .08q^4 - .05q^3 - .01q^2 - .006q - .004 = 0$$

$$q = .935$$

$$.231600q^5 - .08q^4 - .05q^3 - .01q^2 - .006q - .004 = 0$$

$$q = .787 .$$

Summarizing the results we find that the 95-percent confidence limits of q are .801 and .912, and that the 99-percent confidence limits are .787 and .935.

PART VII

MISCELLANEOUS REMARKS[1]

1. Factors that may vary from combat to combat but influence the probability of surviving a hit. The factors that influence the probability of surviving a hit may be classified into two groups. The first group contains those factors that do not vary from combat to combat. This does not necessarily mean that the factor in question has a fixed value of all combats; the factor may be a random variable whose probability distribution does not vary from combat to combat. The second group comprises those factors whose probability distribution cannot be assumed to be the same for all combats. To make predictions as to the proportions of planes that will be downed in future combats, it is necessary to study the dependence of the probability q of surviving a hit on the factors in the second group. In part V we have already taken into account such a factor. In part V we have considered a subdivision of the plane into several equi-vulnerability areas $A_1,\dots,A_k$ and we expressed the probability of survival as a function of the part of the plane that received the hit. Since the probability of hitting a certain part of the plane depends on the angle of attack, this probability may vary from combat to combat. Thus, it is desirable to study the dependence of the probability of survival on the part of the plane that received the hit. In addition to the factors represented by the different parts of the plane, there may also be other factors, such as the type of gun used by the enemy, etc., which belong to the second group. There are no theoretical difficulties whatsoever in extending the theory in part V to any number and type of factors. To illustrate this, let us assume that the factors to be taken into account are the different parts $A_1,\dots,A_k$ of the plane and the different guns $g_1,\dots,g_m$ used by the enemy. Let $q(i,j)$ be the probability of surviving a hit on part A_1 knowing that the bullet has been fired by gun g_j. We may order the km pairs (i,j) in a sequence. We shall denote $q(i,j)$ by $q(u)$ if the pair (i,j) is the u-th element in the ordered sequence of pairs. The problem of determining the unknown probabilities $q(u)$ $(u = 1,\dots,km)$ can be treated in exactly the same way as the problem discussed in

[1]This part of "A Method of Estimating Plane Vulnerability Based on Damage of Survivors" was published as SRG memo 109 and AMP memo 76.7.

part V assuming that the plane consists of km parts. Any hit on part A_i by a bullet from gun g_j can be considered as a hit on part A_u in the problem discussed in part V where (i,j) is the u-th element in the ordered sequence of pairs.

2. Non-probabilistic interpretation of the results. It is interesting to note that a purely arithmetic interpretation of the results of parts I through V can be given. Instead of defining q_i as the probability of surviving the i-th hit knowing that the previous i - 1 hits did not down the plane, we define q_i as follows: Let M_i be the number of planes that received at least i hits and the i-th hit did not down the plane, and let N_i be the total number of planes that received at least i hits. Then $q_i = \frac{M_i}{N_i}$. Thus, q_i is defined in terms of what actually happened in the particular combat under consideration. To distinguish this definition of q_i from the probabilistic definition, we shall denote the ratio $\frac{M_i}{N_i}$ by $\bar{q}_i$. The quantity $\bar{q}$ is unknown, since we do not know the distribution of hits on the planes that did not return. However, it follows from the results of part I that these quantities must satisfy equation 26. If we can assume that in the particular combat under consideration we have $\bar{q}_i = \ldots = \bar{q}_n$ then the common value $\bar{q}$ of these quantities is the root of the equation

$$\sum \frac{a_j}{\bar{q}^j} = 1 - a_o \quad .$$

Assuming that $\bar{q}_1 \geq \bar{q}_2 \geq \ldots \geq \bar{q}_n$, the minimum value Q_i^o of Q_i derived in parts III and IV can be interpreted as the minimum value of $\bar{Q}_i = \bar{q}_1 \ldots \bar{q}_i$.

The minimum and maximum values of Q_i derived in part IV can also be interpreted as minimum and maximum values of $\bar{Q}_i = \bar{q}_1 \ldots \bar{q}_i$ if we assume that the inequalities $\lambda_1 \bar{q}_j \leq \bar{q}_{j+1} \leq \lambda_2 \bar{q}_j$ $(j = 1, \ldots, n-1)$ are fulfilled. Similarly, a pure arithmetic interpretation of the results of part V can be given.

3. <u>The case when $\gamma(i)$ is unknown.</u> In part V we have assumed that the probabilities $\gamma(1),\ldots,\gamma(k)$ are known. Since the exposed areas of the different parts $A_1,\ldots,A_k$ depend on the angle of attack, and since this angle may vary during the combat, it may sometimes be difficult to estimate the probabilities $\gamma(1),\ldots,\gamma(k)$. Thus, it may be of interest to investigate the question whether any inference as to the probabilities $q(1),\ldots,q(k)$ can be drawn when $\gamma(1),\ldots,\gamma(k)$ are entirely unknown. We shall see that frequently a useful lower bound for $q(i)$ can still be obtained. In fact, the value $q^*(i)$ of $q(i)$, calculated under the assumption that the parts A_j $(j \neq i)$ are not vulnerable $(q(j) = 1)$, is certainly a lower bound of the true value $q(i)$. Considering only the hits on part A_i, a lower bound of $q^*(i)$, and therefore also of $q(i)$, is given by the root of the equation

$$\sum_{r=1}^{n} \frac{a_r^*}{q^r} = 1 - a_o^* \quad , \qquad (100)$$

where a_r^* $(r = 0,1,\ldots,n)$ is the ratio of the number of planes returned with exactly r hits on part A_i to the total number of planes participating in combat.

The lower limit obtained from equation 100 will be a useful one if it is not near zero. The root of equation 100 will be considerably above zero if $\sum_{r=1}^{n} a_r^*$ is not very small as compared with $1 - a_o^*$. This can be expected to happen whenever both $\gamma(i)$ and $q(i)$ are considerably above zero.

PART VIII

VULNERABILITY OF A PLANE TO DIFFERENT TYPES OF GUNS[1]

In part V we discussed the case where the plane is subdivided into several equi-vulnerability areas (parts) and we dealt with the problem of determining the vulnerability of each of these parts. It was pointed out in part VII that the method described in part V can be applied to the more general problem of estimating the probability $q(i,j)$ that a plane will survive a hit on part i caused by a bullet fired from gun j. However, this method is based on the assumption that the value of $\gamma(i,j)$ is known where $\gamma(i,j)$ is the conditional probability that part i is hit by gun j knowing that a hit has been scored. In practice it may be difficult to determine the value of $\gamma(i,j)$ since the proportions in which the different guns are used by the enemy may be unknown. On the other hand, it seems likely that frequently we shall be able to estimate the conditional probability $\gamma(i|j)$ that part i is hit knowing that a hit has been scored by gun j. The purpose of this memorandum is to investigate the question whether $q(i,j)$ can be estimated from the data assuming that merely the quantities $\gamma(i|j)$ are known a priori. In what follows we shall restrict ourselves to the case of independence, i.e., it will be assumed that the probability of surviving a hit does not depend on the non-destructive hits already received.

Let $\delta(i,j)$ be the conditional probability that part i is hit by gun j knowing that a hit has been scored and the plane survived the hit. Furthermore, let q be the probability that the plane survives a hit (not knowing which part was hit and which gun scored the hit). Then, similar to equation 82, we shall have

$$q(i,j) = \frac{\delta(i,j)}{\gamma(i,j)}\, q \; . \tag{101}$$

Let $q(j)$ be the probability that the plane will survive a hit by gun j (not knowing the part hit). Then obviously

$$q(j) = \sum_{1} \gamma\,(i|j)q(i,j) \quad . \tag{102}$$

Let $\delta(i|j)$ be the conditional probability that part i is hit by gun j knowing that a hit has been scored by gun j and the plane survived the hit. Clearly

[1]This part of "A Method of Estimating Plane Vulnerability Based on Damage of Survivors" was published as SRG memo 126 and AMP memo 76.8.

$$\delta(i|j) = \frac{\gamma(i|j)q(i,j)}{\sum_i \gamma(i|j)q(i,j)} = \frac{\gamma(i|j)q(i,j)}{q(j)} \quad . \tag{103}$$

From equation 103, we obtain

$$q(i,j) = \frac{\delta(i|j)}{\gamma(i|j)} q(j) \quad . \tag{104}$$

The quantity $\delta(i|j)$ can be estimated on the basis of the observed hits on the returning planes. The best sample estimate of $\delta(i|j)$ is the ratio of the number of hits scored by gun j on part i of the returning planes to the total number of hits scored by gun j on the returning planes. Thus, on the basis of equation 104, the probability $q(i,j)$ can be determined if $q(j)$ is known.

Now we shall investigate the question whether $q(j)$ can be estimated. First, we shall consider the case when it is known a priori that a certain part of the plane, say part 1, is not vulnerable. Then $q(i,j) = 1$ and we obtain from equation 104

$$1 = \frac{\delta(1|j)}{\gamma(1|j)} q(j) \quad . \tag{105}$$

Hence,

$$q(j) = \frac{\gamma(1|j)}{\delta(1|j)} \quad . \tag{106}$$

Thus, in this case our problem is solved. If no part of the plane can be assumed to be invulnerable, then we can still obtain upper limits for $q(j)$. In fact, since $q(i,j) \leq 1$, we obtain from equation 104

$$q(j) \leq \frac{\gamma(i|j)}{\delta(i|j)} \quad . \tag{107}$$

Denote by $\rho(j)$ the minimum of $\frac{\gamma(i|j)}{\delta(i|j)}$ with respect to 1. Then we have

$$q(j) \leq \rho(j) \quad . \tag{108}$$

If there is a part of the airplane that is only slightly vulnerable (this is usually the case), then $q(j)$ will not be much below $\rho(j)$. Let the part i_j be the part of the plane least vulnerable to gun j. If $q(i_j, j)$ has the same value for any gun j, then $q(j)$ is proportional to $\rho(j)$. Thus, the error is perhaps not serious if we assume that $q(j)$ is proportional to $\rho(j)$, i.e.,

$$q(j) = \lambda\rho(j). \tag{109}$$

The proportionality factor λ can be determined as follows. From equations 101 and 104 we obtain

$$\frac{\delta(i,j)}{\gamma(i,j)}\, q = \lambda\rho(j)\, \frac{\delta(i|j)}{\gamma(i|j)} \quad . \tag{110}$$

Hence,

$$\lambda\gamma(i,j) = q\, \frac{\delta(i,j)\gamma(i|j)}{\delta(i|j)\rho(j)} \quad . \tag{111}$$

Denote $\sum_i \delta(i,j)$ by $\delta(j)$. Then,

$$\delta(i|j) = \frac{\delta(i,j)}{\delta(j)} \; . \tag{112}$$

From equations 111 and 112 we obtain

$$\lambda\gamma(i,j) = q\, \frac{\delta(j)\gamma(i|j)}{\rho(j)} \; . \tag{113}$$

Since

$$\sum_i \gamma(i|j) = 1,$$

we obtain from equation 113

$$\lambda \sum_j \sum_i \gamma(i,j) = q \sum_j \frac{\delta(j)}{\rho(j)} \quad . \tag{114}$$

But

$$\sum_j \sum_i \gamma(i,j) = 1.$$

Hence,

$$\lambda = q \sum_j \frac{\delta(j)}{\rho(j)} \quad . \tag{115}$$

Since $\delta(j)$ and $\rho(j)$ are known quantities, the proportionality factor λ can be obtained from equation 115. The probability q is the root of the equation

$$\sum_{j=1}^{n} \frac{a_j}{q^j} = 1 - a_o,$$

where a_j denotes the ratio of the number of planes returned with exactly j hits to the total number of planes participating in combat.

NUMERICAL EXAMPLE

In part V, the case of a plane subdivided into several equi-vulnerability areas was discussed, and the vulnerability of each part was estimated. The same method can be extended to solve the more general problem of estimating the probability that a plane will survive a hit on part i caused by a bullet fired from gun j, if assumptions corresponding to those of part V are made. The first three of the four assumptions that must be made to apply the method of part V directly are identical with those made in part V. They are:

- The number of planes participating in combat is large so that sampling errors can be neglected.
- The probability that a hit will not down the plane does not depend on the number of previous non-destructive hits. That is, $q_1 = q_2 = \ldots = q_o$ (say), where q_i is the conditional probability that the i-th hit will not down the plane, knowing that the plane is hit.
- The division of the plane into several parts is representative of all planes of the mission.

The fourth assumption necessary to apply the method of part V directly usually cannot be fulfilled in practice. It is:

- Given that a shot has hit the plane, the probability that it hit a particular part, and was fired from a particular type of gun, is known.

These probabilities depend upon the proportions in which different guns are used by the enemy. To overcome this difficulty a method that does not depend on these proportions is developed in part VIII. The assumptions necessary for the method of part VIII differ from those of part V only in that the fourth assumption is replaced by:

- Given that a shot has hit the plane, and given that it was fired by a particular type of gun, the probability that it hit a particular part is known.

The information necessary to satisfy this assumption is more readily available, and in the numerical example that follows a simplified method is suggested for estimating these probabilities.

The Data

The numerical example will be an analysis of a set of hypothetical data, which is based on an assumed record of damage of surviving planes of a mission of 1,000 planes dispatched to attack an enemy objective. Of the 1,000 planes dispatched, 634 (N) actually attacked the objective. Thirty-two planes were lost (L=32) in combat and the number of hits on returning planes was:

A_i = number of planes returning with i hits

$$A_o = 386 \qquad\qquad (A)$$
$$A_1 = 120$$
$$A_2 = 47$$
$$A_3 = 22$$
$$A_4 = 16$$
$$A_5 = 11$$

The total number of hits on all returning planes is

$$A_1 + 2A_2 + 3A_3 + 4A_4 + 5A_5 = \qquad\qquad (B)$$

$$120 + 2x47 + 3x22 + 4x16 + 5x11 = 399 \quad .$$

These 399 hits were made by three types of enemy ammunition:

B_1	Flak
B_2	20-mm aircraft cannon
B_3	7.9-mm aircraft machine gun

and the hits by these different types of ammunition were also recorded by part of airplane hit:

C_1	Forward fuselage
C_2	Engine
C_3	Full system
C_4	Remainder

The necessary information from the record of damage is given in table 7.

TABLE 7

NUMBER OF HITS OF VARIOUS TYPES BY PARTS

	Forward fuselage, C_1	Engine, C_2	Fuel system, C_3	Remainder, C_4	Total all parts
Flak, B_1	17	25	50	202	294
20-mm cannon, B_2	8	7	17	18	50
7.9-mm machine gun, B_3	7	13	17	18	55
Total all types	32	45	84	238	399

A Method of Estimating the Probability of Hitting a Particular Part Given That a Shot of a Particular Ammunition Has Hit the Plane[1]

The conditional probability that a plane will be hit on the i-th area, knowing that the hit is of the j-th type, must be determined from other sources of information than the record of

[1] Necessary for fourth assumption.

damage. Although a simplified method is used in this example, more accurate estimates can be made if more technical data is at hand. The first step is to make definite boundaries for the areas C_1, C_2, C_3, C_4. Next, assume that each type of enemy fire B_1, B_2, B_3 has an average angle of fire θ_1, θ_2, θ_3. Finally, assume that the probability of hitting a part of the plane from a given angle is equal to the ratio of the exposed area of that part from the given angle to the total area exposed from that angle.

In this example it is assumed that flak (B_1) has the average angle of attack of 45 degrees in front of and below the plane, whereas 20-mm cannon and 7.9-mm machine gun fire both hit the plane head-on on the average. The area C_1 is so bounded that it includes areas which, if hit, will endanger the pilot and co-pilot. Area C_2 includes the engine area and area C_3 consists essentially of the area covering the fuel tanks. The results of computations, based on the above assumptions, are assumed to be as follows, where $\gamma(C_i|B_j)$[1] represents the probability that a hit is on part C_i knowing it is of type B_j (as estimated by determining the ratio of the area of C_i to the total area as viewed from the angle θ_j associated with ammunition B_j).

(C)

$\gamma(C_1\|B_1) = .058$	$\gamma(C_1\|B_2) = .143$	$\gamma(C_1\|B_3) = .143$
$\gamma(C_2\|B_1) = .092$	$\gamma(C_2\|B_2) = .248$	$\gamma(C_2\|B_3) = .248$
$\gamma(C_3\|B_1) = .174$	$\gamma(C_3\|B_2) = .303$	$\gamma(C_3\|B_3) = .303$
$\gamma(C_4\|B_1) = .676$	$\gamma(C_4\|B_2) = .306$	$\gamma(C_4\|B_3) = .306$

[1]This notation differs from the previous notation of part VIII. In the first part of part VIII, $\gamma(i|j)$ is used with the understanding that the first subscript refers to the part hit and the second subscript refers to the type of bullet. In the numerical example, the relationship is made explicit by letting C_i stand for the i-th part (or component) and B_j for the j-th type of bullet. The same device is used throughout this example.

Computations for Method of Part VIII

Let $q(C_i, B_j)$ be the probability of surviving a hit on part C_i by gun B_j. By equation 104, we have

$$q(C_i, B_j) = \frac{\delta(C_i | B_j)}{\gamma(C_i | B_j)} q(B_j) \quad , \tag{D}$$

where $\delta(C_i | B_j)$ is the probability of being hit on part C_i, knowing that the hit was scored by a bullet from gun B_j and that the plane survived; $\gamma(C_i | B_j)$ is the probability of being hit on part C_i, knowing that the hit was scored by a bullet of type B_j; and $q(B_j)$ is the probability that a plane will survive a hit of type B_j, knowing that the plane is hit. This can be estimated by taking the ratio of the number of hits of type B_j on part C_i to the total number of hits of type B_j on returning planes. Applying this method to the table we obtain

(E)

$\delta(C_1 \| B_1) = .058$	$\delta(C_1 \| B_2) = .160$	$\delta(C_1 \| B_3) = .127$
$\delta(C_2 \| B_1) = .085$	$\delta(C_2 \| B_2) = .140$	$\delta(C_2 \| B_3) = .236$
$\delta(C_3 \| B_1) = .170$	$\delta(C_3 \| B_2) = .340$	$\delta(C_3 \| B_3) = .309$
$\delta(C_4 \| B_1) = .687$	$\delta(C_4 \| B_2) = .360$	$\delta(C_4 \| B_3) = .327$

The final quantity required to calculate $q(C_i, B_j)$ by equation D is $q(B_j)$. By equation 109, we have

$$q(B_j) = \lambda \rho(B_j) \quad , \tag{F}$$

where $\rho(B_j)$ is the minimum of $\dfrac{\gamma(C_i | B_j)}{\delta(C_i | B_j)}$ with respect to i.

$$\rho(B_j) = \min \left\{ \frac{\gamma(C_1|B_j)}{\delta(C_1|B_j)}, \frac{\gamma(C_2|B_j)}{\delta(C_2|B_j)}, \frac{\gamma(C_3|B_j)}{\delta(C_3|B_j)}, \frac{\gamma(C_4|B_j)}{\delta(C_4|B_j)} \right\}$$

$$\begin{aligned}
\rho(B_1) &= \min \left\{ \frac{.058}{.058}, \frac{.092}{.085}, \frac{.174}{.170}, \frac{.676}{.687} \right\} \\
&= \min \{ 1, >1, >1, .984 \} \\
&= .984 \\
\rho(B_2) &= \min \left\{ \frac{.143}{.160}, \frac{.248}{.140}, \frac{.303}{.340}, \frac{.306}{.360} \right\} \\
&= \min \{ .894, >1, .891, .850 \} \\
&= .850 \\
\rho(B_3) &= \min \left\{ \frac{.143}{.127}, \frac{.248}{.236}, \frac{.303}{.309}, \frac{.306}{.327} \right\} \\
&= \min \{ >1, >1, .981, .936 \} \\
&= .936
\end{aligned} \qquad (G)$$

The constant multiplier λ is defined by equation 115

$$\lambda = q \sum \frac{\delta(B_j)}{\rho(B_j)}, \qquad (H)$$

where $\delta(B_j)$ is the conditional probability that a hit is of type B_j.

The determination of q is identical with the procedure of part VII. From equation 26

$$\sum \frac{A_j}{q^j} = N - A_o$$

we substitute the values of equation A:

$$248q^5 - 120q^4 - 47q^3 - 22q^2 - 16q - 11 = 0 \qquad (I)$$

The root is .930 (= q_o, say).

The values $\delta(B_j)$ are obtained directly from table 7 by taking the ratio of hits of type B_j on returning planes to the total number of hits on returning planes.

$$\begin{aligned}\delta(B_1) &= \frac{294}{399} = .737\\ \delta(B_2) &= \frac{50}{399} = .125\\ \delta(B_3) &= \frac{55}{399} = .138\end{aligned} \qquad \text{(J)}$$

Substituting the results of equations G, I, and J in equation H, we obtain:

$$\begin{aligned}\lambda &= q_o \sum \frac{\delta(B_j)}{\rho(B_j)}\\ &= .930 \left\{\frac{.737}{.984} + \frac{.125}{.850} + \frac{.138}{.936}\right\}\\ &= .930\ (1.0433)\\ &= .9703\end{aligned}$$

Substituting in equation F

$$\begin{aligned}q(B_1) &= (.9703)\ (.984) = .955\\ q(B_2) &= (.9703)\ (.850) = .825\\ q(B_3) &= (.9703)\ (.936) = .908\end{aligned} \qquad \text{(K)}$$

The probabilities $q(C_i,B_j)$ can now be determined from equation D by using the values given in equations C, E, and K.

$$q(C_i,B_j) = \frac{\delta(C_i|B_j)}{\gamma(C_i|B_j)}\, q(B_j)$$

$$q(C_1,B_1) = (.058)\ (.955)/.058 = .955$$
$$q(C_2,B_1) = (.085)\ (.955)/.092 = .882$$
$$q(C_3,B_1) = (.170)\ (.955)/.174 = .933$$
$$q(C_4,B_1) = (.687)\ (.955)/.676 = .971$$

$$q(C_1,B_2) = (.160)\ (.825)/.143 = .923$$
$$q(C_2,B_2) = (.140)\ (.825)/.248 = .466$$
$$q(C_3,B_2) = (.340)\ (.825)/.303 = .926 \qquad (L)$$
$$q(C_4,B_2) = (.360)\ (.825)/.306 = .971$$

$$q(C_1,B_3) = (.127)\ (.908)/.143 = .806$$
$$q(C_2,B_3) = (.236)\ (.908)/.248 = .864$$
$$q(C_3,B_3) = (.309)\ (.908)/.303 = .926$$
$$q(C_4,B_3) = (.327)\ (.908)/.306 = .970$$

Comments on Results

The vulnerability of a plane to a hit of type B_j on part C_i is the probability that a plane will be destroyed if it receives a hit of type B_j on part C_i. Let $P(C_i,B_j)$ represent this vulnerability. The numerical value of $P(C_i,B_j)$ is obtained from the set L and the relationship

$$P(C_i,B_j) = 1 - q(C_i,B_j) \qquad (M)$$

The vulnerability of a plane to a hit to type B_j on part C_i is given in table 8.

This analysis of the hypothetical data would lead to the conclusion that the plane is most vulnerable to a hit on the engine area if the type of bullet is not specified, and is most vulnerable to a hit by a 20-mm cannon shell if the part hit is not specified. The greatest probability of being destroyed is .534, and occurs when a plane is hit by a 20-mm cannon shell on the engine area. The next most vulnerable event is a hit by a 7.9-mm machine gun bullet on the cockpit. These, and other conclusions that can be made from the table of vulnerabilities derived by the method of analysis of part VIII, can be used as guides for locating protective armor and can be used to make a prediction of the estimated loss of a future mission.

TABLE 8

VULNERABILITY OF A PLANE TO A HIT OF A SPECIFIED TYPE ON A SPECIFIED PART

	Forward fuselage	Engine	Fuel system	Remainder	Vulnerability to specified type of hit when area is unspecified
Flak, B_1	.045	.118	.067	.029	.045
20-mm cannon, B_2	.077	.534	.074	.029	.175
7.9-mm machine gun, B_3	.194	.136	.074	.030	.092
Vulnerability to hit on specified area when type of hit is unspecified[a]	.114	.179	.074	.038	.070[b]

[a]These vulnerabilities are calculated using the method of part V, and assuming that the $\gamma(C_i)$, the probability that part C_i is hit, knowing that the plane is hit, are as follows:

$$\gamma(C_1) = .084 \quad \gamma(C_2) = .128 \quad \gamma(C_3) = .212 \quad \gamma(C_4) = .576 \quad .$$

[b]This is the probability that a plane will be destroyed by a hit, when neither the part hit nor the type of bullet is specified.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: RC 432	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): A REPRINT OF "A METHOD OF ESTIMATING PLANE VULNERABILITY BASED ON DAMAGE OF SURVIVORS" BY ABRAHAM WALD	5. TYPE OF REPORT & PERIOD COVERED: Reprint - 1943
	6. PERFORMING ORG. REPORT NUMBER: See Block 18
7. AUTHOR(s): Abraham Wald	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Statistical Research Group/National Defense Research Committee, Applied Mathematics Panel	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Center for Naval Analyses, 2000 N. Beauregard Street, Alexandria, Virginia 22311	12. REPORT DATE: July 1980
	13. NUMBER OF PAGES: 91
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Office of the Chief of Naval Operations (Op03), Department of the Navy, Washington, D.C. 20350	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

1) SRG Memo 85/AMP Memo 76.1
2) SRG Memo 87/AMP Memo 76.2
3) SRG Memo 88/AMP Memo 76.3
4) SRG Memo 89/AMP Memo 76.4
5) SRG Memo 96/AMP Memo 76.5
6) SRG Memo 103/AMP Memo 76.6
7) SRG Memo 109/AMP Memo 76.7
8) SRG Memo 126/AMP Memo 76.8

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

aircraft, equations, kill probabilities, probability, vulnerability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This research contribution consists of a series of eight memoranda originally published by the Statistical Research Group at Columbia University for the National Defense Research Committee in 1943 on methods of estimating the vulnerability of various parts of an aircraft based on damage to surviving planes. The methodology presented continues to be valuable in defense analysis and, therefore, has been reprinted by the Center for Naval Analyses in order to achieve wider dissemination.

02 043200.00